



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Eđitim Bilimleri Anabilim Dalı

Eđitimde Ölçme ve Deđerlendirme Bilim Dalı

**DIKEY ÖLÇEKLEMEDE MADDE TEPKİ KURAMINA DAYALI  
FARKLI KALİBRASYON VE YETENEK KESTİRİM  
YÖNTEMLERİNİN KARŞILAŞTIRILMASI**

Ayşegül ALTUN

Doktora Tezi

Ankara, 2013

DİKEY ÖLÇEKLEMEDE MADDE TEPKİ KURAMINA DAYALI FARKLI  
KALİBRASYON VE YETENEK KESTİRİM YÖNTEMLERİNİN  
KARŞILAŞTIRILMASI

Ayşegül ALTUN

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü  
Eğitim Bilimleri Anabilim Dalı  
Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Doktora Tezi

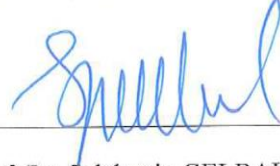
Ankara, 2013

## KABUL VE ONAY

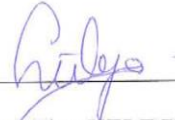
Ayşegül ALTUN tarafından hazırlanan “Dikey Ölçeklemede Madde Tepki Kuramına Dayalı Farklı Kalibrasyon ve Yetenek Kestirim Yöntemlerinin Karşılaştırılması” başlıklı bu çalışma 04.07.2013 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Doktora Tezi olarak kabul edilmiştir.



Prof. Dr. Şener BÜYÜKÖZTÜRK (Başkan)



Prof. Dr. Selahattin GELBAL



Doç. Dr. Hülya KELEÇİOĞLU (Danışman)



Doç. Dr. Nuri DOĞAN



Yrd. Doç. Dr. Burecu ATAR

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Yusuf ÇELİK

Enstitü Müdürü

## BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezin/raporumun kağıt ve elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezin/Raporumun tamamı her yerden erişime açılabilir.
- Tezim/Raporum sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezin/Raporumun ..... yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezin/raporumun tamamı her yerden erişime açılabilir.

04/07/2013

  
Ayşegül ALTUN

## TEŞEKKÜR

Öncelikle bu çalışmam sırasında defalarca tezimi gözden geçiren, umutsuzluğa düştüğüm, paniklediğim zamanlarda sabırla ve sakince beni dinleyip destek olan ve yol gösteren danışmanım Doç.Dr. Hülya Kelecioğlu'na desteği, fikirleri ve önerileri için çok teşekkür ederim. Sizin katkılarınız olmadan bu tez tamamlanamazdı.

İhtiyacım olan her zaman ve her konuda bilgisine ve tecrübesine danıştığım desteğini ve yardımlarını hiçbir zaman benden esirgemeyen hocam Prof. Dr. Aysun Umay'a ne kadar teşekkür etsem azdır.

Doktora eğitimim boyunca bana kattığı her şey için Prof.Dr. Selahattin Gelbal'a, teşekkür ederim.

Bana inandıkları, yaptığım her şeyde ve her zaman bana destek oldukları, bunu her zaman bana hissettirdikleri için, beni ben olduğum için seven ve olduğum gibi kabul eden annem ve babama çok teşekkür ederim.

Sadece doktora sürecinde değil tanıştığımız andan itibaren beni dinleyen, anlayan, eleştiren, düzelten seçtiğim kardeşim Gülçin çok teşekkürler.

Ayrıca desteğini, sevgisini bilgisini benimle paylaşan arkadaşlarım, meslektaşlarım ve akrabalarım teşekkür ederim.

Doktora süresince birçok problemle ve beni zorlayan durumla karşılaştım. Bunların bir kısmıyla baş edebildim, bazen de hatalar yaptım. Ancak her durumda birçok şey öğrendim. Şuanda olduğum ben olmama yardımcı olan bu sürece ve bu sürecin tüm aktörlerine de teşekkür ederim.

## ÖZET

ALTUN, Ayşegül. *Dikey Ölçeklemede Madde Tepki Kuramına Dayalı Farklı Kalibrasyon ve Yetenek Kestirim Yöntemlerinin Karşılaştırılması*, Doktora Tezi, Ankara, 2013.

Bu araştırma kapsamında; ortak madde deseninde madde tepki kuramına dayalı ayrı ve eş zamanlı kalibrasyon ve Expected A Posteriori (EAP), Maximum A Posteriori (MAP) ve Maksimum Olabilirlik (MO) yetenek kestirim yöntemleri kombinasyonu ile elde edilen dikey ölçekleme sonuçlarının karşılaştırılması yapılmıştır. Bu karşılaştırmayı yapabilmek için öncelikle 2008-2010 yılları arasında yapılan SBS sınavında sorulan sorulardan 6., 7. ve 8. sınıf matematik testleri oluşturulmuştur. Daha sonra 6. sınıftaki 503, 7. sınıftaki 502 ve 8. sınıftaki 500 öğrencinin zorluk düzeyleri farklı olan matematik testlerine verdikleri cevaplardan elde edilen puanlar kullanılarak dikey ölçekler geliştirilmiştir. Dikey ölçekleme süreciyle aynı ölçeğe yerleştirilen bu puanlardan elde edilen ölçek puanları kullanılarak ortalamalar, ortalamalar arasındaki fark, etki büyüklükleri ve yatay uzaklıklar hesaplanmış ve böylece ölçekleme sürecinde yapılan seçimlerin dikey ölçeklemeyi nasıl etkilediği belirlenmeye çalışılmıştır.

Kalibrasyon yöntemleri karşılaştırıldığında her iki kalibrasyon yönteminde de 6. sınıftan 7. sınıfa başarıda azalma, 7. sınıftan 8. sınıfa ise artma olduğu görülmektedir. Standart sapmalar incelendiği zaman her iki kalibrasyon yöntemi içinde 7. sınıf daha homojen iken her üç sınıf düzeyi için eş zamanlı kalibrasyon ayrı kalibrasyona kıyasla daha büyük standart sapma değerleri vermektedir. Etki büyüklükleri incelendiği zaman ise her iki kalibrasyon yönteminde de 6 ve 7. sınıflar arasında negatif, 7 ve 8. sınıflar arasında ise pozitif değerler aldığı görülmektedir. Ancak ayrı kalibrasyondaki etki büyüklükleri eş zamanlı kalibrasyona göre daha düşüktür. Ortalama yatay uzaklıklar etki büyüklüğü ile benzer sonuçlar vermektedir. Yatay uzaklıklar incelendiği zaman ayrı kalibrasyonda 6 ve 7. sınıflar arasında başarı düzeyi arttıkça yetenek düzeyleri arasındaki fark 6. sınıf lehine artmaktadır. 7 ve 8. sınıflarda

ise başarı düzeyi arttıkça fark da artmaktadır. Eş zamanlı kalibrasyonda yatay uzaklıkların 6 ve 7. sınıflar arasında negatif ve mutlak değerce azalan değerler alması 6 ve 7. sınıflar arasındaki farkın yedinci sınıf lehine yavaş yavaş arttığı 7 ve 8. sınıflar arasında pozitif azalan değerler alması da yetenek düzeyleri arasındaki farkın 7 ve 8. sınıflar arasında azaldığı anlamına gelmektedir.

Yetenek kestirim yöntemleri incelendiği zaman, ortalamalar açısından MAP çoğunlukla EAP ve MO'ya kıyasla daha düşük değerler üretirken EAP ile MO karşılaştırıldığında EAP ile hesaplanan değerlerin daha düşük olduğu görülmektedir. Standart sapma değerleri incelendiği zaman ayrı kalibrasyonda MO, EAP ve MAP'a göre daha büyük değerler verirken eş zamanlı kalibrasyonda yetenek kestirimleri için böyle bir örüntü söz konusu değildir. Etki büyüklükleri ise ayrı kalibrasyonda EAP > MAP > MO şeklinde iken eş zamanlı kalibrasyonda EAP MO'ya kıyasla daha büyük değerler almaktadır. Ortalama yatay uzaklık değerleri ise ayrı kalibrasyonda EAP > MO > MAP şeklinde iken eş zamanlı kalibrasyonda MAP > EAP > MO şeklinde değerler almaktadır. Yatay uzaklıklarda ise genellikle ayrı kalibrasyonda EAP > MO > MAP şeklinde bir örüntü varken eş zamanlı kalibrasyonda MAP > EAP > MO şeklinde bir örüntü söz konusudur.

### **Anahtar Sözcükler**

Dikey Ölçekleme, Madde Tepki Kuramı (MTK), Seviye Belirleme Sınavı (SBS)

## ABSTRACT

ALTUN, Ayşegül. *A Comparison of Different Calibration Methods and Proficiency Estimators Based on Item Response Theory in Vertical Scaling*,, Phd thesis, Ankara, 2013.

In this study the comparison of the vertical scales, which are obtained through the combination of separate and concurrent calibration based on item response theory and EAP, MAP and MO proficiency estimation methods, take place. For this comparison firstly, math tests for the 6<sup>th</sup>, 7<sup>th</sup> and 8<sup>th</sup> grades were composed from the questions asked in SBS (high school entrance exam) between 2008-2010 years. Then, the vertical scales were developed by using the scores obtained from the answers of 503 6th grade, 502 7th grade and 500 8th grade students to the math tests in different difficulty levels. By using the scale scores which were obtained from these scores placed in the same scale with the vertical scaling process, means and the difference between the means, effect sizes and horizontal distances were calculated so it was tried to be determined how the choices in scaling process affected the vertical scaling.

When calibration methods are compared it is noticed that there is a decrease in the achievements from the 6<sup>th</sup> to the 7<sup>th</sup> grade however there is an increase from the 7<sup>th</sup> to the 8<sup>th</sup> grade in both calibration methods. Examining the standard deviations while the 7<sup>th</sup> grade is more homogenous in both calibration methods, concurrent calibration gives larger standard deviation values than separate calibration for all three grades. When the effect sizes are examined it is noticed that it takes negative values between the 6<sup>th</sup> and 7<sup>th</sup> grades however it takes positive values between the 7<sup>th</sup> and 8<sup>th</sup> grades. On the other hand effect sizes in the separate calibration are lower than those in the concurrent calibration. Mean horizontal distances give the same results with the effect size. When



the horizontal distances are examined, as the achievement level between the 6<sup>th</sup> and 7<sup>th</sup> grades increases the difference between the proficiency levels also increases in favor of the 6<sup>th</sup> grades in separate calibration. Besides, in the 7<sup>th</sup> and 8<sup>th</sup> grades the difference also increases with the increase in the achievement level. In concurrent calibration the horizontal distances between the 6<sup>th</sup> and 7<sup>th</sup> grades taking negative and decreasing values (according to absolute value) which means that the difference between the 6<sup>th</sup> and the 7<sup>th</sup> grades increases slowly in favor of the 7<sup>th</sup> grades besides taking positively decreasing values between the 7<sup>th</sup> and the 8<sup>th</sup> grades which means that the difference between the proficiency levels decreases between the 7<sup>th</sup> and the 8<sup>th</sup> grades.

When proficiency estimation methods are examined, in terms of means while MAP usually creates lower values comparing with EAP and MO it is noticed that values calculated with EAP are lower. According to the examination of the standard deviation values while MO gives larger values than EAP and MAP such a pattern is not there for the proficiency estimations in the concurrent calibration. Although in the separate calibration the effect sizes are in the form of  $EAP > MAP > MO$ , in concurrent calibration EAP takes larger values comparing to MO. While mean horizontal distance values are in the form of  $EAP > MO > MAP$  in the separate calibration, they take values in the form of  $MAP > EAP > MO$  in the concurrent calibration. In horizontal distances there is a pattern in the form of  $EAP > MO > MAP$  in separate calibration while the pattern is in the form of  $MAP > EAP > MO$  in concurrent calibration.

**Key Words:** Vertical scaling, Item and Response Theory (IRT), SBS(high school entrance exam)

## İÇİNDEKİLER

<b>KABUL VE ONAY</b> .....	<b>i</b>
<b>BİLDİRİM</b> .....	<b>ii</b>
<b>TEŞEKKÜR</b> .....	<b>iii</b>
<b>ÖZET</b> .....	<b>iv</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>İÇİNDEKİLER</b> .....	<b>viii</b>
<b>KISALTMALAR DİZİNİ</b> .....	<b>xii</b>
<b>TABLolar DİZİNİ</b> .....	<b>xiii</b>
<b>BÖLÜM I</b>	<b>1</b>
<b>GİRİŞ</b>	<b>1</b>
<b>1.1 DİKEY ÖLÇEKLEME</b> .....	<b>1</b>
<b>1.1.1 Veri Toplama Desenleri</b> .....	<b>5</b>
<b>1.1.1.1 Ortak madde deseni</b> .....	<b>5</b>
<b>1.1.1.2 Ölçekli test deseni ( Scaling design)</b> .....	<b>6</b>
<b>1.1.1.3 Eşit grup deseni</b> .....	<b>7</b>
<b>1.1.1.4 Veri toplama desenlerinin karşılaştırılması</b> .....	<b>7</b>
<b>1.1.2 Ölçekleme Yöntemleri</b> .....	<b>8</b>
<b>1.1.2.1 Hieronymus ölçekleme</b> .....	<b>8</b>
<b>1.1.2.2 Thurstone ölçekleme</b> .....	<b>9</b>
<b>1.1.2.3 MTK ile ölçekleme</b> .....	<b>10</b>
<b>1.1.2.3.1 Madde tepki kuramı (MTK)</b> .....	<b>11</b>
<b>1.1.2.3.2 Kalibrasyon yöntemleri</b> .....	<b>12</b>
<b>1.1.2.3.2.1 Ayır kalibrasyon (Separate Calibration)</b> ....	<b>12</b>
<b>1.1.2.3.2.2 Eş zamanlı kalibrasyon (Concurrent Calibration)</b> .....	<b>15</b>
<b>1.1.2.3.2.3 Kalibrasyon yöntemlerinin karşılaştırılması</b> .....	<b>16</b>
<b>1.1.2.3.3 Yetenek kestirimi</b> .....	<b>17</b>
<b>1.1.2.3.3.1 Maksimum olabilirlik kestirimi (Maximum Likelihood Estimation)</b> .....	<b>18</b>
<b>1.1.2.3.3.2 Maximum a posteriori (MAP)</b> .....	<b>19</b>

1.1.2.3.3.3 Expexted A posteriori (EAP).....	21
1.1.2.3.4 Değerlendirme ölçütleri.....	23
1.1.2.3.4.1 Bir sınıf düzeyinden diğer sınıf düzeyine olan büyüme.....	23
1.1.2.3.4.2 Sınıf düzeyleri arasındaki farklılık (grade to grade variability).....	24
1.1.2.3.4.3 Sınıf dağılımları arasındaki ayırım (separation of grade distrubution).....	25
1.1.2.3.4.3.1 Etki Büyüklüğü (Effect Size)...	25
1.1.2.3.4. 3. 2 Yatay Uzaklık (Horizantal Distance).....	26
1.1.3 Dikey ölçeklemedeki tehditler.....	28
1.1.3.1 Ölçek duyarlılık farkları (Score precision diferences)...	28
1.1.3.2 Çok boyutluluk (multidimensionality).....	28
1.1.3.3 Yapı farklılıkları (Construct diferences ).....	29
1.1.3.4 Bilmediği içeriğe ait soruyla karşılaşmak.....	29
1.1.3.5 Pratik ve yorgunluk etkisi.....	29
1.2 PROBLEM DURUMU.....	30
1.3 ARAŞTIRMANIN AMACI VE ÖNEMİ.....	31
1.4 PROBLEM CÜMLESİ	32
1.4.1 Alt problemler.....	32
1.5 SAYILTI.....	33
1.5 SINIRLILIK.....	33
1.5 İLGİLİ ARAŞTIRMALAR.....	33
<b>BÖLÜM II</b>	<b>54</b>
<b>YÖNTEM</b>	<b>54</b>
2.1 ARAŞTIRMA TÜRÜ.....	54
2.2. ÇALIŞMA GURUBU.....	54
2.3. ARAŞTIRMA DESENİ.....	55
2.4 VERİ TOPLAMA ARACININ HAZIRLANMASI VE VERİLERİN	

<b>TOPLANMASI.....</b>	<b>57</b>
<b>2.5 VERİLERİN ANALİZİ.....</b>	<b>59</b>
<b>2.5.1 Test ve Madde İstatistikleri.....</b>	<b>59</b>
<b>2.5.2 MTK Varsayımlarının Test Edilmesi.....</b>	<b>62</b>
<b>2.5.2.1 Tek Boyutluluk.....</b>	<b>63</b>
<b>2.5.2.2 Yerel Bağımsızlık.....</b>	<b>65</b>
<b>2.5.3 Model Veri Uyumu.....</b>	<b>65</b>
<b>2.6 KULLANILAN YÖNTEMLER.....</b>	<b>67</b>
<b>BÖLÜM III</b>	<b>72</b>
<b>BULGULAR ve YORUM</b>	<b>72</b>
<b>3.1 ALT PROBLEM 1'E AİT BULGULAR VE YORUM.....</b>	<b>72</b>
<b>3.2 ALT PROBLEM 2'YE AİT BULGULAR VE YORUM.....</b>	<b>78</b>
<b>3.3 YETENEK KESTİRİM YÖNTEMLERİ VE KALİBRASYON</b>	
<b>YÖNTEMLERİNİN KARŞILAŞTIRILMASI.....</b>	<b>84</b>
<b>BÖLÜM IV</b>	<b>91</b>
<b>SONUÇ VE ÖNERİLER</b>	<b>91</b>
<b>4.1 SONUÇLAR.....</b>	<b>91</b>
<b>4.2 ÖNERİLER.....</b>	<b>93</b>
<b>4.2.1 Araştırma Sonuçlarından Çıkan Öneriler.....</b>	<b>93</b>
<b>4.2.2 Araştırmacılara Yönelik Öneriler.....</b>	<b>93</b>
<b>KAYNAKÇA</b>	<b>95</b>
<b>ÖZGEÇMİŞ</b>	<b>103</b>
<b>EK-1 : Eş zamanlı kalibrasyona ait veri giriş şeması.....</b>	<b>104</b>
<b>EK-2. 6. Sınıf Testi.....</b>	<b>105</b>

<b>EK-3.</b> 7. Sınıf Testi.....	<b>113</b>
<b>EK-4 :</b> 8. Sınıf Testi.....	<b>122</b>
<b>EK-5 :</b> 6-8 Eş Zamanlı Kalibrasyona Ait Örnek BILOG-MG Komutları.....	<b>130</b>

## KISALTMALAR DİZİNİ

SBS	: Seviye Belirleme Sınavı
MTK	: Madde Tepki Kuramı
MO	: Maksimum Olabilirlik
MAP	: Maksimum A Posteriori
EAP	: Expected A Posteriori
SL	: Stocking Lord Dönüşümü
MM	: Mean/Mean (Ortalama/ Ortalama) Dönüşümü
MS	: Mean/Sigma (Ortalama/Sigma) Dönüşümü
QD	: Quadrature distribution
HD	: Horizontal Distance (Yatay Uzaklık)

## TABLOLAR DİZİNİ

Tablo 1 Ortak ve ortak olmayan maddelerin testlere dağılımı.....	56
Tablo 2 Testlerdeki ortak olmayan maddelerin öğrenme alanlarına göre dağılımları.....	58
Tablo 3 Testlerde bulunan ortak madde sayıları ve öğrenme alanları.....	59
Tablo 4 Sınıflara Göre Puanların Betimsel İstatistikleri.....	60
Tablo 5 Uygulanan Test Maddelerinin Madde Güçlük ve Madde Ayrıcılık Değerleri.....	61
Tablo 6. Ayrı Kalibrasyona Ait Sınıflara Göre Puanların Uyum İstatistikleri...	65
Tablo 7 MTK'nın 2 parametrelili modeline göre hesaplanan madde parametreleri ölçek değerleri özeti.....	66
Tablo 8 SL dönüşümü için Kullanılan A ve B sabitleri.....	70
Tablo 9 Ayrı Kalibrasyonda Sınıf Düzeylerine Göre Ortalamaları, Ortalamalar Arası Farkları ve Standart Sapmaları.....	73
Tablo 10 Ayrı Kalibrasyonda Ard arda Gelen Sınıf Düzeyleri için Etki Büyüklükleri ve Ortalama Yatay Uzaklıklar.....	76
Tablo 11 Ayrı Kalibrasyonda Seçilen Yüzdelerdeki Yatay Uzaklıklar.....	77
Tablo 12 Eş Zamanlı Kalibrasyonda Sınıf Düzeylerine Göre $\theta$ Ortalamaları, Ortalamalar Arası Farkları ve Standart Sapmaları.....	79
Tablo 13 Eş Zamanlı Kalibrasyonda Ard arda Gelen Sınıf Düzeyleri için Etki Büyüklükleri ve Ortalama Yatay Uzaklıklar.....	81
Tablo 14 Eş Zamanlı Kalibrasyonda Seçilen Yüzdelerdeki Yatay Uzaklıklar.....	83

# BÖLÜM 1

## GİRİŞ

Öğrencilerin herhangi bir konu alanındaki başarılarını belirlemek için çeşitli testler uygulanmaktadır. Bu testlerden alınan ham puanlardan elde edilen bilgiler ise sınırlıdır. Hemen hemen tüm geniş ölçekli değerlendirme raporlarında ölçek puanları rapor edilir çünkü ölçek puanları ham puanların sağlamadığı bilgiyi sağlamaktadır. Ham puanlar ölçek puanlarına dönüştürüldükten sonra farklı formlardan elde edilen puanlar arasında karşılaştırma yapılabilmektedir. Ölçekleme, ölçek puanı elde edilirken sınava girenlerin performansının sayılar ya da diğer göstergelerle eşleştirilmesi sürecidir. Bu sayılar veya göstergelerin artan başarı veya yeterlik düzeylerini yansıttıkları varsayılmaktadır. Ölçekleme süreci ölçek puanlarına ek olarak farklı bilgiler de içermektedir. Örneğin sınava girenlerin puanları ilgili referans grubunun puanıyla karşılaştırılacaksa normatif bilgi içerir. Sınava girenlerin performansı daha önceden belirlenen performans standartlarıyla karşılaştırılacaksa içerik bilgisi eklenmiş olur. Testler bir test bataryasına ait ise ölçek puanları öğrencilerin konu alanlarındaki güçlü ve zayıflıklarının yorumlanmasına da olanak sağlar (Tong ve Kolen, 2010). Öğrencilerin bir sınıf düzeyinden diğer sınıf düzeyine geçişte başarılarındaki değişim ise dikey ölçekleme ile belirlenmektedir.

### 1.1 DİKEY ÖLÇEKLEME

Ana sınıfından 12. sınıfa kadar olan büyük çaplı değerlendirmelerde test geliştiriciler öğrencilerin akademik gelişimlerini belirlemekle ilgilenmektedirler. Yıllar arasında



karşılaştırma yapabilmek veya sınıf düzeyine bakılmaksızın tüm test puanlarını aynı ölçekte gösterebilmek için sınıf düzeyleri arasında (örneğin 3. sınıftan 8. sınıfa kadar), geniş bir aralıktaki tüm öğrenci performansları için tek bir ölçek puanı elde etmek gerekmektedir. Böyle bir ölçek geliştirme sürecine ve bütün sınıf düzeylerindeki değerlendirme puanlarını böyle bir ölçeğe yerleştirmeye dikey ölçekleme denir (McBride ve Wise, 2001). Yani dikey ölçekleme aynı yapıyı ölçen fakat farklı içerik ve güçlük düzeyine sahip iki ve ya daha fazla testten elde edilen puanların aynı ölçek puanına yerleştirilmesidir (Kim, 2007; Tong ve Kolen, 2008). Dikey ölçekleme birçok kararı içeren karmaşık bir süreçtir. Dikey ölçek geliştirilirken verilen bu kararlar öğrenci başarısını yorumlamayı etkileyebilir (Meng, 2007).

Öğrenci başarısını ölçen test bataryalarını dikey ölçeklemenin gerekliliğinin temel nedeni test geliştiricilere öğrencilerin eğitim hayatları boyunca başarılarındaki ilerlemeyi izleyebilecek gelişimsel ölçek puanı (developmental score scale) sağlamaktır (Loyd ve Hoover, 1980).

Dikey ölçekleme, aynı sınıf düzeyinin yıldan yıla gelişimini göstermek için elde edilen puanların aynı metrik üzerine yerleştirilmesi için kullanılan ölçek kalibrasyonu yöntemiyle benzerdir fakat materyal olarak farklıdır. Madde kalibrasyon süreçleri benzerdir, ancak prensiplerindeki farklılıklar önemlidir. Yatay kalibrasyon aynı zorluk düzeyine sahip formlardan elde edilen puanları aynı ölçeğe yerleştirir. Bu yerleştirme işlemi yapılırken de test eşitleme sürecinden faydalanılır (McBride ve Wise, 2001).

Psikometrik eşitleme sürecinde farklı test formlarından elde edilen ham puanlar arasında ilişki kurulur. Test eşitleme yapıldıktan sonra farklı formlardan elde edilen ham puanları ölçek puanına dönüştürmek için ölçekleme işlemi uygulanır. Dönüştürme işlemi lineer veya non-lineer olabilir. (Tong ve Kolen, 2010). Dikey ve yatay süreçlerin amacı test puanlarını bir ölçekten diğerine dönüştürmek için matematiksel bir eşitlik geliştirmektir. Bu eşitlikler genellikle lineerdir ve bu eşitliklerin elde edilme süreçleri benzerdir. Bu benzerliklerinin yanında dikey ve yatay süreçlerin çok önemli iki farkı vardır. Bunlardan biri test formlarının güçlük düzeylerinin farklı olması diğeri ise farklı sınıf düzeylerindeki öğrencilerin genel yetenek düzeyleri arasındaki farktır.

Yatay kalibrasyon probleminde farklı test formları içerik olarak ve güçlük düzeyi olarak benzerdir ve formlar arasındaki öğrencilerin genel yetenek düzeyleri arasındaki fark çok azdır. Ancak dikey ölçekleme sınıf düzeyi değiştikçe öğrencilerin değişen yetenek düzeyleriyle ve güçlük düzeyleri bakımından farklı olan testlerle ilgilenmektedir. Farklı sınıf düzeyindeki öğrencileri değerlendirmek için hazırlanan testlerin içeriği eğitim programındaki farklılıkları yansıtabilmesi için farklı olmak zorundadır. Ard arda olmayan sınıf düzeylerinde, örneğin 3. ve 7. sınıf gibi, eğitim programının ve içeriğin farklı olması önemli olabilir. Bu farklılık genellikle aynı sınıf düzeyinde yapılan ölçek kalibrasyonunda görülmeyen dikey ölçekleme girişimine ait bir problemdir (McBridge ve Wise, 2001).

Dikey ölçekleme süreci farklı düzeylerden elde edilen test puanlarını ortak bir ölçekte ilişkilendirmede kullanılan bir süreçtir. Aynı sınıf düzeyinde yapılan ölçeklemelerden farklıdır. Ölçekleme sürecindeki kullanılacak veri toplama deseni ve ölçekleme yöntemi

gibi çeşitli kararların verilmesi gerekmektedir. Bu kararların dikey ölçeklemeyi dolayısıyla da öğrenci başarısındaki gelişmeyi gösteren örüntüleri etkilediği görülmüştür (Tong ve Kolen, 2007).

Dikey ölçeklemenin sonunda aşağıdaki bilgiler elde edilir.

- 1) Öğrencilerin bireysel gelişimlerinin değerlendirilmesini sağlar. Örneğin 3. sınıftan 4. sınıfa geçerken A öğrencisinin puanının artıp artmadığının belirlenmesi gibi.
- 2) Maddelerle becerilerin sınıf düzeyleri arasında ve aynı sınıf düzeyinde tek bir ölçek üzerinde eşleştirilmesini sağlar.
- 3) Farklı sınıf düzeylerindeki yeterlik standartlarını tek bir ölçekle ilişkilendirir. Örneğin, 4. sınıfta yüksek bir başarı elde etmiş bir öğrencinin 5. sınıfta yeterli düzeyi elde etmiş bir öğrenciden daha başarılı olup olmadığına ilişkin bilgi verir.

Dikey ölçekleme çok karmaşık bir süreçtir. Bunun nedeni ise ölçek geliştirme sürecinin birçok faktöre bağlı olması ve bu faktörler arasından seçim yapılması gerekliliğidir (Tong ve Kolen, 2010). Dikey ölçekleme süreci; sınava giren grubun doğası, sınıf düzeyleri farklılaştıkça testlerin zorluk düzeylerinin de farklılık göstermesi, dikey ölçekleme sürecinde kullanılan ölçekleme yöntemi, ölçeği oluşturmak için kullanılan veri toplama deseni, yeterlik tahminleri ve kullanılan bilgisayar programları, madde tepki kuramı (MTK) ndaki kalibrasyon yöntemleri, aynı sınıf düzeyinde ve sınıflar arasında puanların boyutluluğu ve başlangıç grubunun seçilmesi gibi faktörlerden etkilenmektedir.

Dikey ölçekleme çalışmasında öncelikle veri toplama deseni tanımlanmalıdır (Tong ve Kolen, 2007).

### **1.1.1 Veri Toplama Desenleri**

Dikey ölçeklemede ortak madde deseni, ölçekli test deseni ve eşit grup deseni en yaygın olarak kullanılan veri toplama desenleridir.

#### **1.1.1.1 Ortak madde deseni**

Ortak madde deseninde, her sınıf düzeyi için o sınıf düzeyine uygun test geliştirilir ve her test ait olduğu sınıf düzeyine uygulanır. Farklı sınıf düzeylerindeki testleri birbirine bağlamak için ise ard arda gelen sınıf düzeylerinde (örneğin 6. ve 7. sınıf) ortak maddeler testte dahil edilir. Öğrencilerin ortak maddelerde gösterdikleri performans, bir sınıf düzeyinden diğer sınıf düzeyine ne kadar gelişme olduğunu göstermektedir. Bu desende en önemli karar ortak madde seçimidir. Bu bağlamda ortak maddelerin her iki sınıf düzeyinden mi, yoksa sadece bir sınıf düzeyinden mi seçileceğinin belirlenmesi önem kazanmaktadır. Öğrencilerin öğrenmedikleri konulara ait sorularla test edilmemeleri gerektiği belirtilerek ilk durum eleştirilmektedir. İkinci durum ise bu şekilde ortak madde setinin seçilmesiyle içeriğin iyi örneklemeyeceği konusunda eleştirilmektedir. Bunlara ek olarak ortak maddelerin seçiminde; eşitleme yapılırken olduğu gibi ortak maddelerin testin genelinin bir minyatürü mü olacağı ya da test edilen alanın iki sınıf düzeyinde ortak olan yapısını en iyi yansıtacak şekilde mi seçilmesi gerektiğinin kararının verilmesi gerekmektedir.

Ortak madde deseni uygulanırken bir sınıfın temel düzey olarak tanımlanması gerekmektedir. Öğrencilerin ortak maddelerdeki performansları ise ard arda gelen sınıf düzeylerinin puanlarının birbiriyle ilişkilendirilmesinde kullanılır. Daha sonra diğer sınıf düzeylerini temel düzeye bağlamak için zincirleme bir süreç başlar. Örneğin eğer 5. sınıf temel düzey olarak seçildiyse 7. sınıf puanlarını temel düzeye bağlamak için; 7. sınıf puanları 6. ve 7. sınıfta ortak olan maddeler yardımıyla 6. sınıfla aynı ölçeğe yerleştirilir. Daha sonra 6. sınıf puanları 5. sınıf puan ölçeğine 5. ve 6. sınıflarda ortak olan maddeler yardımıyla yerleştirilir. Bu zincirleme kullanılarak 7. sınıf puanları temel düzey olan 5. sınıf ölçeğine yerleştirilmiş olur. K-12 değerlendirmelerinde kullanılan dikey ölçeklemelerin çoğunda ortak madde deseni kullanılmaktadır (Tong ve Kolen, 2007).

Pratikte en çok kullanılan yöntem ortak madde desenidir. Örneğin kopya çekilmesine karşı önlem olarak alternatif formlar hazırlanmaktadır. A formunu alan bir öğrencinin B formunu alan öğrenciden daha çok ya da daha az avantajlı olmaması gerekmektedir bu nedenle formlar eşitlenmelidir. Bu yöntemin pratikteki bir diğer uygulaması ise kullanılan testlerin tamamen yenilenmesi durumunda eski testlerdeki maddelerin bir kısmının anchor madde olarak alınması şeklinde olmaktadır. Böylece bu eski teste ait maddelerin parametreleri sabit parametreler olarak alınıp eski ve yeni testlerin eşitlenmesinde kullanılabilir (Yu ve Popp, 2005).

#### **1.1.1.2 Ölçekli test deseni (Scaling design)**

Ölçekli test deseninde iki test geliştirilmesi gerekmektedir. Bu testlerden ilki tüm sınıf düzeylerindeki ilgileri kapsayacak şekilde geliştirilen özel bir testtir ve bu test ‘ölçek

testi' olarak adlandırılır. Buna ek olarak her sınıf düzeyi için o sınıf düzeyine uygun test geliştirilir. Bu desende bütün öğrencilere 'ölçek testi' ve ayrıca kendi düzeylerine uygun test uygulanır. Ölçek puanı ölçek testinden elde edilen puanlar kullanılarak tanımlanır. Bu desenin zorlayıcı kısmı en düşük sınıf seviyesinden en yüksek sınıf seviyesindeki öğrenme alanlarını kapsayan özel ölçek testi geliştirmektir (Tong ve Kolen, 2007).

### **1.1.1.3 Eşit grup deseni**

Eşit grup deseninde her sınıf seviyesindeki öğrenciler kendi sınıf seviyelerine uygun veya ardışık sınıf seviyesindeki testleri alırlar. Hangi öğrencinin hangi sınıf düzeyindeki testi alacağı rastgele belirlenir. Rastgele eşitlenen gruplar boyunca arda arda gelen sınıf seviyelerinde öğrencilerin gösterdikleri performans bir sınıf düzeyinden diğer sınıf düzeyine ne kadar gelişme gösterdiğinin belirlenmesine yardımcı olur. Bu desende de bir grubun temel grup olarak belirlenmesi gerekmektedir. Sınıf düzeyleri arasında zincirleme yapılarak bu uygulamadan elde edilen veri tüm test düzeylerindeki puanların aynı ölçeğe yerleştirilmesinde kullanılır. Bu desen ard arda gelen sınıf düzeyleri için ortak madde kullanımı gerektirmez. Ortak madde yerine ortak kişileri kullanarak bağlantıyı kurar. Bu desende öğrenciler kendi sınıf düzeyleri dışında testler aldıkları için pratikte sıkça kullanılan bir yöntem değildir (Tong ve Kolen, 2007).

### **1.1.1.4 Veri toplama desenlerinin karşılaştırılması**

Ortak madde deseni var olan gruplara uygulanabilmektedir. Eşit grup deseni ise özel uygulama gerektirmektedir. Ölçekli test deseni ise hem özel uygulama hem de özel olarak hazırlanmış bir test gerektirmektedir. Ölçekli test deseninde ve eşit grup deseninde öğrenciler kendi sınıf düzeylerine uygun olmayan birçok test almaktadırlar.

Uygulaması en zor olanı ölçekli test deseni olsa da bu desenin avantajı, konu alanındaki büyüme tanımlarını açıkça göz önüne alarak tüm sınıf düzeyindeki her öğrenciyi tek bir test üzerinde sıralamasıdır. Diğer iki desen açıkça sıralama yapılmasına imkan vermez çünkü farklı sınıf düzeyindeki öğrenciler farklı test sorularını cevaplamaktadırlar. Ölçekli test deseniyle elde edilen ölçekleme sonuçları özellikle konu alanları öğretim programıyla sıkı bir bağlantısı olduğu durumda diğer iki desene göre farklılık göstermektedir. Farklı desen kullanılması durumunda öğrenci başarıları için farklı ilerleme sonuçları elde edilmektedir. Yani öğrenci başarısındaki ilerleme sayıtları, desenler arasında farklılık göstermektedir (Tong ve Kolen, 2007).

### **1.1.2 Ölçekleme Yöntemleri**

Test geliştirilip veri toplandıktan sonra dikey ölçeği geliştirmek için istatistiksel yöntemler kullanılır. Dikey ölçekleme sürecinde hangi ölçekleme metodunun kullanılacağına da belirlenmesi gerekmektedir. Tarihsel olarak baktığımızda ilk önce Thustone -1925, Thustone -1938 yöntemi kullanılmaktadır. MTK'nın gelişmesiyle son dönemlerde sıkça MTK ölçekleme metodu kullanılmaktadır. En çok kullanılan üç istatistiksel ölçekleme yöntemi Hieronymus ölçekleme, Thurstone ölçekleme ve MTK ile ölçeklemedir. Bu genel istatistiksel yaklaşımlarda kullanılan spesifik süreç, kullanılan veri toplama desenine bağlıdır (Tong ve Kolen, 2010).

#### **1.1.2.1 Hieronymus ölçekleme**

Hieronymus ölçekleme metodu herhangi bir veri toplama deseniyle kullanılsa da bu desen ölçekli test deseniyle kullanmak için geliştirilmiştir. Ölçeklemeyi yapabilmek için her sınıf düzeyi için ölçek testi üzerindeki puanların medyanı ön ölçek puan değerleri

olarak belirlenir. Bu ön değerler ölçülen yapıya bağlı olarak test geliştiricinin medyan değerinin artış miktarındaki değişiminden beklentisidir. Örneğin ITBS (Iowa Test of Basic Skills) formunda test geliştiricilerin büyüme modeli şu şekildedir; öğrencilerin bir seneden diğer seneye başarılarındaki artış miktarı azalırken akademik başarılarını etkileyen değişkenlerin arttığı gözlenmiştir ve dikey ölçekleme de bu duruma bağlı kalınarak yapılmıştır. Büyüme azalarak devam etmektedir. Bir başka deyişle küçük sınıflarda öğrencilerin başarılarındaki artış miktarı daha fazla iken sınıf düzeyi arttıkça başarıdaki artış devam etmekle birlikte artış miktarı azalmaktadır. Sınıflar arası standart sapma sınıf düzeyi arttıkça artmaktadır. Ölçekli test puanlarının ölçek puanlarına dönüştürülmesinden sonra her sınıf düzeyindeki test puanları ölçek testi boyunca dikey ölçekle ilişkilendirmektedir. Bu testlerin bağlantısını kurma sürecinde Hieronymus ölçeklemede gerçek puanlar ölçek testi ve düzey testinden tahmin edilirler. Düzey testleri ait olduğu sınıf düzeyini hedeflediği için ölçek testlerine göre daha güvenilir puanlar sağlamaktadır. Tahmin edilen gerçek puanlar kullanılarak güvenilirlikler arasındaki bu farklılıkla baş edilmeye çalışılmaktadır (Tong ve Kolen, 2010).

### **1.1.2.2 Thurstone ölçekleme**

Thurstone ölçeklemede her sınıf düzeyindeki toplam puanların normalleştirilmesini içeren bir ölçekleme metodu tanımlanmaktadır. Bu yöntem veri toplama desenlerinin üçü için de kullanılmaktadır. Ancak hangi veri toplama deseni kullanılırsa kullanılsın bu yöntem her sınıf düzeyinde puanların normal dağıldığını ve sınıflar arasında lineer bir bağlantı olduğunu varsaymaktadır (Tong ve Kolen, 2010).



### 1.1.2.3 MTK ile ölçekleme

MTK kullanan test geliřtiriciler test performansı ile yetenek arasında matematiksel bir fonksiyon tanımlarlar. Bu fonksiyon geniş bir yetenek aralığını tahmin etmede kullanılabilir. Bu yöntem test tarafından ölçülen yetenek veya özellik ile madde arasındaki ilişkiye odaklanmıştır ve her madde sınava giren birey hakkında anlamlı bilgi sağlamalıdır (Reid, Kolakowsky-Hayner, Lewis ve Armstrong, 2007). MTK ile ölçeklemede sınava girenlerin test sorularına verdikleri tüm cevap setleri kullanılır. Bu ölçekleme yöntemi herhangi bir veri toplama deseninde uygulanabilmektedir. MTK ölçekleme metodu olarak kullanıldığında, ölçekleme süreci ve yetenek tahmin yönteminin de seçilmesi gerekmektedir. Bütün bunlar da dikey ölçeklemenin karmaşıklığını artırmaktadır.

MTK genellikle olasılık modelini içeren madde düzeyinde yapının boyutluluğunu ve cevaplayıcıların maddelere verdikleri cevapların yerel bağımsızlığı varsayımında bulunur. Dikey ölçekleme için MTK kullanıldığı zaman boyutluluk, model çeşidi, madde parametre sayısı, kalibrasyon metodu (ayrı veya eş zamanlı), puanlama yöntemi (toplam puan veya örüntü puanlama), yeterlik tahminleri (Maksimum olabilirlik, Bayesian ve Expected a posteriori...) konularında karar verilmesi gerekmektedir (Tong ve Kolen, 2007).

MTK'da yetenek ölçeği için başlangıç ve birim (unit) seçimi rastgele yapılmaktadır ve bu durum MTK'da *ölçek belirsizliği* olarak adlandırılmaktadır.

Pratikte MTK ile ölçekleme yapılırken  $\theta$  dağılımının ölçek belirsizliği ortalamasının 0 ve standart sapmanın 1 olarak ayarlanmasıyla ortadan kaldırılır. Yetenek ölçeğini bu

şekilde hazırlamak madde parametrelerinin ölçeğinin belirlenmesini sağlar (Mislev ve Block, 1990, akt; Yen, 2007). BILOG, MULTILOG, PARSCALE, WINSTEP gibi birçok MTK parametrelerini tahmin eden programlar bu uygulamayı kullanmaktadırlar (Yen, 2007).

#### **1.1.2.3.1 Madde tepki kuramı (MTK)**

Ölçme analistlerinin kullanabilecekleri MTK modellerinin sayısı bilgisayarların gelişmesi ve karmaşık data yapısının temelini daha zengin ve daha anlamlı bir şekilde yorumlanmasına olan ihtiyaç nedeniyle son 15 yılda artmıştır (Rupp, 2003).

Tek boyutlu MTK'ya dayalı olarak geliştirilen dikey ölçekleme desenlerinin en önemli sayıtlıları, diğer MTK uygulamalarında olduğu gibi, tek boyutluluk ve yerel bağımsızlıktır.

Madde tepki kuramında norm dayanaklı değerlendirme yapmak mümkündür ancak maddenin içeriğiyle ilgili olarak yetenek düzeyinin değerlendirilmesine de (yorumlanmasına da) olanak sağlamaktadır. Bu da sınava girenlerin yetenek düzeyleri ve madde güçlüklerinin aynı ölçekte olması nedeniyle karşılaştırılabilmektedir.

1PL ve 2PL modellerde bir maddenin sağladığı bilgi madde güçlüğü etrafında maksimize olur. Sonuç olarak bireyin yetenek düzeyi ile zor olarak eşleşen madde (madde birey için zor bir madde ise) en çok bilgi veren maddedir. 3PL modelde ise maksimum madde bilgisi yetenek düzeyinin madde güçlüğü altında olduğu durumda ortaya çıkmaktadır. Genel olarak c parametresi bir maddenin sağladığı bilginin diğer modellere göre daha düşük bilgi vermesine yol açar.

İkinci prensip de bir maddenin verdiği bilgi düzeyi madde ayırt edicilik indeksiyle belirlenir. Madde ne kadar ayırt edici ise maddenin sağladığı psikometrik bilgi madde güçlüğü etrafında o kadar fazladır. Yüksek ayırt ediciliğe sahip madde tepe yapmış bir bilgi eğrisi sunar ve bu tür maddeler dar bir yetenek düzeyi aralığında çok bilgi sağlar. Tersine düşük ayırt ediciliğe sahip madde daha düz ve yayılmış bir bilgi verir. (Embretson ve Reise, 2000).

### **1.1.2.3.2 Kalibrasyon yöntemleri**

Dönüşüm yoluyla parametre tahminlerini aynı ölçeğe yerleştirme süreci kalibrasyon olarak adlandırılmaktadır (Kolen & Brennan, 2004). Farklı gruplardan elde edilen verileri ortak bir ölçeğe yerleştirmek için çeşitli kalibrasyon yöntemleri kullanılmaktadır. Bunların içinde en yaygın olarak kullanılanları ayrı kalibrasyon ve eş zamanlı kalibrasyon yöntemleridir (Meng, 2007).

#### **1.1.2.3.2.1 Ayrı kalibrasyon (Separate Calibration)**

Ayrı kalibrasyonda her sınıf düzeyi için madde ve yetenek kestirimleri ayrı ayrı hesaplanmaktadır. Daha sonra bütün sınıf düzeyleri için yapılan parametre tahminlerini başlangıç ölçeğine yerleştirmek için bir dizi lineer dönüşüm yapılmaktadır. Ölçek dönüşümü yapılırken ardışık sınıf düzeyleri arasındaki ortak maddeler için iki grup madde parametre tahminleri kullanılmaktadır.

Ayrı kalibrasyon uygulanırken her sınıf düzeyi için yeterlik puan ortalamaları 0 ve standart sapmaları 1 olarak alınır. Bu ölçek (0,1) ölçeği olarak adlandırılır. Ancak (0,1) ölçeği gruba bağlı olduğu için ve aynı başlangıç ve aynı birime sahip olmadığı için parametre tahminleri aynı ölçek üzerinden olmayabilir. Bu durumla baş edebilmek için

parametre tahminleri (0,1) ölçeğinde yapılan bir temel grup seçilmekte ve diğer sınıf düzeyleri için yapılan parametre tahminleri bazı dönüşüm yöntemleriyle temel ölçeğe yerleştirilmektedir.

Ayrı kalibrasyonda sonuçlarını ortak bir ölçeğe yerleştirmek için birçok bağlama yöntemi geliştirilmiştir. Bunlar ortalama/ortalama yöntemi (MM) (Lyod ve Hover, 1980), ortalama/sigma yöntemi (MS) (Macro, 1977), Stocking-Lord metodu (SL) (Stocking ve Lord,1983) ve Haebara yöntemidir (Haebara, 1980; akt. Meng, 2007). Ortalama/ortalama ve ortalama/sigma yöntemleri moment yöntemi; Stocking-Lord ve Haebara yöntemleri de karakteristik eğri yöntemleri olarak adlandırılmaktadırlar (Meng, 2007).

MM, MS ve SL yöntemlerinden herhangi biri uygulanırken, uygun lineer dönüşüm için tahmin edilen bir grup parametre diğer parametre tahminlerinin ölçeğine yerleştirilerek eğim ( A ) ve kesişim ( B ) tanımlanır.

Lineer dönüşüm ile (0,1)  $\theta$  ölçeği bir başka (0,1)  $\theta^*$  ölçeğine dönüştürülür.

$$\theta_i^* = A \theta_i + B$$

A ve B bağlama sabitleri veya lineer fonksiyonun eğim ve kesişimidir. Buna paralel olarak ayırt edicilik parametresi  $a_j$  ve güçlük parametresi  $b_j$ 'nin de dönüştürülmesi gerekmektedir.

$$a_j^* = a_j/A$$

ve

$$b_j^* = Ab_j + B$$

Madde karakteristik eğrisinde  $\theta$ 'nın  $\theta^*$  ile  $a_j$ 'nin  $a_j^*$  ile ve  $b_j$ 'nin  $b_j^*$  ile yer değiştirilmesiyle; i. kişinin bir soruyu doğru cevaplama olasılığıyla aynı kalır (Stocking ve Lord,1983). SL metodunda verilen  $\theta$  için ortak maddelerden elde edilen test karakteristik eğrileri (Test charachteristic curve,TCC) arasındaki farkların karesi kullanılmaktadır. Stocking-Lord (SL) yöntemi de üç parametrelili modelle daha sık kullanılmaktadır (Kim, Lee, Kim, Kelly, 2009) .

Ortalama/sigma yönteminde ortak maddelerden kestirilen b parametrelerine ait ortalama ve standart sapmalara dayalı olarak A ve B katsayıları şu şekilde elde edilir (Marco, 1977).

$$A = \frac{\sigma(b_j)}{\sigma(b_i)}$$

$$B = \mu(b_j) - A\mu(b_i)$$

$\sigma(b_j)$  : Ortak maddelere ait J testinden elde edilen b parametrelerine ait standart sapma.

$\sigma(b_i)$  : Ortak maddelere ait I testinden elde edilen b parametrelerine ait standart sapma.

$\mu(b_j)$  : Ortak maddelere ait J testinden elde b parametrelerine ait ortalama.

$\mu(b_i)$  : Ortak maddelere ait I testinden elde b parametrelerine ait ortalama.

Ortalama/sigma (mean/sigma-MS) bağlama yöntemi Rasch modelle daha sık kullanılmaktadır. MS metodu ortak maddeler için b parametre tahmininin ortalama ve standart sapmasını kullanmaktadır (Kim, Lee, Kim, Kelly, 2009).

Ortalama/ortalama (mean/mean-MM) yönteminde A değeri şu şekilde elde edilir (Loyd ve Hoover, 1980).

$$A = \frac{\mu(a_i)}{\mu(a_j)}$$

$\mu(a_i)$  : Ortak maddelere ait I testinden elde edilen a parametrelerinin ortalaması.

$\mu(a_j)$  : Ortak maddelere ait J testinden elde edilen a parametrelerinin ortalaması.

MM metodu a ve b parametre tahminlerinin ortalamalarını, bağlanacak iki test arasındaki ortak maddeler için kullanmaktadır. Ortalama/ortalama bağlama yöntemi Rasch modelle daha sık kullanılmaktadır. Bu yöntem Rasch modelinde sadece b parametresini dikkate alsa da Rasch model aynı zamanda mean shift metot olarak da tanımlanmaktadır (Kim, Lee, Kim, Kelly, 2009).

#### **1.1.2.3.2.2 Eş zamanlı kalibrasyon (Concurrent Calibration)**

Eş zamanlı kalibrasyonda bütün sınıf düzeylerindeki verinin aynı anda kalibre edilmesiyle dikey ölçekleme yapılmaktadır (Meng, 2007). Bunun için kullanılan program bir kere çalıştırılmakta ve her seviyeye ait madde parametreleri ve yetenek kestirimleri dönüşümü tek bir ölçekte elde edilmektedir. Bu kalibrasyon yöntemi bütün madde parametre tahminlerinin aynı ölçek üzerinde olduğunu varsaymaktadır. Eş zamanlı kalibrasyon yapılırken öncelikle herhangi bir sınıf, referans sınıf olarak belirlenmekte ve ardışık sınıf seviyelerine ait ölçeklenmiş yetenek kestirimleri ortalaması 0 ve standart sapması 1 olan bir ölçeğe dönüştürülmektedir (Kim, 2007). Hedef testteki ortak maddeler için madde parametreleri referans testin değerlerine ayarlanıp tekrar tahmin edilmemektedir. Ortak maddeler için madde parametreleri bilinmekte ve MTK kalibrasyonları hedef testte ortak olmayan maddeleri referans test ölçeğine yerleştirmektedir (Meng 2007). Bu

nedenle eş zamanlı kalibrasyonda fazladan bir dönüştürme ve A-B sabitlerinin elde edilmesine ihtiyaç yoktur. Eş zamanlı kalibrasyon kullanıldığında bir testteki yetenek düzeyinin diğer teste dönüştürülmesi, BILOG-MG 3 gibi bilgisayar programları tarafından yapılmaktadır (Kim, Lee, Kim ve Kelly, 2009).

#### **1.1.2.3.2.3 Kalibrasyon yöntemlerinin karşılaştırılması**

Madde parametre tahminleri için en sık kullanılan yöntemler ayrı kalibrasyon ve eş zamanlı kalibrasyon yöntemleridir. Ayrı kalibrasyonda madde parametreleri her sınıf düzeyi için ayrı ayrı hesaplanmaktadır. Her sınıf düzeyini aynı ölçüğe yerleştirmek için zincirleme süreci uygulanmaktadır. Eş zamanlı kalibrasyonda ise sınıf düzeyleri boyunca sınava giren bütün öğrencilerden elde edilen veriler bir defada kalibre edilmektedir. Sınıf grupları veri üzerinde belirtilerek her sınıf düzeyi için ayrı ayrı yeterlik dağılımı hesaplanabilmektedir. Eş zamanlı kalibrasyondan sonra bütün madde parametreleri aynı MTK ölçüğindedir ve başka bir dönüştürmeye ihtiyaç yoktur.

MTK modellerinin gerekleri sağlandığı zaman eş zamanlı kalibrasyonun uygulanması daha kolaydır, daha fazla miktarda bilgi kullanır ve daha kararlı bağlama sonuçları üretir. Fakat MTK modellerinin varsayımları sağlanmadığı zaman ayrı kalibrasyon daha “doğru sonuçlar vermektedir. Kolen ve Bernana (2004) ayrı kalibrasyonun MTK’daki tek boyutluluk sayılısının sağlanamamasının etkilerini azalttığını bu nedenle bu yöntemin tercih edilebileceğini vurgulamışlardır. Ayrı kalibrasyon arda arda gelen sınıf düzeylerinde farklı çalışan ortak maddeler için farklı madde parametre tahminlerinin karşılaştırılmasına da olanak sağlamaktadır. Bunlara ek olarak eş zamanlı kalibrasyonda

özellikle de bir çok sınıf düzeyi boyunca uygulanırsa sıklıkla yakınsaklık problemi ile karşılaşmaktadır (Tong ve Kolen, 2010).

### **1.1.2.3.3 Yetenek kestirimi**

Yetenek kestiriminde değişik yaklaşımlar vardır. Madde tepki kuramı modellerinden herhangi birini uygularken toplam puanların bir fonksiyonu olarak ham puanların mı (toplam puan) yoksa bütün cevapların örüntüsünün (örüntü puanlama) mü kullanılacağına karar verilmelidir. Toplam puan basitçe madde puanlarının toplamıdır. Örüntü puanlama ise maddelere verilen belirli cevap örüntüleridir. Bir testte, verilebilecek olası cevap örüntüleri çıkarılır ve her bir örüntü için yetenek kestirimi yapılır. Toplam puan yaklaşımı ham puanlar ile ölçek puanları arasında bire bir ilişki üretir ve ham puanların ilgili kişilere açıklanması daha kolaydır. Örüntü puanlama ise daha fazla bilgi kullanır, daha doğrudur ancak kullanıcılara açıklanması zordur. Örüntü puanlama ile bireylerin madde cevaplarından yola çıkılarak yetenek düzeyleri tahmin edilmektedir. kullanarak sınav olanların yetenek düzeyleri tahmin edilmektedir. Bu yöntemde, aynı sayıda fakat farklı maddeleri doğru cevaplayanlar farklı yetenek tahminine sahip olabilmektedirler. Toplam puan ile ise toplam doğru cevap sayıları kullanılarak yetenek kestiriminde bulunmaktadır (Thissen ve Orlando, 2001). Bu yöntemde aynı doğru cevap puanına sahip öğrenciler her zaman aynı yetenek düzeyine sahip olmaktadır. Toplam puan, örüntü puanlamaya göre daha fazla ölçme hatası içerir.

Bütün MTK dayanaklı puanlama stratejileri tahmin edilen madde parametreleriyle birlikte madde cevap örüntülerini kullanarak cevaplayıcının örtük özellik boyutundaki



yerini tahmin etmeye çalışır. İki kategorili maddeleri puanlarken üç strateji kullanılır.

Bunlar

- a) Maksimum Olabilirlik (Maksimum Likelihood) (MO)
- b) Maximum a posteriori (MAP)
- c) Expected a posteriori (EAP)

MTK kullanılarak dikey ölçekleme yapılırken, kalibrasyon işlemi bitirildikten sonra model-veri uyumu sağlandıysa; sınava giren her birey için yetenek tahmininin yapılması gerekmektedir. MTK'ya dayalı modellerde bireylerin yeteneklerini kestiren birçok yol vardır. Bazı kestiriciler toplam puana dayalıdır, diğerleri ise sınava girenlerin maddelere verdikleri cevap örüntülerine dayanır (Tong ve Kolen, 2010). Maksimum olabilirlik kestirimi (MO), Expected a posteriori (EAP) ve Maximum a posteriori (MAP) kestirimlerinden biri kullanılarak gerekli yetenek kestirimleri yapılabilmektedir. Bu yöntemler farklı özelliklere sahiptirler. Aşağıda her bir yöntemin özellikleri açıklanmıştır.

#### **1.1.2.3.3.1 Maksimum olabilirlik kestirimi (Maximum Likelihood Estimation)**

MO ile puanlama bir araştırma sürecidir ve cevaplayıcının madde cevap örüntüsünün olasılığını maksimum yapan  $\theta$  değerinin bulunmasına dayanır.

Bireysel olarak MO tahmini belirlenirken  $\theta$  değeri, gözlenen madde cevaplarının ( $u_s$ ) olasılıklarının toplamı (log birimi içinde) ile belirlenir. Daha sonra en büyük olasılık fonksiyonu (örneğin mod) belirlenir ve bu değer  $\theta$  olarak alınır (Embretson ve Reise, 2000). Bir diğer ifadeyle MO ile yetenek kestirilirken,  $\theta$ 'yı tahmin etmek için olasılık fonksiyonu maksimum yapılır.

Bireysel  $\theta$ ' lar için MO bir yetenek tahminidir ve MO tahminleri uzun testler için yansızdır. Ancak MO sıfır puan alındığı veya tam puan alındığı durumlarda hesaplanamamaktadır. Bu durum, MO kestiriminin bir sınırlılığı olarak kabul edilmektedir. Aynı zamanda MO yöntemleri doğal olarak asimtotiktir ve büyük örneklemlerde (örneğin 50 madde) en iyi sonucu vermektedir. MO tahmininin değişkenliği iki bileşenden etkilenmektedir. İlk bileşen yetenek ( $\theta$ ) değişkenliği ve ikinci bileşen de ölçmedeki hataların değişkenliğidir. Sınava giren grup için MO tahmininin değişkenliği bu iki bileşenin toplamıdır. Sonuçta MO'nun değişkenliği quadrature dağılıma göre daha büyüktür ve bu büyüklük ölçme hatalarındaki değişime daha çok bağlıdır (Thissen ve Orlando, 2001). MO kestiriminin yaygın olarak kullanılmasının nedeni, asimtotik tutarlılık ve asimtotik normallik gibi özelliklerinin olmasıdır.

#### **1.1.2.3.3.2 Maximum a posteriori (MAP)**

Sınava girenlerin yetenek düzeyleri belirlenirken önsel dağılım kullanılıyorsa bu prosedüre maximum a posteriori (MAP) puanlama stratejisi denilmektedir. Önsel dağılım, cevaplayıcıların yetenek düzeyleri tahminleri sırasında belirlenebilmektedir. MAP tahminini anlamak için birçok anahtar kavram vardır. Bunların ilki önsel dağılım kavramıdır. Önsel dağılım, araştırmacının sınava girenlerin rastgele dağıldığını varsaydığı hipotetik bir olasılık dağılımıdır. Bu dağılım, tahminin daha etkili olmasını ve tüm cevap örüntüleri için kestirim elde edilmesini sağlar. MTK yetenek düzeyindeki en yaygın önsel dağılımı standart normal dağılımdır. Bu durum, sınava girenlerin ortalaması 0 ve varyansı 1 olan bir dağılımdan örneklendiği anlamını taşımaktadır. İkinci önemli kavram log-olasılık fonksiyonudur. Üçüncü ve son önemli kavram ise

sonsal dağılımdır. Sonsal dağılım; önsel dağılım fonksiyonunun olasılık fonksiyonu ile çarpımından elde edilmektedir. MAP puanlamanın amacı sonsal dağılımı en büyük yapan  $\theta$  değerini bulmaktır. Sonsal dağılımı en büyük yapan  $\theta$  değeri, dağılımın modu olacağı için, bu dağılıma Bayes model tahmini de denilmektedir.

Bireylerin yetenek düzeylerini tahminine öncelikle  $\theta$ 'nın ne olduğunun tahmin edilmesiyle başlanır. Daha sonra kısmi cevap örüntüleri ve madde parametreleri kullanılarak log-olasılık fonksiyonu ve bu log-olasılık fonksiyonunun birinci ve ikinci türevleri hesaplanır. Birinci ve ikinci türevlerin oranını ( $\epsilon$ ) almadan önce önsel dağılım birinci ve ikinci türevleri kapsayacak şekilde yeniden düzenlenir. Bu ayarlardan sonra  $\epsilon$  hesaplanır ve yetenek-düzeyi tahmini buna göre düzenlenir. Test ne kadar kısa olursa önsel dağılımın yetenek düzey dağılımına etkisi o kadar artmaktadır. Uzun testlerde ise önsel dağılım olasılık fonksiyonu tarafından bastırıldığı için, önsel dağılımın yetenek düzeyi tahminine etkisi azdır.

Önsel dağılım sınava girenlerin yetenek seviye tahminlerini önsel dağılımın ortalamasına doğru çeker ve tahmin edilen standart hataları azaltır. Standart hataların azalmasının nedeni sınava girenler hakkında daha fazla bilginin olmasıdır. Spesifik olarak hangi dağılımdan örneklendiği bilinmektedir. MAP kestiriminin güçlü yanlarından biri yetenek düzeyini sınava giren her birey için kestirebilmesidir. Tam puan ve sıfır puan alan bireyler için de mükemmel tahminlerde bulunur. İkinci güçlü yanı ise önsel bilgiyi kapsayarak yetenek düzey tahmininin değerini artırmasıdır. Diğer taraftan, madde sayısının 20'den az olduğu durumlarda MAP puanlar yanlıdır. Bu da MAP tahmininin gerçek parametre değerine eşit olmadığı anlamına gelmektedir. Dahası

önsel dağılım için belirli bir form varsaymak zorunludur. Eğer yanlış önsel dağılım kullanılırsa puanlar ciddi bir şekilde yanlış ve yanlış olmaktadır. Bu etki madde ayırt ediciliklerinin düşük olduğu durumlarda artmaktadır. Çünkü madde ayırt edicilikleri düşük ise daha az psikometrik bilgi var demektir bu durumda önsel dağılım daha önemli hale gelmektedir ve sınava girenlerin yetenek düzey tahminleri üzerindeki etkisi büyümektedir.

### 1.1.2.3.3 Expected A posteriori (EAP)

EAP'ta her bir madde grubu (test) için belirli bir olasılık serisi (ağırlık) belirli  $\theta$  değerleri için hesaplanır. Sonsuz sayıdaki bu  $\theta$  değerlerine Quadrature noktalar  $W\varphi_r$  ve her bir noktadan elde edilen değerlere de ağırlıklar  $W\varphi_r$  adı verilir. Genelde bu ağırlıklar standart normal dağılım yoluyla elde edilir:

$$F(\theta) = \left[ \frac{1}{\sqrt{2.3,14}} \right] \exp(-\theta^2 / 2)$$

Quadrature dağılım (QD) gerçek performansa dayalı doğru yeterlik dağılımını tahmin etmektedir. Genellikle önsel dağılım altta yatan dağılım olarak varsayılmaktadır. QD sadece bütün dağılımı tahmin etmekte, bireysel yetenek kestirimini vermemektedir. Yani QD bireyler üzerinden ( $\theta$ ) dağılımlarının tahminidir. QD bireysel yetenek kestirimi sağlamamakta fakat sonsal dağılımın yetenek kestirimlerinin ortalaması ve standart sapması elde edilebilmektedir. Bu da dikey ölçeklemede gelişimsel ölçeği izleyebilmek için yeterlidir (Tong, 2005).

Noktalar ve ağırlıklar elde edildikten sonra örneğin 40 quadrature nokta için, EAP kestirimi şu şekilde yapılır (Embretson ve Reise, 2000):

$$\theta = \sum_{r=1}^{40} [\varphi_r L(\varphi_r) W(\varphi_r)] / \left\{ \sum_{r=1}^{40} [L(\varphi_r) W(\varphi_r)] \right\}$$

EAP sonsal dağılımının yanlı ortalama yetenek kestirimidir. EAP kestiriminde evrendeki ortalama ve standart sapma değerleri, MO da dahil diğer tüm kestiricilerden daha küçüktür (Zimowski ve diğerleri, 2003). EAP kestirimi, önsel dağılımının kullanılmasından dolayı, 0 ve tam puan alan bireyler üzerinde bile yetenek kestirimine olanak sağlamaktadır, çünkü önsel dağılımın kullanılması,  $\theta$  kestirimlerinin düşük bir aralıkta kalmasını sağlamaktadır (Lord, 1980). EAP ayrıca toplam puanlar üzerinden madde ve yetenek kestirimine olanak sağlamaktadır. EAP kestirimi MAP'tan büyük çıkar çünkü MAP da sonsal dağılımın modunun kullanılmasına karşılık EAP da önsel dağılımın ortalaması kullanılmaktadır. Bu durum da EAP'ın hesaplanmasının daha hızlı olmasını sağlamaktadır.

EAP yetenek popülasyonu üzerinde en küçük hata karelerinin ortalamasına sahiptir ve ortalama doğruluğa göre bundan daha fazla artırılmaz. Fakat bu özellik ancak önsel dağılım doğru olduğu zaman uygulanabilmektedir. EAP tahmini eğer madde sayısı sınırlı ise yanlıdır. Madde sayısı fazla olmadığı zaman yetenek düzey tahminleri ortalamaya doğru geri dönmektedir. Madde sayısının ne büyüklükte olması gerektiği bilinmemektedir. İteratif olmaması, kolay hesaplanması, en küçük hata kareleri tahmini ve her cevap örüntüsü için yetenek düzeyi tahminini sağlaması EAP'ı MO'ya göre daha avantajlı yapan özellikleridir.

#### **1.1.2.3.4 Değerlendirme ölçütleri**

Ölçekleme/eşitleme çalışmalarının son aşaması, sonuçların hangi ölçütler bağlamında değerlendirileceğine karar vermektir (Çetin, 2009). Alan yazında genel olarak kabul edilmiş bir büyüme modeli tanımlanmadığı için dikey ölçeklemenin karakteristik özellikleri kesin bir ölçüt ile karşılaştırılmamaktadır. Dikey ölçeğin özellikleri kendi arasında karşılaştırılabilmektedir. Dikey ölçeklemedeki araştırmalar bir seneden diğer seneye ne kadar büyüme olduğuna ve aynı sınıf düzeyinde sınıflar içindeki dağılımın zamanla nasıl değiştiğine ve sınıf düzeyleri arasındaki dağılımı ayırmaya (separation of grade distribution) odaklanmaktadır. Ortalama, standart sapma, etki büyüklüğü ve yatay uzaklık (Horizontal Distance-HD) gibi istatistikler birçok ölçek puan dağılımını değerlendirmek için kullanılmaktadırlar (Tong ve Kolen, 2007).

##### **1.1.2.3.4.1 Bir sınıf düzeyinden diğer sınıf düzeyine olan büyüme**

Bir sınıfta öğretilen bir bilginin diğer bir sınıfta değişmesi olarak tanımlanır.

Bu büyümeye bakmanın ilk kriteri de ard arda gelen sınıf düzeylerindeki ortalamalar arasındaki farktır. Bunun için de

- Ard arda gelen sınıf düzeylerinde öğrencilerin ortalama olarak ne kadar büyüdüklerine;
- Bir yıldan diğer bir yıla öğrencilerin ortalama olarak ne kadar büyüdüklerine;
- Sınıf düzeyleri arttıkça ortalamalar arasındaki farkın artmakta mı yoksa azalmakta mı olduğuna;
- Öğrencilerin küçük sınıflardaki büyümesinin mi daha fazla yoksa büyük sınıflardaki büyümesinin mi daha fazla olduğuna bakılmalıdır.

Farklı dikey ölçekleme sonucunda elde edilen arda arda gelen sınıf düzeyleri arasındaki ortalama tahminleri ve ortalamalar arasındaki farklar birbirleriyle karşılaştırılır. Ölçekleri karşılaştırmak için kesin bir ölçüt yoktur ancak ortalama tahminlerinin konu alanı, kalibrasyon yöntemleri ve yetenek kestirimlerine bakılmaksızın artması beklenmektedir (Kim, 2007). Bu çalışmada ortalamalar arasındaki fark değerleri mutlak değerlerine göre yorumlanmıştır.

#### **1.1.2.3.4.2 Sınıf düzeyleri arasındaki farklılık (grade to grade variability)**

Sınıf düzeyleri arttıkça her sınıf düzeyinin kendi içerisindeki değişkenliğidir.

Bu değişkenliği inceleyebilmek için şu sorular cevaplanmalıdır:

Sınıf düzeyi arttıkça bu değişkenlik artıyor mu?

Başarılı öğrenciler sınıf düzeyi arttıkça başarısız öğrencilere göre nasıl bir büyüme eğilimindedir?

Sınıf düzeyi arttıkça öğrenciler daha homojen mi daha heterojen mi olma eğilimindedir?

Bu değişkenliğe bakmanın bir kriteri her sınıf düzeyi için elde edilen dikey ölçeklemelerden elde edilen standart sapmalardır. Bu çalışmada her sınıf düzeyi için farklı olarak geliştirilen dikey ölçeklemeler sonucunda hesaplanan standart sapmaların, kalibrasyon metotları ve yetenek kestirimlerine bağlı olup olmadığına bakılmıştır.

Farklı dikey ölçeklemeler sonucunda elde edilen SD tahminlerinin örüntüsü birbirleriyle karşılaştırılmıştır. Ölçekleri karşılaştırabilmek için kesin ölçütler yoktur. Dahası aynı sınıf düzeyleri içindeki standart sapma (örneğin 6. sınıfların arasındaki) farklı sınıf düzeyleri arasında (6., 7. ve 8. sınıflar arasında) artabilir, azalabilir veya aynı kalabilir. Her sınıf düzeyinde (6. sınıflar kendi içinde, 7. sınıflar kendi içinde, 8. sınıflar kendi

içinde) dramatik değişim (10 katı büyük veya küçük) ölçeğin iyi işlemediğini gösterir (Kim,2007).

### 1.1.2.3.4.3 Sınıf dağılımları arasındaki ayırım (separation of grade distrubution)

Ard arda gelen sınıf düzeyleri arasındaki ölçek puan dağılımlarının üst üste gelme derecesidir. Bunu hesaplayan indekslerden biri etki büyüklüğü, diğeri ise yatay uzaklıktır.

#### 1.1.2.3.4.3.1 Etki Büyüklüğü (Effect Size)

Yen (1986) sınıflar arasındaki olası değişkenlik farklılıklarını göz önüne alarak sınıf dağılımları arasındaki ayrışmayı değerlendirmek için etki büyüklüğü indekslerini kullanmayı önermektedir. Eğer  $X_{üst}$  ve  $S^2_{üst}$  üst sınıf için örneklem büyüklüğü ve varyansı;  $X_{alt}$  ve  $S^2_{alt}$  alt sınıflar için örneklem ortalaması ve varyansı ise etki büyüklüğü

$$\text{Etki büyüklüğü} = \frac{\bar{x}_{üst} - \bar{x}_{alt}}{\sqrt{S^2_{üst} + S^2_{alt}}}$$

$\bar{x}_{üst}$  : Üst sınıfın ortalama

$\bar{x}_{alt}$  : Alt sınıfın ortalaması

$S^2_{üst}$  : Üst sınıfın standart sapması

$S^2_{alt}$  : Alt sınıfın standart sapması

Etki büyüklüğü arda arda gelen sınıflar arasındaki ortalama farklarını sınıf düzeyleri içindeki varyansların karekökünü kullanarak standardize etmektedir. Etki büyüklüğü 0-0.20 arasında değer alıyorsa zayıf etki, 0.21-0.50 arasındaki değerler için küçük etki, 0.51-1 orta etki ve 1'den büyük değerler için ise güçlü etki olarak tanımlanmaktadır



(Cohen, Manion ve Morrison, 2007). Bu çalışmada etki büyüklükleri işaretine bakılmaksızın (mutlak değerlerine göre) yorumlanmıştır. Etki büyüklüğü ne kadar büyük ise başarıdaki artış o kadar fazla ve sınıf düzeyleri arasında daha fazla ayırım var demektir.

Etki büyüklüğü farklı ölçekler için ard arda gelen sınıf düzeylerinde hesaplanmıştır. Etki büyüklüğü tahminleri, birbirleriyle kalibrasyon yöntemleri ve yetenek kestirimlerini karşılaştırmak için hesaplanmıştır. Ölçekleri karşılaştırmak için bir ölçüt bulunmamaktadır fakat sınıflar arasındaki herhangi bir dramatik değişim (10 kat daha fazla veya daha az) ölçeğin iyi işlemediğini gösterir. Bu ayrımı gösteren ikinci indeks ard arda gelen yıllar arasındaki yatay uzaklıktır (Holland,2002).

#### **1.1.2.3.4.3.2 Yatay Uzaklık (Horizontal Distance)**

Puan ölçeği üzerindeki iki dağılımın aynı yüzdeler için yüzdeler farkı olarak tanımlanmaktadır (Holland, 2002). Uzaklık ne kadar büyük ise bir düzeyden diğer düzeye o kadar büyüme olmuş demektir. Örneğin 5, 10, 25, 50, 75, 90 ve 95 yüzdeler dilimlerde HD hesaplanıp büyüme miktarı hesaplanabilmektedir. Böylece büyümenin başarılı öğrencilerde mi yoksa başarısız öğrencilerde mi daha fazla olduğu belirlenebilmektedir. Yani birbiriyle ilişkili kalibrasyon yöntemleri ve yetenek kestirimleri için HD'ler birbiriyle karşılaştırılır. Ölçekleri karşılaştırmak için bir ölçüt bulunmamaktadır fakat sınıflar arasındaki herhangi bir dramatik değişim (10 kat daha fazla veya daha az) ölçeğin iyi işlemediğini gösterir.

Holland (2002), yatay uzaklığı hesaplarırken, test puan dağılımlarının Kümülatif Dağılım Fonksiyonlarını (KDF) temel eleman olarak kullanılmıştır. Tanım olarak eğer bazı

grupların veya popülasyon yada alt popülasyonların test puan dağılımının KDF'si  $F$  şu şekilde tanımlanmaktadır;

$$F(x) = \text{Test puanları } x \text{ 'e eşit veya } x \text{ 'ten küçük olan grubun oranı}$$

Hesaplmalarda ondalık sayılarla uğraşmamak için grup oranı yerine yüzde kullanılmıştır. Örneğin %100 ile 1.00 aynı şeyi ifade etmektedir. Yatay aralık hesabı puan dağılım yüzdesinin tanımlanmasını gerektirmektedir. Eğer  $p$  0 ile 1 arasında bir oran ise  $F$ 'nin yüzde-  $p$ 'si  $X(p)$  şu şekilde tanımlanmaktadır;

$$p = G[X(p)] \text{ veya } Y(p) = G^{-1}(p)$$

$p$  yüzdeyi gösterdiğinde  $X(p)$ 'de  $F$ 'nin  $p$ . yüzdesi olarak adlandırılmaktadır.  $F^{-1}$ 'de  $F$  fonksiyonunun ters fonksiyonunu göstermektedir. Başka bir grubun kümülatif dağılım fonksiyonu da  $G$  olsun ve  $G$ 'nin yüzdesi  $Y(p)$  olarak gösterilsin. Yukarıdaki eşitlikte bunları gösterirsek eşitlik;

$$p = G[Y(p)] \text{ veya } Y(p) = G^{-1}(p)$$

şeklinde gösterilir. Bu iki gurup arasındaki fark

$$D(p) = Y(p) - X(p)$$

olarak gösterilir ve burada  $D$  ile gösterilen  $F$  ile  $G$  arasındaki yatay uzaklıktır ve bu iki dağılımın  $p$ . yüzdesi arasındaki farktır.

Ortalama puanları farkı ile iki puan dağılımı arasındaki farkı hesaplamak her  $p$ 'ye eşit ağırlık verilerek olası tüm değerlerinin ortalaması ile  $D(p)$  değerlerinin özetlemekle

hemen hemen aynıdır. Sonuç olarak ortalamalar arasındaki fark ile iki dağılım arasındaki farkı ölçmek iki KDF arasındaki ortalama yatay uzaklık ölçmenin bir versiyonudur. Bu çalışmada yatay uzaklıklar işaretlerine bakılmaksızın mutlak değerlerine göre yorumlanmıştır.

### **1.1.3 Dikey ölçeklemedeki tehditler**

Dikey ölçeklemenin doğruluğunu olumsuz olarak etkileyebilecek faktörler vardır.

Dikey ölçeklemenin geçerliğini etkileyen faktörlerden bazıları şunlardır:

#### **1.1.3.1 Ölçek duyarlılık farkları (Score precision differences)**

Dikey ölçeklemede testler eşitlenirken, eğer testlerin duyarlılıkları eşit değilse eşitlenememe durumu doğar. Puan eşitleme veya kalibrasyon bu farklılıkları telafi edememektedir. Dikey ölçeklemede birçok eşitleme modelinde testlerin duyarlılıklarının farklı olması yeniden ölçeklemeden önce ölçeklerin karşılaştırılabilmesine zıt etki yapabilmektedir. Bu istenmeyen ciddi bir problemdir.

#### **1.1.3.2 Çok boyutluluk (multidimensionality)**

Maddelerin içerik olarak ait oldukları alanlara göre çok boyutlu olması ölçeklemenin doğru olmamasına neden olmaktadır. Bu durum MTK ile ölçekleme yapıldığı durumlarda ortaya çıkmaktadır. Thurstone 1938 modelinde ise bu durum sorun yaratmamaktadır.

### **1.1.3.3 Yapı farklılıkları (Construct differences )**

Dikey ölçeklemenin temel sayıtlısı ölçeklenen testin aynı niteliği veya yapıyı ölçmesidir. Eğer bu sayıtlı sağlanmıyorsa, dikey ölçekleme sonucunda elde edilen puanlarının yorumlanabilirliği azalır. Bu da farklı düzeylerden elde edilen benzer ölçek puanlarının aynı anlam taşımaması demektir. Bu durum ise farklı sınıf düzeylerinde özellikle ard arda gelmeyen sınıflarda eğitim programlarının farklı öğrenme alanlarını konu alması durumunda sorun olmaktadır. Eğer yapı geçerliliğinden kaynaklanan bir çok boyutluluk varsa bu ikisi farklı şeylerdir. Örneğin ardışık sınıflardaki öğrenme alanları çok boyutlu olabilir; fakat ölçülen yapı bakımından farklılık göstermeyebilir. Diğer taraftan iki sınıf için hazırlanan testin yapılarının düzeyi farklı olabilir; ancak her sınıf düzeyinin öğrenme alanı tek boyutlu olabilir.

### **1.1.3.4 Bilmediği içeriğe ait soruyla karşılaşmak**

Dikey ölçekleme sürecinde bazı öğrenciler sınıf düzeylerine ait olmayan sorular çözmek zorunda kalabilirler. Üst düzeyden soru sorulması durumunda bu sorular alt sınıftakilerin bilgi ve seviyesinin üstünde fakat üst sınıftakilerin bilgi ve seviyesine uygun olacaktır ve bu durumda ölçekleme etkilenecektir.

### **1.1.3.5 Pratik ve yorgunluk etkisi**

Bazı veri toplama desenleri her öğrencinin iki test almasını gerektirmektedir. Bu durumda, pratik etkisi ikinci alınan testteki puanların şişirilmesine ya da yorgunluk etkisi puanların azalmasına neden olabilmektedir. Bu etkiler ortaya çıkar ve kontrol edilmezse ölçeklemenin sonucunu etkileyebilir.

## 1.2 PROBLEM DURUMU

Öğrenme süreci boyunca öğrencilerin kazandığı özellikler çeşitli yöntemlerle ölçülmekte ve bu ölçüm sonuçlarına dayanarak değerlendirmeler yapılmaktadır. Özellikle No Child Left Behind Act of 2001 (NCLB; Public Law 107-110) gibi bazı anasınıfından 12. sınıfa kadar olan uzun ölçekli değerlendirme programlarının uygulamasından sonra test geliştiricileri ve uygulayıcıları öğrenci gelişimini izleyen programlarla ilgilenmeye başlamışlardır (Tong ve Kolen, 2010). Dolayısıyla öğrencilerin sınıf düzeyleri arttıkça başarılarındaki gelişimlerinin ortaya konulmasına olanak sağlayan dikey ölçekleme konusu son yıllarda daha fazla önem kazanmıştır ve bu konuda daha fazla çalışma yapılmaya başlanmıştır.

Herhangi bir konu alanında öğrencilerin bireysel olarak gelişimlerinin belirlenmesi önemlidir. Öğrencilerin başarısında hangi sınıf düzeyinde daha fazla artış olduğu örneğin üçüncü sınıftan dördüncü sınıfa geçerken mi yoksa yedinci sınıftan sekizinci sınıfa geçerken mi daha fazla gelişim gösterdiği dikey ölçekleme ile belirlenebilmektedir. Buna ek olarak dikey ölçeklemede çok boyutlu yaklaşımın kullanımı spesifik olarak hangi öğrencinin hangi konularda ilerleyip hangi konularda ilerleme göstermediğinin belirlenmesine olanak sağlamaktadır. Öğrencilerin başarılarındaki artma testlerden elde edilen ölçek puanlarının farklı sınıf düzeyleri arasında karşılaştırılmasıyla belirlenebilmektedir. Bu karşılaştırmanın yapılabilmesi zorluk düzeyleri farklı olan testlere farklı sınıf seviyesindeki öğrencilerin verdikleri cevaplardan elde edilen puanların aynı ölçeğe yerleştirilmesini gerektirmektedir. Bu süreç dikey ölçeklemedir ve karmaşık bir süreçtir. Ölçekleme süreci boyunca kullanılacak ölçekleme deseni, ölçekleme yöntemi gibi birçok kararların verilmesi

gerekmektedir. Farklı kararlar farklı ölçeklerin ortaya çıkmasına neden olmaktadır. Alan yazında hangi yöntemin ya da yöntemlerin öğrencilerin başarılarındaki artışı en iyi ve doğru olarak ortaya koyduğu konusunda bir uzlaşma yoktur. Buna rağmen dikey ölçekleme birçok test geliştiricisi tarafından kullanılmaktadır. Ancak her test geliştirici geliştirdiği ölçek için dikey ölçekleme süreçlerini kendisi belirlemektedir (Tong ve Kolen, 2007).

Öğrencilerin eğitim hayatları boyunca başarılarındaki ilerlemeyi izleyebilecek gelişimsel ölçek puanı (developmental score scale) sağlaması açısından dikey ölçekleme önemli ve gerekli bir süreçtir. Bu nedenle dikey ölçeklemede kullanılan yöntemlerin karşılaştırılması, bu sürecin daha az hata içerecek biçimde yapılmasını sağlayacaktır. Bu çalışmada MTK'ya dayalı dikey ölçekleme yöntemleri değerlendirme ölçütlerine dayalı olarak karşılaştırılmıştır.

### **1.3 ARAŞTIRMANIN AMACI VE ÖNEMİ**

Bu araştırmanın amacı Madde Tepki Kuramının (MTK)'nın tek boyutluluk sayıtlısı sağlandığında MTK'nın ayrı (seperate) ve eş zamanlı (concurrent) kalibrasyon yöntemlerinin, ve MTK yeterlik tahminlerinin (MO, EAP ve MAP) kombinasyonlarıyla oluşturulan dikey ölçekleme sonuçlarının karşılaştırılması ve öğrenci başarısındaki gelişimin tahmin edilmesine olan etkisinin araştırılmasıdır.

Alan yazında yapılan çalışmalar incelendiğinde konuyla ilgili çalışmaların çoğunluğunun simülasyon verisiyle veya simülasyon verisiyle desteklenen gerçek veriyle yapıldığı görülmüştür. Simülasyon verisi gerçek veride olması mümkün

olmayan mükemmel bir uyum sağlar. Bu arařtırmada ise gerek veri setinden elde edilen veriler kullanılmıřtır. Gerek veri setiyle yapılan alıřma sayısı daha azdır yapılan bu alıřmanın ilgili alan yazına da katkısı olacađı dūřunılmektedir.

Ülkemizde öđrencilerin bařarılarındaki geliřimi ölen bir sistem yoktur. 6., 7. ve 8. sınıflarda yapılan SBS sınavları bu konuda fikir verebilir ancak ortak soru sorulmadıđı için dikey ölekleme yapılamamaktadır. Bu alıřma öđrenci bařarısındaki geliřimin izlenmesi konusunda örnek teřkil edebilir.

Bu alıřmada farklı eđitim seviyelerindeki hedefleri kapsamak ve böylece oluřturulan testlerden elde edilen puanları aynı öleđe yerleřtirip karřılařtırma yapabilmek için testler geliřtirilmiřtir. Oluřturulan bu testler yardımıyla elde edilen gerek veriler kullanılarak parametre tahminleri aynı öleđe yerleřtirildikten sonra analizler yapılmıřtır. Ortalama puanları, standart sapma, etki büyüklüđu ve yatay uzaklık hesaplanmıřtır. Bu indeksler dikey öleđin hangi kalibrasyon yöntemiyle ve hangi yeterlik tahminiyle daha dođru olarak sürdürüldüđünü belirleyen bilgileri sađlamaktadır.

#### **1.4.PROBLEM CÜMLESİ**

Dikey ölekleme sürecinde madde parametreleri ayrı ve eř zamanlı kalibre edildiđinde MO, EAP ve MAP ile kestirilen yetenek parametrelerinden elde edilen deđerlendirme ölütleri nasıldır?

##### **1.4.1 Alt problemler**

1. Ayrı kalibrasyon yöntemiyle elde edilen MO, EAP ve MAP yetenek kestirimlerinin

- a. ard arda gelen sınıf düzeyleri arasındaki ortalama farkları,
- b. standart sapmaları,
- c. etki büyüklükleri,
- d. yatay uzaklıkları nasıldır?

2. Eş zamanlı kalibrasyon yöntemiyle elde edilen MO, EAP ve MAP yetenek kestirimlerinin

- a. ard arda gelen sınıf düzeyleri arasındaki ortalama farkları,
- b. standart sapmaları,
- c. etki büyüklükleri,
- d. yatay uzaklıkları nasıldır?

### **1.5 SAYILTI**

Çalışmaya katılan 6., 7. ve 8. sınıf öğrencileri benzer sosyo kültürel ve ekonomik özelliklere sahiptirler.

### **1.6 SINIRLILIK**

Bu çalışma 2007-2010 yılları arasında uygulanan 6., 7. ve 8. sınıf Seviye Belirleme Sınavlarında (SBS) sorulan sorulardan seçilen 54 soruyla sınırlıdır.

### **1.7 İLGİLİ ARAŞTIRMALAR**

Dikey ölçekleme öğrencilerin sınıf düzeyleri arasında başarılarındaki ilerlemeyi ölçmek için kullanılan ve birçok kararı içeren karmaşık bir süreçtir. Dikey ölçekleme sürecinde verilen bu kararlar öğrenci başarısını yorumlamayı etkileyebilir. Alan yazında 1980'li



yıllarda yapılan dikey ölçekleme ile ilgili çalışmalarda eşit yüzdelliklerle ve gerçek veriler eşitleme gibi konular çalışılırken 1990'lı yıllarda ölçek dönüşüm yöntemlerini içeren çalışmalara rastlanmaktadır. Daha sonra ağırlıklı olarak simülasyon çalışmaları ile dikey ölçeklemeyi etkileyen süreçler daha ayrıntılı olarak çalışılmaya başlanmıştır.

Alan yazında ölçekleme süreci boyunca ölçekleme yöntemi, ortak madde sayısı, ölçekleme deseni, kalibrasyon yöntemi, kullanılan bilgisayar programları gibi ölçeklemeye etki eden faktörleri araştıran çalışmalar yer almaktadır. Alan yazında dikey ölçekleme sürecini ve bu süreç boyunca alınan kararları araştıran çalışmalardan önemli görülenler aşağıdaki paragraflarda sunulmuştur.

Wysel ve Reckase (2011) tarafından gerçek veri seti ve simülasyon verisi kullanılarak yapılan çalışmada test eşitlemede farklı kalibrasyon yöntemlerinin karşılaştırılması amaçlanmıştır. Araştırma sonucunda sabit kalibrasyon ve SL prosedürünün en iyi sonucu verdiği bulunmuştur. Eş zamanlı kalibrasyonda daha fazla veri kullandığı için genellikle daha iyi sonuç vermektedir. Ancak eş zamanlı kalibrasyon ölçülen yapıdan kayma olması yani test formları arasında ölçülen yapının eşit olmaması veya test verisinin çok boyutlu olmasından etkilenmektedir bu nedenle bu kalibrasyon yönteminde yanlılık görülebilmektedir.

Kenyon, Gregor, Li ve Cook (2011) anasınıfından 12. sınıfa kadar geniş bir aralıktaki sınıf düzeylerinden elde ettikleri verileri kullanarak WINSTEP ile eş zamanlı kalibrasyon yöntemiyle elde edilen dikey ölçekleme sonucunda alt sınıflardaki

öğrencilerin büyüme oranının üst sınıflardaki öğrencilerin büyüme oranından daha fazla olduğu sonucuna varmışlardır.

Lee ve Ban (2010) tarafında yapılan çalışmada gerçek veri setine benzer olarak üretilen simülasyon verisi kullanılarak test karakteristik eğrisi ve beklenen gözlem puanlarının dağılımı (the expected observed score distributions) ile bağlama yöntemlerinin karşılaştırılması amaçlanmıştır. Aynı kalibrasyonda Heabera'nın SL'ye kıyasla daha düşük bağlama hatası ürettiğini bulunmuştur. Madde sayısı ve örneklem sayısının küçük olduğu durumlarda yeterlik dönüşüm prosedürü aynı kalibrasyon kadar bazen de daha iyi sonuçlar vermektedir.

Simülasyon verisi kullanarak yapılan çalışmada Kim, Lee, Kim ve Kelley (2009) tarafından a) farklı kalibrasyon yöntemleri (örneğin ayrı ve eş zamanlı kalibrasyon) ve farklı bağlama yöntemleri (ortalama/ortalama, ortalama/sigma ve SL) , b) farklı yetenek kestirimleri ve tahminleri (örneğin, WINSTEP'de JML ve BILOG'da MO ve EAP) son olarak da c) Rasch modelde farklı sınıfların temel sınıf olarak alınmasıyla oluşturulan dikey ölçeklemelerin sonuçları karşılaştırılmıştır. Ayrıca her sınıf düzeyi için farklı ortak madde sayısı ( tüm madde sayısının %25'i veya %50'si) ve her sınıf düzeyi için örneklem büyüklüğü (250, 500 ve 1000 olarak) manipüle etmişlerdir. Araştırma sonucunda hataların örneklem büyüklüğü fazla olduğu zaman veya ortak madde sayısı fazla olduğu zaman küçük olduğu görülmüştür ve örneklem büyüklüğünün 250'den daha fazla ve ortak madde sayısının tüm testteki madde sayısının % 25'ten daha fazla olarak Rasch modelde dikey ölçekleme yapılması önerilmektedir. Bunlara ek olarak bağlama sayısı arttıkça (örneğin bütün sınıfları 6. sınıf ölçeğine yerleştirmek için 6, 4.

sınıf ölçeğine yerleştirmek için ise 10 bağlamaya ihtiyaç vardır) hatalar hızlı bir şekilde artmaktadır. Rasch modelde MS dönüşümünün kullanılması önerilmektedir. Ayrıca MM, SL kullanılan bilgisayar programına bağlı olmaksızın benzerdir. İki kalibrasyon metodunda da sonuçlar hangi programla tahmin edilse edilsin benzerdir. Yetenek kestirimleri için genel olarak test karakteristik eğrisi ve ortalama kriteri her iki kalibrasyon metodunda benzer sonuçlar vermektedir. Standart sapma kriteri ise EAP da en düşük sonucu vermektedir. EAP' MMO ve JMO izlemektedir. JMO veya MMO da MS dönüşümü kullanıldığında kalibrasyon ve yetenek kestirimi kombinasyonlarına göre daha büyük hata üretmektedirler. Sonuç olarak en düşük standart sapma hata tahmini dikkate alınacaksa EAP kullanımı önerilmektedir. Bunun dışında Rasch modelle dikey ölçekleme yapılması için herhangi bir yetenek kestirim yöntemi kullanılabilir.

Yaptıkları araştırmada Kang ve Petersen (2009) üç farklı (eş zamanlı, ayrı ve sabit) kalibrasyon yöntemini karşılaştırmışlardır. Bu karşılaştırma gerçek veriye uygun şekilde simüle edilen veriler kullanılarak yapılmıştır. Bu yöntemler karşılaştırılırken düzeltilmiş yetenek dağılımı, madde karakteristik eğrileri ve test karakteristik eğrileri kullanılmıştır. Simülasyon verisinde örneklem büyüklüğü, yetenek dağılımı ve ortak madde sayısı manipüle edilmiştir. 3 parametrelili model kullanılmış ve temel grubun ortalaması 0 ve standart sapmasının 1 olduğu varsayılmıştır. Örneklem büyüklüğü 500'den 2000'e çıkarıldığı zaman standart hataların ortalaması azalmıştır. Benzer şekilde ortak madde sayısı artırıldığında standart hataların ortalaması azalmıştır. Ancak toplam madde sayısının % 20'si kadar ortak madde sayısının güvenilir bir bağlama için yeterli olabileceğini belirtmektedirler. Ayrı ve eş zamanlı kalibrasyon tamamıyla kabul edilebilir sonuçlar üretmektedir. Sabit kalibrasyonda ise iki farklı süreç kullanılmış ve

bu süreçlerinde ayrı ve eş zamanlı kalibrasyonla karşılaştırılabilecek kadar doğru sonuçlar ürettiği görülmüştür. Ancak sabit kalibrasyonun diğer uygulamalarının yanı sıra sonuçlar verebileceği belirtilmektedir.

Briggs ve Weeks (2009) tarafından yapılan araştırmada 2003 ve 2006 yılları arasında Colorado'da uygulanan testlerden elde edilen iki veri setini (biri 3. ve 6. sınıflar arasında diğeri ise 4. ve 7. sınıflar arasında ) kullanarak MTK modelleri (1PLM/PCM ve 3PLM/GPCM- genelleştirilmiş kısmi puan modeli), bağlama yaklaşımı (ayrı ve eş zamanlı kalibrasyon) ve yetenek kestirimleri ( MO ve EAP) kullanılarak oluşturulan sekiz farklı dikey ölçekleme ile öğrencilerin büyüme örüntüleri belirlenmeye çalışılmıştır. Bu araştırmanın sonucunda EAP ile hesaplanan etki büyüklükleri MO ile hesaplanan etki büyüklüklerine göre daha büyük olarak bulunmuştur. MTK modellerinde etki büyüklükleri benzer sonuçlar vermiştir. Genel olarak büyüme örüntüleri bir yıldan diğeri yıla lineer bir değişim göstermemektedir. 3PLM/GPCM modelde hibrit kalibrasyon kullanılarak yapılan analizlerin sonucunda büyüme doğrultusu azalma şeklindedir. Ayrıca 3PLM/GPCM ve ayrı kalibrasyon kombinasyonu sonucunda diğer kombinasyonlara göre daha fazla büyümenin olduğu belirlenmiştir. Ayrı kalibrasyonda ölçek değişkenliği diğer kalibrasyonlara göre daha büyükken, MO'de EAP'a göre daha büyük ölçek değişkenliği üretmektedir.

Çetin (2009) tarafından yapılan çalışmada gerçek veri seti üzerinden Thurstone ve MTK'ya dayalı dikey ölçeklemede farklı örneklem büyüklüklerinin, sınıf seviyelerine göre ortalama, standart sapma ve etki büyüklüğü üzerine etkilerinin sınanması yapılmıştır. Bunun sonucunda Thurstone ölçeklemede, matematik testinde ortalamaların

belirli bir örüntü göstermezken, standart sapmaların ise tutarlı artma eğiliminde olduğu bulunmuştur. Türkçe testinde ise sınıf seviyesi arttıkça ölçeklenmiş ortalamalar düşerken, standart sapmalar ise artış göstermiştir. Ayrıca ortalama örüntüsünden bağımsız olarak standart sapmaların Thurstone ölçeklemede her iki derste de sınıf seviyesi ile arttığı gözlenmiştir. MTK'ya dayalı ölçekleme sonucunda ise matematik dersinde, sınıf seviyesi arttıkça ortalamalar tutarlı bir artış göstermemiş ancak, standart sapmalar sınıf seviyesi arttıkça artış göstermiştir. Türkçe testine MTK ölçeklemede gerek ortalama gerekse standart sapmalarda düzenli bir örüntü meydana gelmemiştir. Örneklem büyüklüklerine göre ölçek sonuçları incelendiğinde Thurstone yönteminde, 1000 kişilik örneklemin, hatayı minimize ettiği ve örneklem küçüldükçe (250), örnekleme bağlı hatanın arttığı gözlenmiştir. Ayrıca Türkçe testinde de standart sapmalara ait hatanın ortalamaya daha bağlı olduğu görülmüştür. KTK ve MTK uygulamalarında, ortalamaların seyrinden farklı olarak standart sapmaların arttığı belirtilmiştir.

Wang ve Jiao (2009) tarafından yapılan araştırmada Stanford 10. Ulusal Araştırma Enstitüsü tarafından 3. sınıftan 10. sınıfa kadar uygulanan veri setinden yapı ve anasınıfından 12. sınıfa kadar olan sınıf düzeylerinde yapı ve ya faktör yapısı ve sınıf düzeyleri arasında yapısal farklılığın olmadığı sınımlanması gerçek veri seti kullanılarak yapılmıştır. Okuma testinde her sınıf düzeyinde yapıların tek boyutlu ve sınıf düzeyleri arasındaki dikey yapılarında tek boyutlu olduğu ayrıca sınıf düzeyleri arasında yapısal atlama (construct shift) olmadığı ve sınıf düzeyleri arasında ortak olan yapıda da farklılık olmadığı, aynı olduğu belirtilmiştir.

Ito, Sykes ve Yao (2008) MTK ile ölçekleme yöntemleri (ayrı ve eş zamanlı) için deneysel kanıt elde etmek amacıyla bir çalışma yapılmıştır. Çoktan seçmeli maddelerden oluşan matematik ve okuma testleri anasınıfından 9. sınıfa kadar olan sınıflara uygulanmıştır. Bu testlerden elde edilen veriler 3 parametrelî model de 4. sınıf referans sınıf olarak alınarak Markov chain Monte Carlo yöntemiyle ayrı ve eş zamanlı kalibre edilerek madde parametreleri hesaplanmıştır. Ayrı kalibrasyonda ise sınıf düzeyleri arasındaki bağlantı SL test karakteristik eğrisi yöntemiyle kurulmuştur. Yetenek kestirimleri de maksimum olabirlik yöntemiyle yapılmıştır. Cronbach's alfa okuma testleri için 0.76 ile 0.90 arasında; matematik testleri için ise 0.78 ile 0.88 arasında değişmektedir. Madde güçlükleri ve ayırt edicilikleri arasındaki bazı farklılıklara rağmen iki ölçekleme yaklaşımının ölçek puanları arasında yüksek korelasyon vardır. Okuma testinde ölçek ortalamaları her sınıf düzeyi için benzerdir. Matematik testlerinde de çoğu sınıf düzeyinde ölçek ortalamaları benzerdir. Ancak büyük sınıflar ve küçük sınıflar arasında varyanslardaki artış eş zamanlı kalibrasyonda daha belirgindir. Ayırt edicilikler, madde güçlükleri ve yetenek kestirimleri karşılaştırıldığında hangi ölçekleme metodunun kullanıldığı matematik testlerinde, okuma testlerine göre daha önemlidir. Sınıf düzeyleri arasında sınıf düzeyi başlangıç olarak kabul edilen sınıf düzeyinden uzaklaştıkça puanlardaki değişkenlik her iki öğrenme alanında da gözlenmiştir. Puanlardaki bu artış eş zamanlı kalibrasyonda ayrı kalibrasyona göre ve düşük sınıf düzeylerinde büyük sınıflara göre daha fazla bahsedilmektedir. Briggs, Weeks ve Wiley'in (2008) tarafından yapılan çalışmanın amacı büyüme ve değer eklenmiş modellerin hassaslığını dikey ölçekleme ile değerlendirmektir. Bu değerlendirme yapılırken dikey ölçekleme sonucu elde edilen ortalamaları ve standart sapmaları arasındaki fark karşılaştırılmıştır. Çoktan seçmeli

sorular için 3PLM ve 1PML ve yapılandırılmış cevaplı sorular için ise genelleştirilmiş kısmi kredi modeli (GKKM) ile ölçekleme yapılmıştır. Kullanılan modellere göre dikey ölçeklemelerin sonuçları karşılaştırıldığında ortalamalar arasındaki farkla gösterilen ortalama büyüme 3PLM ve GKKM kullanıldığında 1PLM ve GKKM kullanılmasına göre daha büyüktür. Ancak ortalamalar arasındaki farklar etki büyüklüğü olarak standardize edilirse fark daha azdır. Ölçek puan değişkenliği tahmininde EAP kullanıldığı zaman etki büyüklüğü MO kullanıldığı zamankine göre daha büyüktür. 1PLM ve GKKM kombinasyonu kullanıldığı zaman korelasyon EAP veya MO bağlı olmaksızın güçlü ve negatiftir. 3PLM ve GKKM kombinasyonu kullanıldığında ise EAP seçildiğinde korelasyon dramatik bir şekilde düşmektedir. Hangi kalibrasyon yaklaşımı ve hangi MTK yöntemi seçilirse seçilsin normatif olarak okul veya öğrenci düzeyindeki rastgele etkiler karşılaştırıldığı zaman benzer sonuçlar gözlenmiştir. Bu yöntemlerin birbiriyle etkileşimi yani birbirleriyle ne şekilde kombine edildiği sonucu etkilemektedir. 3PLM ve GKKM kombinasyonu aynı sınıf düzeyinde okullar daha güvenilir bir şekilde belirlenmektedir. Benze şekilde EAP ile MO'ya göre öğrenci başarılarına göre okullar daha iyi tanımlanmaktadır.

Tong ve Kolen (2008) tarafından yapılan araştırmada 2005 ve 2006 yıllarında 3. sınıftan 8. sınıf düzeyine kadar uygulanan matematik ve İngilizce testleri kullanılarak 2006 testlerini aynı sınıf düzeylerinde 2005 testlerine yatay olarak eşitlenip sonradan sonuçlar temel yıl olan 2005 dikey ölçeğine yerleştirilmiştir. Ayrıca buna ek olarak aynı verilerle 2006 verisi kullanılarak dikey ölçekleme ve yatay eşitleme ile 2006 verileri temel düzey olan 2005 düzeyine eşitlenmiştir. İngilizce testinde genel olarak sınıf düzeyi arttıkça ortalamalarda artmıştır ancak 6. sınıfın ortalaması 7. sınıfın ortalamasından daha büyük

olarak bulunmuştur. Her iki model içinde benzer sonuçlar bulunmuştur. Standart sapmalar ise her iki yıl içinde yükselip alçalıyor ve düzleşiyor. Etki büyüklüklerinde ise sınıf düzeyi arttıkça azalma gözlenmiştir. 2005 verisinde 6. sınıf ile 7. Sınıf arasında ise negatif büyüme görülmüştür. Matematik testinde de benzer şekilde sınıf düzeyi arttıkça ortalamalar artmıştır ve bu durum büyük sınıflarda başarının daha fazla olduğunu göstermektedir. Standart sapmalarda artıp azalma eğilimindedir. 3. ve 4. sınıflar arasında azalmakta sonra ise düzleşmektedir. Böyle bir örüntü puan dağılımının yayılma eğiliminde olduğunu ve sınıf düzeyleri arasında sabit kaldığını göstermektedir. Bu gözlem büyüme oranının başarılı ve başarısız öğrenciler arasında sınıf seviyesi arttıkça benzer olduğu şeklinde yorumlanmıştır. Etki büyüklüklerinde İngilizce testi ile benzer şekilde azalma eğilimindedir ancak İngilizce testine kıyasla bu azalma daha az belirgindir.

Beard (2008) tarafından yapılan araştırmada adaptive çok kademeli test yapısını madde tepki kuramı ölçekleme teknikleriyle kullanmanın dikey ölçeklemeye etkisi simülasyon verisi kullanılarak araştırılmıştır. Ortalamalar incelendiği zaman EAP ve MO birbirine yakın sonuçlar vermekteyken varyanslarda ise MO kestirimi EAP'a göre büyük değerler vermektedir. Ayrıca EAP tahmini MO ile kıyasla daha az hata üretmektedir. MO'da hatalar sistematik olarak artmaktadır.

Burg (2008) tarafından yapılan araştırmada matematiksel içerik standartları farklı yapılar gösterdiği için tek boyutludur sayılısı araştırılmıştır. Tek boyutluluk sayılısının sağlanmamasının parametre tahmini, dikey ölçekleme ve geçerlik kanıtının toplanmasını içeren test geliştirme süreci üzerinde önemli etkileri vardır. Test maddeleri



ve öğrenci performansları belirli boyutsal yapıyı varsayan Madde Tepki Kuramı veya Çok Boyutlu Madde Tepki Kuramı matematiksel modellerle analiz edilmektedirler. Bu çalışmanın sonucunda 3. ve 8. sınıflar arasında geliştirilen matematik testlerinde karmaşık ve bazı çok boyutluluklar olsa da boyutluluğun içerik için geliştirilen standartlara (NCTM standartları) veya testlerin farklı sınıf düzeylerinde soru içermesiyle ilişkili olmadığı bulunmuştur. Ayrıca veri yapısının karmaşık olmasıyla birlikte bilinen matematik becerilerinin üst üste binmesiyle matematik testlerinin temelde tek boyutlu bir yapı gösterdiği sonucuna ulaşılmıştır.

Kim (2007) tarafından yapılan çalışmada farklı kalibrasyon yöntemleri (ayrı, eş zamanlı ve sabit kalibrasyon) ve yetenek kestirim yöntemlerinin (Örüntü puanlamaya dayalı MO, örüntü puanlamaya dayalı EAP, toplam puanlamaya dayalı pseudo-MO ve toplam puanlamaya dayalı Pseudo EAP ve QD) eşit olmayan gruplarda ortak madde deseniyle geliştirilen dikey ölçeklerin özelliklerini nasıl etkilediği araştırılmıştır. Çalışmada geniş ölçekli kelime, okuma, matematik ve fen testlerinden elde edilen veriler kullanılmıştır. Çalışmanın sonucunda eş zamanlı kalibrasyonda ayrı ve sabit kalibrasyona kıyasla başarıdaki artışın daha az olduğu ve sınıf düzeyi arttıkça bu artış miktarının azaldığı belirlenmiştir. Okuma ve fen testleri kalibrasyon yöntemlerinden daha çok etkilenmektedir. Ayrı ve sabit kalibrasyonda standart sapmalar sınıf düzeyi arttıkça azalmaktadır, eş zamanlı kalibrasyonda da azalma gözlenmekte ancak bu azalma miktarı daha azdır. Kalibrasyon yöntemlerinde sınıf düzeylerinin farklılığı MO veya pseudo-MO kullanıldığı zaman EAP, pseudo-EAP ve QD'ye göre daha da artmaktadır. Yatay uzaklıklar ve ortalama yatay uzaklıklarda benzer sonuçlar vermektedir. Okuma ve fen testlerinde ayrı ve sabit kalibrasyonda eş zamanlı kalibrasyona kıyasla sınıf

düzeyleri arasındaki farkın daha fazla azaldığı yani başarısız öğrencilerin daha hızlı büyüdüğü belirlenmiştir. Yetenek kestirim yöntemleri sınıf düzeyleri arasındaki büyüme örüntüsünü çok fazla etkilememektedir. Sadece sabit kalibrasyonda MO 3 ve 4. sınıflar arasındaki büyümeyi diğer yöntemlere kıyasla daha fazla olarak tahmin etmektedir. Yetenek kestirim yöntemleri standart sapmaları etkilemekte ancak sınıf düzeyi attıkça standart sapmalar azalmakta, azalma miktarı yetenek kestirim yönteminden etkilenmemektedir. Her konu alanında standart sapmalar için genel olarak  $\text{pseudo-MO} \geq \text{MO} > \text{QD} > \text{EAP} \geq \text{pseudo-EAP}$  şeklinde bir örüntünün olduğu belirlenmiştir. Yetenek kestirim yöntemlerinde etki büyüklükleri için genel olarak  $\text{pseudo-EAP} \geq \text{EAP} > \text{QD} > \text{MO} \geq \text{pseudo-MO}$  şeklinde bir örüntü elde edilmiştir. Ortalama yatay uzaklık ve yatay uzaklıklar incelendiği zaman ise sadece okuma ve fen testlerinde sabit kalibrasyonda MO 3 ve 4. sınıflar arasında daha büyük farklılıklar olduğu görülmektedir. Diğer durumlarda yetenek kestirim yöntemleri benzer sonuçlar vermektedir.

Lee ve Ban (2007) tarafından yapılan ve simülasyon verisinin kullanıldığı çalışmada eş zamanlı kalibrasyon ayrı kalibrasyona ve sabit kalibrasyon yöntemleri karşılaştırılmış ve eş zamanlı kalibrasyonun örneklem sayısı ve madde sayısının büyük olduğu durumlarda daha kötü sonuçlar verdiği bulunmuştur. Ayrı ve sabit kalibrasyonun ise benzer sonuçlar verdiği görülmüştür. Madde sayısı azaltıldığında ise ayrı ve eş zamanlı kalibrasyon farklı sonuçlar vermektedir. Sabit kalibrasyon sürecinde EAP eğer rasgele grup seçimi yapıldıysa MO'ya göre daha iyi sonuç vermektedir. Genel olarak sayıtlılar sağlandığında MO ayrı kalibrasyonda daha doğru ve daha yaygın kullanılan sonuçlar vermektedir. Bu çalışma ile farklı durumlarda farklı sonuçlar elde edildiği için herhangi bir yöntemin kullanılması önerilmemektedir.

Meng (2007) tarafından iki kategorili ve çok kategorili doğru-yanlış ve açık uçlu sorulardan oluşan karma ve çok düzeyli (multilevel) simülasyon verisi kullanarak dört farklı kalibrasyon yöntemini (ayrı, ikili eş zamanlı kısmi eş zamanlı ve eş zamanlı) karşılaştırmak üzere bir çalışma yapılmıştır. İkili verilerde üç parametrelili model kullanılırken açık uçlu sorular için genelleştirilmiş kısmi puanlar modeli kullanılmıştır. Her dört kalibrasyon yöntemi de farklı örneklem büyüklüklerinde benzer sonuçlar vermişlerdir. Ortalama ve etki büyüklüklerinde en iyi tahmin ikili eş zamanlı kalibrasyon ile yapılan tahmindir. Ancak bu yöntem standart sapma tahminlerinde sorun yaratmaktadır. Parametre tahminlerinde de güvenilir sonuçlar vermemektedir. İkinci en iyi yöntem olarak yarı eş zamanlı kalibrasyon bulunurken SL yönteminin kullanıldığı ayrı kalibrasyon ise en kötü sonucu vermektedir.

Ayrı kalibrasyonda karakteristik eğri yöntemiyle Heabera ve SL yöntemlerini karşılaştırdıkları çalışmalarında Kim ve Kolen (2007) verinin dağılımını ve simetrisini göz önüne alarak ürettikleri simülasyon verisini kullanmışlardır. Ölçüt fonksiyonları simetri ilişkili (ileri, geri ve simetrik) ve dağılımsal ilişkili (dağılımsal ilişkili ve eşit oranda ağırlıklandırılmış) şemaların 6 olası kombinasyonundan oluşmaktadır. Çalışmanın sonucunda hatayı azaltan optimal bir kombinasyon bulunamamıştır

Gerçek ve simülasyon verisi kullanarak yaptıkları çalışmalarında Tong ve Kolen (2007) tarafından Thurstone ve MTK yöntemlerinin karşılaştırılması yapılmıştır. Gerçek veri setinde Thurstone ölçeklemenin çeşitlilik gösterdiği yani sınıf düzeyleri arttıkça büyümenin de arttığı ve başarılı öğrencilerde bu artışın daha fazla olduğu gözlenmiştir.

MTK da ise puanlar artıp azalmaktadır. Yani başarısız öğrencilerin başarı düzeyleri küçük sınıflarda daha fazla artarken büyük sınıflarda daha az artmaktadır. Etki büyüklükleri incelendiği zaman ölçekleme metotlarına göre büyümenin benzer olduğu belirlenmiştir. Simülasyon verisinde ise Thurstone ve MTK yöntemleri benzer dikey ölçekleme sonuçları üretmekte ve MTK ölçekleme marjinal olarak daha doğru tahminde bulunmaktadır. Puanlama yöntemleri ve yeterlik tahminleri karşılaştırıldığı zaman ise; standart sapmalar için :  $MLE > QD > EAP\_PS > EAP\_SS > MAP$  sonucu elde edilirken etki büyüklükleri için  $MLE < QD < EAP\_PS < EAP\_SS < MAP$  şeklinde bir örüntü elde edilmiştir. Gerçek veride çeşitli yetenek kestirim yöntemleri ve puanlama yöntemleri benzer sonuçlar vermektedir. Ortalamalar ve standart sapmalar arasındaki farklılıklar daha az iken etki büyüklükleri için fark MAP ve MO'da önemsizdir. QD, EAP ve MAP benzer sonuçlar verirken MO diğerlerinden biraz daha farklı sonuçlar vermektedir. Simülasyon verisinde ise puanlama yöntemleri ve yetenek kestirimleri için benzer sonuçlar elde edilmiştir. Thurstone yöntemi ile geliştirilen dikey ölçeklerde öğrencilerin başarısının yıllar boyunca arttığını gösterirken MTK ile ölçeklemede öğrencilerin başarılarındaki dağılım sınıf düzeyleri boyunca kararsızlık göstermekte veya azalmaktadır. MO kestirimi EAP ve MAP'a göre daha büyük standart sapma ve daha büyük hata üretmektedir. Örüntü ölçekleme ve toplam puan ölçekleme arasında ise fark gözlenmemiştir.

Yapılan çalışmada Chin, Kim ve Nering (2006) tarafından Monte Carlo simülasyon yöntemiyle 10000 kişi ve 60 maddeden oluşan veri üretilmiş ve bu veri kullanılarak ortalama/sigma yöntemiyle ayrı kalibrasyon ve eş zamanlı kalibrasyon karşılaştırılmıştır. Her iki kalibrasyonda da EAP kullanılmıştır. Her iki yöntem büyüme

örüntüsü veya büyüme farklılıklarını belirlemek için her sınıf düzeyinin sınıf içi varyansı ve yeteneklerin ortalamaları hesaplanmıştır. Tahminin doğruluğunu belirlemek için a, b, c ve teta parametreleri için RMSE hesaplanmıştır. Sonuç olarak her durumda diğerlerinden daha iyi sonuç veren tek bir dizayn yada metot bulunamamıştır. Fakat yine de eş zamanlı kalibrasyon ortak madde sayısından ve ortak maddelerin zorluk derecesinden daha az etkilenmekle beraber bağlanacak form sayısı ve güçlük arttıkça değişken sonuçlar vermektedir. Bir başka deyişle MTK ölçekleme yöntemlerine göre bir sınıf düzeyinden diğer sınıf düzeyine artan yada azalan bir örüntü bulunamamıştır.

Boughton, Lorié ve Yao (2005) tarafından yapılan çalışmada öğrencilerin büyüme örüntüsünün tek boyutlu yapıya mı yoksa çok boyutlu yapıya mı uygun yol izlediği araştırılmıştır. Bu çalışmada Bayesian çok gruplu çok boyutlu MTK modeli dikey olarak ölçeklenen başarı testinin boyutunu belirlemek için kullanılmıştır ve kalibrasyon yöntemleri gerçek veri seti kullanılarak karşılaştırılmıştır. Matematik başarısındaki artış her bilgi ve yetenek düzeyinde tek düze değildir. Öğrencilerin ölçek puanları bir sınıf düzeyinden diğer sınıf düzeyine genel bir büyüme göstermekte fakat bu büyüme aynı değil ve böyle bir büyümenin olacağına da garantisi yoktur. Yapılan çalışmada farklı yıllardaki verilerde farklı büyüme örüntüleri vardır örneğin 2002 yılındaki veride 8 ve 9. sınıflarda başarıda azalma varken 2003 yılındaki veride sadece 9. sınıfların başarısında azalma gözlenmiştir. 2004 yılındaki veride ise hiçbir sınıf düzeyinde azalma görülmemiştir. 2002 verisinde ise en başarılı sınıf 7. sınıftır. Bir yıldan diğer yıla başarıdaki artışı tek boyutlu dikey ölçekleme ile tahmin ederken çok boyutlu dikey ölçekleme ile tahmin edilemeyebilir. Sınıf düzeyleri arasındaki büyüme karmaşıktır fakat çok boyutlu MTK modelleri böyle karmaşık durumları ayırt etmeye yardım

etmektedir. Ancak test maddelerinin çeşitli boyutları kapsayacak şekilde ve öğrenci büyümesini modelleyebilmek için yazılması gerekmektedir.

Yu ve Popp (2005) ortak maddelerle ve ortak bireylerle test eşitleme çalışması yapmışlardır. Ortak maddelerle yaptıkları eşitlemede tüm testlerde ortak olan bağlantı maddeleri vardır aynı zamanda farklı olan maddelerde vardır. Ortak maddeler kullanılarak iki çeşit eşitleme yapılmıştır. İlkinde tüm maddeler aynı anda eşitlenmiştir. İkincisinde ise farklı maddeler ortak maddelerin önceden kalibre edilmesiyle ayrı ayrı analiz edilmiştir. Ortak bireylerin kullanıldığı eşitlemede ise bireyler farklı testleri cevaplamışlardır. Bu testlerde farklı sorular olsa da aynı kazanımları ölçecek şekilde hazırlanmıştır ancak bu testlerde ortak maddeler yoktur. Bu analizler BILOG-MG ve WINSTEP programları kullanılarak yapılmıştır. Böylece ortak madde deseniyle ortak bireylerin kullanıldığı desenler iki programda da analiz edilmiştir. Bu çalışmaların sonucunda araştırmacılar tek bir en iyi eşitleme modeli olmadığı sonucuna varmışlardır.

Pomplun, Omar ve Custer (2004) tarafından yapılan araştırmada BILOG-MG ve WINSTEPS programlarında Rasch model için eş zamanlı kalibrasyonla elde edilen dikey ölçekleme sonuçları karşılaştırılmıştır. 2. sınıftan 8. sınıfa kadar olan matematik testleri için madde ve yetenek parametreleri gerçek ve simülasyon verileri kullanılarak ölçeklenmiştir. BILOG-MG ve WINSTEPS programlarından elde edilen sonuçlar tahmin edilen madde ve yetenek parametreleri arasındaki farklara, korelasyonlara ve gerçek ve tahmin edilen parametreler arasındaki farka göre karşılaştırılmıştır. Genel olarak WINSTEPS madde ve bireylerin ortalamalarını daha doğru olarak tahmin ederken BILOG-MG ise standart sapmaları daha doğru olarak tahmin etmektedir. Bu

farklılık her iki programında farklı tahmin prosedürleri kullanmalarından kaynaklanıyor olabilir. WINSTEPS de BILOG-MG'nin ikinci aşamasına kıyasla yetenekler arasındaki farklar daha büyüktür ve dolayısıyla bu durum bir sınıf düzeyinden diğer sınıf düzeyine daha fazla büyüme olduğu sonucunu vermektedir. Bunun nedeni de WINSTEPS de ortalamaların daha fazla yayılım göstermesidir. Bu durum özellikle en düşük ve en yüksek seviyelerde daha belirgindir. Ancak büyüme örüntüsü sınıf düzeyi arttıkça azalmaktadır ve bu bütün tahminlerde böyle bulunmuştur.

Custer, Omar ve Pomplun (2004) tarafından Rasch modelle BILOG-MG ve WINSTEPS kullanılarak geliştirilen dikey ölçeklemelerin sonuçları karşılaştırıldığı bir araştırma yapılmıştır. Madde ve yetenek parametreleri anasınıfında 11. sınıfa kadar olan sınıf düzeyleri için gerçek verilere uygun şekilde geliştirilen kelime testi hem normal hem de çarpık dağılıma uygun olarak iki farklı şekilde üretilmiştir. BILOG-MG ve WINSTEPS programlarında default ve convergence olan ayarlar kullanılmıştır. Tahin edilen madde ve yetenek parametreleri arasındaki farklar ve korelasyonlar ile sonuçlar karşılaştırılmıştır. İki program arasındaki fark; BILOG-MG'de marjinal maksimum olabilirlik tahminini grup seçeneğiyle kullanılmasına karşılık WINSTEPS'de grup seçeneği olmadan joint maksimum olabilirlik kullanılmasıdır. Bu çalışmada hangisinin daha doğru sonuçlar ürettiği araştırılmış ve bu karşılaştırmayı yaparken de her iki programın kendilerinde var olan (default) ayarları kullanılmıştır. Default ayarlarda BILOG-MG WINSTEPS'e göre madde ve yetenek parametrelerini daha doğru olarak tahmin ederken, yakınsaklık değerleri (0.005; 0.003 ve 0.0001 olarak) azaltıldığı zaman her iki programda daha doğru ve birbirine benzer madde ve yetenek parametreleri üretmektedir. Verinin normal yada çarpık dağılmış olması sorun yaratmamaktadır.

Karkee, Lewis, Hoskens, Yao ve Haug (2003) tarafından yapılan çalışmada ortak madde deseniyle ve ard arda gelen sınıf düzeylerinde 2002 Colorada Student Assessment Program (CSAP) verileri kullanılarak ayrı, eş zamanlı ve ikili- eş zamanlı kalibrasyon (ayrı ve eş zamanlı kalibrasyonun karması) yöntemleri karşılaştırılmıştır. Bir sınıf düzeyinden diğer sınıf düzeyine olan büyümenin ölçek puanlarının ortalamasındaki değişimine göre değerlendirildiğinde 5 ve 6. sınıflar arasında daha fazla büyüme 9 ve 10. Sınıflar arasında ise daha az büyüme olduğu gözlemlenmiştir. 6-7-8 ve 9. sınıflarda her üç düzeyde de benzer büyüme örüntüsü gözlenmiştir. Fark büyük olmadığı için yetenek kestirimlerinden birini diğerine tercih edilebilecek bir önceliği yoktur. Sınıf düzeylerinde ayrı ve eş zamanlı kalibrasyon arasında farklılık görülmezken karma yöntemde 5. ve 6. sınıflar arasındaki farkın büyük olduğu görülmüştür. Ayrı kalibrasyonun diğer iki yönteme göre daha etkili olduğunu söylenebilir. Parametre tahminlerinde yetenek tahminleriyle benzer sonuçlar vermektedir. Karma model diğer iki modelde 5 ve 6. sınıf seviyelerinde farklı sonuç vermektedir. Ayrı ve eş zamanlı kalibrasyonun model veri uyumu ve DIF dışında sınıf düzeyleri arasındaki farklar ve büyüme örüntülerinde benzer sonuçlar vermektedir. İki model arasındaki korelasyonda yüksektir. Test puanlarından çıkarılan sonuçların geçerliliğinin temeli olan model veri uyumu sonuçlarına göre ayrı kalibrasyon eş zamanlı kalibrasyona tercih edilebilir.

Hanson ve Béguin (2002) tarafından yapılan çalışmada simülasyon verisi kullanılarak ayrı ve eş zamanlı kalibrasyon ile madde tepki kuramının madde parametreleri ortak madde eşitleme deseniyle elde edilen veriden tahmin edilmiş ve sonuçları karşılaştırılmıştır. Bu karşılaştırmada kullanılan programlar (BILOG-MG ve



MULTILOG), örneklem büyüklüğü (3000 ve 1000), ortak madde sayısı (20 ve 10) ve eşit ve eşit olmayan gruplar (ortalamaları farklı olmayan ve ortalamaları arasında 1 standart sapmalı fark olan) dikkate alınmıştır. Bunlara ek olarak ayrı kalibrasyonda SL, Haebera, MM ve MS parametre ölçekleme metotları kullanılmıştır. Yapılan karşılaştırmada SL metodunun diğer metotlara göre daha az hata ürettiği görülmüştür. Genel olarak BILOG-MG ve MULTILOG'un benzer sonuçlar verdiği görülmüştür. MULTILOG da ortalamanın 1 olduğu durumun dışında eş zamanlı kalibrasyon ayrı kalibrasyona göre daha az hata üretmektedir. Standart hataların ortalaması örneklem büyüdükçe azalmaktadır. 10 ortak maddenin olduğu ve ortalamanın sıfır olduğu durumda varyansın artmasından dolayı 20 ortak maddenin olduğu duruma kıyasla daha büyük hata vermektedir. 10 ortak maddenin olduğu ve ortalamanın 1 olduğu durumda varyans da yanlılık da artmaktadır. Ortalamanın 1 olduğu durumda yanlılığın artması nedeniyle ortalamanın 0 olduğu duruma göre daha büyük hata vermektedir. Genel olarak bu çalışmanın sonucunda ayrı kalibrasyona göre eş zamanlı kalibrasyonun kullanılması önerilse de daha kesin sonuçlar için bu konuda daha fazla çalışma yapılması gerektiği belirtilmektedir.

Becker ve Forsyth (2002) tarafından yapılan araştırmada 9-12. sınıflar arasındaki 4000 öğrenciye uygulanan testlerden elde edilen veriyle Thurstone ve MTK ölçekleme yöntemleri kullanılarak geliştirilen ölçeklerin karşılaştırılması amaçlanmıştır. Standart sapma veya dağılımın %80 orta aralıkları (the ranges of the middle 80% of the distributions) ile gösterilen değişkenlikler benzer sonuçlar vermektedir. Büyüme analizi sonuçları ise ölçekleme yöntemleri seçilen yüzdelerde benzer sonuçlar vermektedir.

Bu sonuçlar başarılı öğrencilerde büyümenin başarısız öğrencilere göre çok daha fazla olduğunu göstermektedir.

Béguin, Hanson ve Glas (2000) tek boyutlu tahmin yöntemleri olan ayrı ve eş zamanlı kalibrasyonların çok boyutlu simülasyon verisi ile sınavasını yapmışlardır. Eşit guruplarda yapılan analizlerde iki boyutlu veriyle elde edilen ortalama standart hatalar tek boyutlu veriyle benzer sonuçlar vermektedir. Eşit olmayan guruplarda ise iki boyutlu veride hatalar çok daha fazla olarak hesaplanmaktadır. Eşit olmayan guruplarda eş zamanlı ve ayrı kalibrasyonda standart hataların ortalaması ve yanlılık iki boyutlu veride tek boyutlu veriye kıyasla çok daha büyüktür. Eşit guruplarda yapılan analizlerde ayrı kalibrasyon da SL kullanılarak elde edilen sonuçlar eş zamanlı kalibrasyona göre daha iyi sonuçlar vermektedir. Eşit olmayan guruplarda ise hata iki boyutlu veride tek boyutlu veriye göre çok büyük olarak bulunmuştur. Bu durumda her iki kalibrasyon yöntemi de tatmin edici sonuçlar vermemektedirler. İki boyutlu eşit olmayan guruplara ait veride ikinci yetenek kestirimi boyutunda varyans ve kovaryans arttıkça hata artmaktadır. Bu sonuç eş zamanlı kalibrasyonda daha belirgin olarak görülmektedir. Verinin çok boyutlu olması eş zamanlı ve ayrı kalibrasyonun performanslarını etkilemektedir.

Williams, Pommerich ve Thissen (1998) çalışmalarında Kuzey Coralina matematik sınıf bitirme testleri ( North Carolina End-of-Grade Mathematics Tests) ile gelişimsel ölçekler oluşturulmuş Thurstone ve MTK teknikleri kullanılarak sınıf düzeyleri arasındaki değişim analiz edilmiştir. MTK ile yapılan ölçeklemede varyanslar arasında kararlı bir şekilde artan bir örüntüye rastlanmamıştır. Thurstone (1925) ölçeklemede ise

sınıf düzeyi arttıkça azalan bir örüntü olduğu görülmüştür. Thurstone -1938 yöntemiyle ise genel olarak varyanslar arasında artan bir örüntü gözlenmiştir.

Kim ve Cohen (1996) tarafından gerçekleştirilen araştırmada simülasyon verisi üzerinde test karakteristik eğrisi yöntemiyle elde edilen lineer dönüşüm sabitlerinin kullanıldığı ayrı kalibrasyon ve eş zamanlı kalibrasyonu karşılaştırılması yapılmıştır. Araştırma sonucunda ortak madde sayısının az olduğu durumlarda ayrı kalibrasyonun tercih edilmesi önerilmektedirler. Parametre üretimi ve üç yöntemle yapılan parametre dönüşümleri sonucunda elde edilen parametre tahminlerinin sonuçları BILOG ve MULTILOG da hesaplanmıştır. Her iki programla ayrı kalibrasyon yöntemi sonucu oluşturulan dikey ölçekleme süreci sonucunda elde edilen standart hataların daha düşük olduğu tespit edilmiştir.

Zeng (1996) yaptığı çalışmada amacı ortak ölçek dönüşüm yöntemini tanımlamaktır. Yöntemin kullanılabilirliğini göstermek için iki örnek verilmiştir. Birinci örnekte farklı madde havuzlarından alınan ve 36 çoktan seçmeli maddeden oluşan ve 11 ortak maddesi olan iki form eşit olmayan iki gruba (1637 kişiden oluşan 1. gruba ve 1637 kişiden oluşan 2. gruba) uygulanmıştır. Veriler BILOG programı ile kalibre edilmiştir. Teta dağılımı -4 ile 4 olarak alınmıştır. Ortak maddeler üzerinden, a ve b parametreleri kullanılarak yapılan dönüşümler yapıldıktan sonra elde edilen sonuçların benzer olduklarını görülmüştür. İkinci örnekte ise 11 ortak maddesi bulunan 1. madde grubu gerçek parametreler olarak kullanılmış ve farklı madde havuzlarından oluşan üç farklı ölçeğe doğrusal olarak dönüştürülmüştür. Bu ortak madde ölçek dönüşümleri sonucunda ortak ölçeğe yerleştirilen parametreler ortak maddeler için üç farklı madde

havuzunda da benzerdir. Bu çalışmanın sonucunda farklı madde havuzları bağımsız olarak kalibre edildiğinde ortak bir ölçek bulmak kullanışlı bir yöntemdir sonucuna ulaşılmıştır.

Lord ve Wingersky (1984) tarafından yapılan araştırmada ilk ve son testin aynı testler olduğu 6 adet SAT ( Scholastic Aptitude Test) kullanılarak MTK ile gözlenen puanları eşitlemede gerçek puan ile eşitleme ve eşit yüzdelliklerle eşitleme karşılaştırılmış ve benzer sonuçlar verdikleri görülmüştür.

Loyd ve Hoover (1980) tarafından yapılan çalışmada Iowa Temel Beceri Testi'nin (ITBS) verileriyle Rasch Modelde dikey ölçeklemenin pratik uygulanmasının yapılması amaçlanmıştır. Eşitlemedeki belirsizlikler bu düzeydeki matematik testlerinin eşitlenmesinin grubun yeteneğinden bağımsız olmadığını göstermektedir. Daha basit düzeydeki testleri alan grubun puanları daha zor yani yetenek düzeyi yüksek gruba eşitlendiği zaman puanlar daha olumlu yani yüksek çıkmaktadır. Zor test alan grubun puanları daha kolay bir başka ifadeyle düşük düzeye eşitlendiği zamanda puanalar benzer şekilde daha yüksek çıkmaktadır. Ard arda olmayan sınıf düzeyleri arasındaki eşitleme sonuçları sınıf düzeyleri arasında yetenek düzeyleri sabit ise benzer sonuçlar vermektedir. Her üç düzeyde de farklı yetenek grupları kullanılırsa eşitleme diğer eşitlemelerle tutarsız sonuçlar vermektedir. Bu çalışmanın amaçlarından biride bir parametrelide farklı kalibrasyon gruplarıyla eşitlemedeki farklılıkları araştırmaktır. Rash model ile eşitleme eşit yüzdelliklerle eşitleme yöntemiyle farklı sonuçlar vermektedir.

## BÖLÜM II

### YÖNTEM

Bu bölümde, araştırma türü, çalışma grubu, araştırma deseni, veri toplama aracının geliştirilmesi ve verilerin toplanmasına yer verilmiştir.

#### 2.1 ARAŞTIRMA TÜRÜ

Farklı MTK kalibrasyon yöntemleri ve MTK yetenek kestirim yöntemlerinin, öğrencilerin sınıf düzeyleri arttıkça başarılarındaki artışın örüntüsünü tahmin etme düzeylerinin belirlenmesi bu araştırmanın amacını oluşturmaktadır. Araştırmada değişkenler üzerinde herhangi bir manipülasyon yapılmadan var olan durumu var olduğu şekliyle açıklamaya çalışıldığı için betimsel araştırmadır (Fraenkel ve Wallen, 2003).

#### 2.2. ÇALIŞMA GURUBU

Bu araştırmanın çalışma grubunu 6., 7. ve 8. sınıf öğrencileri oluşturmaktadır. Çalışmaya Ankara ili Mamak, Etimesgut, Sincan ve Yenimahalle (2 adet) ilçelerinde bulunan ve okul idaresi tarafından testlerin uygulanmasına izin verilen beş farklı ilköğretim okulundan toplam 1504 öğrenci katılmıştır. Testler 6. sınıfta 229 erkek, 245 kız öğrenci ve 24 de cinsiyetini belirtmeyen toplamda 503 (%33) öğrenciye, 7. sınıfta 207 erkek, 242 kız ve 52'de cinsiyetini belirtmeyen toplamda 501 (%33) öğrenciye, 8. sınıfta ise 221 erkek, 175 kız ve cinsiyetini belirtmeyen 103 toplamda 500 (%33) uygulanmıştır.

Bu beş farklı okulda çalışan öğretmenler genel olarak okulların bulunduğu çevrenin sosyo-kültürel ve ekonomik yapısının düşük olduğunu belirtmişlerdir. Ayrıca velilerinde sürece çok fazla dahil olmadığını, ilgisiz olduğunu vurgulamışlardır.

### 2.3. ARAŞTIRMA DESENİ

Bu araştırmada eşitleme desenlerinden, ortak madde deseni kullanılmıştır. Her testte ait olduğu sınıf düzeyine uygun maddeler ve farklı sınıf düzeyindeki testlerde de bulunan ortak maddeler bulunmaktadır. Örneğin 7. sınıf testinde 7. sınıf düzeyindeki maddelerin yanı sıra 6. sınıf testinde de yer alan ortak maddeler bulunmaktadır. Ortak maddelerde gösterilen performans bir sınıf düzeyinden diğer sınıf düzeyine ne kadar gelişme olduğunu belirlemede kullanılmaktadır. Kolen (2004) ortak madde sayısının testteki madde sayısının %20'si kadar olmasının yeterli olduğunu belirtirken Lorie ve Yao (2005) yaptıkları çalışmada ortak madde sayısındaki artışın testteki ölçmenin standart hatasını azalttığını bulmuşlardır. Kim, Lee, Kim ve Kelley (2009) tarafından yapılan araştırmada ortak madde sayısındaki artışın eşitleme hatasını örneklem büyüklüğüne göre daha çok azalttığı bulunmuştur. Bu nedenle Rasch modelde dikey ölçekleme yapılırken ortak madde sayısının tüm madde sayısının %25'inden fazla olarak alınması önerilmektedirler. Bu çalışmada ortak madde sayısı her üç testte de testteki toplam madde sayısının %25'inden fazladır.

Aşağıdaki tabloda ortak ve ortak olmayan maddelerin testlere dağılımlarını göstermektedir. Ancak örneğin 6. sınıf testindeki 1. soru ile 7. sınıf testindeki 1. soru 6. Sınıf düzeyinde ve her üç testte de ortak olan sorular olsa da aynı soru değildir. Bu diğer testler ve sorular içinde geçerlidir.

**Tablo 1 Ortak ve ortak olmayan maddelerin testlere dağılımı**

Testler	Ortak olmayan maddeler	Ortak maddelerin düzeylere göre dağılımı		
		6. sınıf düzeyi	7. sınıf düzeyi	8. sınıf düzeyi
6. sınıf	6,8,9,10,11,12,13,14, 15,16,17,18,19,20,21	1*,2*,3**	4*,5*,7**	-
7. sınıf	9,10,11,12,13,14,15,16 ,17,18,19,20,21,22,23	1*,2*,4**	5**,6***,7*,8*	3***
8. sınıf	7,8,9,10,11,12,13,14, 15,16,17,18,19,20,21	2*,3*	1*,5**,6*	4***

\*: Her üç testte de ortak olan maddeler

\*\* : 6 ve 7. Sınıf testlerinde ortak olan maddeler

\*\*\*: 7 ve 8. Sınıf testlerinde ortak olan maddeler

Tabloda da gösterildiği gibi ortak maddelerin 4'ü her üç sınıf düzeyinde de ortaktır. Bu dört ortak maddeye ek olarak iki madde 6 ve 7. sınıf testlerinde, iki madde de 7. ve 8. sınıf testlerinde ortaktır. Arda arda gelen sınıf düzeylerinde aşamalı olarak ortak olan maddelerden, 1 tane 6. sınıf sorusu ve 1 tane 7. sınıf sorusu 6. ve 7. sınıfta ortak; 1 tane 7. sınıf sorusu ve 1 tane 8. sınıf sorusu 7. ve 8. sınıfta ortak olacak şekilde seçilmiştir. Farklı sınıf düzeyleri için hazırlanan testler bu ortak maddeler kullanılarak birbiriyle bağlanmakta, böylece farklı testlerden elde edilen puanlar aynı ölçüğe yerleştirilmektedir. Altıncı sınıf temel sınıf olarak alındığında 6. ve 7. sınıftaki ortak maddeler bu iki sınıf düzeyindeki bağlantıyı sağlamakta, 7. ve 8. sınıftaki ortak maddeler de 7. ve 8. sınıf düzeyleri arasındaki bağlantıyı sağlamaktadır. Böylece 6. ve 8. sınıf arasındaki bağlantı 7. sınıf üzerinden yapılmaktadır. Bu çalışmada bu şekilde

zincirleme ölçekleme işlemi uygulanmıştır. Ortak madde deseninin eleştirilerinden biri olan küçük sınıflardaki öğrencilerin üst sınıflara ait soru cevaplamak zorunda olmasının dezavantajı veya büyük sınıfların alt sınıflara ait soruları cevaplayıp avantajlı olmaları durumuyla baş edebilmek için bu şekilde karma bir desen oluşturulmuştur. Ayrıca 6. sınıf testinde bulunan 7. sınıf düzeyindeki ortak sorular 6. sınıf öğrencilerinin de cevaplayabilecekleri düzeydedir. Benzer şekilde 7. sınıf testinde olan 8. sınıf düzeyindeki ortak sorularda 7. sınıf öğrencilerinin cevaplayabileceği düzeydedir. Bu sorular seçilirken 2 matematik öğretmeni ve bir akademisyenden görüş alınmıştır. Bu desene ait eş zamanlı kalibrasyona ait veri girişi ek 1’de verilmiştir.

#### **2.4 VERİ TOPLAMA ARACININ HAZIRLANMASI VE VERİLERİN TOPLANMASI**

Araştırmada kullanılan testler, 2007-2010 yılları arasında 6., 7. ve 8. sınıflara uygulanan SBS sınavlarındaki matematik testlerinden ayırt edicilikleri yüksek olan maddelerden seçilerek oluşturulmuştur. Madde güçlükleri açısından ise testlerde zor ve kolay maddeler bulunmaktadır, ancak testin geneli ortalama zorluktaki maddelerden oluşmaktadır. Test maddeleri seçilirken 6., 7. ve 8. sınıf matematik programlarındaki öğrenme alanları dikkate alınmıştır. Bu öğrenme alanları sayılar, cebir, olasılık ve istatistik, geometri ve ölçmedir (MEB, 2009). 6. ve 8. sınıf testinde 21 madde, 7. sınıf testinde ise 23 madde yer almaktadır. Testler ortak madde desenine göre hazırlandığı için 6. ve 8. sınıf testlerindeki 6 madde, 7. sınıf testindeki 8 madde ortak maddelerdir. Her üç testte de ortak maddeler dışında kalan maddeler ait olunan sınıf düzeyine uygun ve sadece bu sınıf düzeyindeki öğrencilerin cevapladıkları maddelerdir. Yani 6. sınıf testinde 6. sınıf SBS sorularından oluşan ve sadece bu sınıf düzeyindeki öğrencilerin



cevapladığı toplam 15 madde ve 6 da ortak madde yer almaktadır. Bu test ek 2’de verilmiştir. 7. sınıf testinde 7. sınıf SBS sorularından oluşturulan ve sadece bu sınıf düzeyindeki öğrencilerin cevapladığı 15 madde ve 8 ortak madde yer almaktadır. 7. Sınıf testi ek 3’de verilmiştir. Benzer şekilde 8. sınıf testinde 8. sınıf SBS sorularından oluşturulan ve sadece bu sınıf düzeyindeki öğrencilerin cevaplayacakları 15 madde ve 6 de ortak madde bulunmaktadır. 8. Sınıf testi ek 4’te verilmiştir. 6., 7. ve 8. sınıf düzeylerinde hazırlanan testlerde ortak olmayan soruların hangi öğrenme alanından kaçar adet olduğu Tablo 2’de verilmiştir.

**Tablo 2 Testlerdeki ortak olmayan maddelerin öğrenme alanlarına göre dağılımları**

Sınıf Düzeyleri	Öğrenme Alanları				
	Sayılar	Cebir	Olasılık ve İstatistik	Geometri	Ölçme
<b>6.Sınıf</b>	5	3	3	2	2
<b>7.Sınıf</b>	5	3	3	2	2
<b>8.Sınıf</b>	5	3	3	2	2

Ortak olmayan maddelere benzer şekilde ortak sorular seçilirken de öğrenme alanları ve alt öğrenme alanları dikkate alınmıştır. Ortak soruların 4’ü 6, 7 ve 8. sınıf testlerinin üçünde de bulunmaktadır. Bu dört ortak soruya ek olarak 1 adet 6. sınıf ve 1 adet 7. sınıf sorusu 6. ve 7. sınıf testlerinde ortak; 1 adet 7. sınıf ve 1 adet 8. sınıf sorusu 7. ve 8. sınıf testlerinde ortaktır.

Ortak soru sayısı ve bu soruların ait olduğu öğrenme alanları Tablo 3’te verilmiştir.

**Tablo 3 Testlerde bulunan ortak madde sayıları ve öğrenme alanları**

<b>Öğrenme Alanı</b>	<b>Sınıf</b>	<b>6. sınıf</b>	<b>7. sınıf</b>	<b>8. sınıf</b>
<b>Düzeı</b>				
Sayılar	6	X	X	X
Cebir	6	X	X	X
Sayılar	7	X	X	X
Cebir	7	X	X	X
Ölçme	6	X	X	-
Ölçme	7	X	X	-
Olasılık	7	-	X	X
Olasılık	8	-	X	X

Yukarıda açıklandığı gibi ortak ve ortak olmayan maddelerden oluşturulan testler 2010–2011 eğitim öğretim yılında, seçilen 5 farklı okuldaki 6., 7. ve 8. sınıf öğrencilerinin tamamına 15 Nisan 2011 ve 25 Mayıs 2011 tarihleri arasında uygulanmıştır. Öğrenciler testleri derslerine giren öğretmenlerinin gözetmenliğinde 1 ders saatlik (40 dakika) sürede cevaplandırmışlardır. Testler uygulanmadan önce araştırmacı öğretmenleri uygulama konusunda bilgilendirmiş ve öğrenciler de öğretmenleri tarafından bilgilendirilip motive edilmiştir.

## **2.5 VERİLERİN ANALİZİ**

Veriler, araştırmanın alt problemlerine göre analiz edilmeden önce klasik test teorisine göre elde edilen test ve madde istatistikleri incelenmiş, daha sonra madde tepki kuramının varsayımlarının sağlanıp sağlanmadığı test edilmiştir.

### **2.5.1 Test ve Madde İstatistikleri**

6., 7. ve 8. sınıf testlerine ait betimsel istatistikler hesaplanıp bulunan sonuçlar Tablo 4’te gösterilmiştir.

**Tablo 4 Sınıflara Göre Puanların Betimsel İstatistikleri**

	6. Sınıf	7. Sınıf	8. Sınıf
Soru Sayısı	21	23	21
Ortak Madde Sayısı	6	8	6
N	503	501	500
Ortalama	10.658	12.840	9.736
Ortalama Güçlük	0.508	0.558	0.464
$S_x$	3.81	4.62	4.44
Çarpıklık	0.21	0.11	0.50
Basıklık	-0.53	-0.70	-0.14
KR-20	0.73	0.79	0.82
$r_{jx}$ Ort	0.50	0.56	0.46
Medyan	10	13	9
Standart Hata	2.04	2.09	1.9

Araştırmaya katılan öğrenci sayıları 6. sınıfta 503, 7. sınıfta 501 ve 8. sınıfta da 500'dür. Her üç testin ham puanlara dayalı ortalama güçlükleri incelendiğinde, 8. sınıfın ortalama güçlüğü 6 ve 7. sınıflara göre daha düşüktür. 7. sınıfın ortalama güçlüğü ise en yüksek değeri aldığı görülmektedir. Ham puanlardan elde edilen ortalamalar incelendiğinde, 6 ve 7. sınıfların ortalamalarının 8. sınıfa göre daha büyük olduğu görülmüştür. En yüksek ortalama 7. sınıf düzeyinde hesaplanmıştır. Standart sapmalar ise 8. sınıflarda diğer sınıflara göre daha düşük iken 7. sınıflarda ise ortalama ile benzer şekilde diğer sınıflara göre daha büyüktür. 6., 7. ve 8. sınıf testlerine ait KR-20 güvenirlik katsayısı incelendiğinde, 6. sınıfa ait testin güvenirliği (.73), 7. sınıfa ait testin güvenirliği (.79) ve 8. sınıfa ait testin güvenirliği de (.82) olarak bulunmuştur. 8. sınıf testine ait güvenirlik 6. ve 7. sınıf testlerine ait olan güvenirliklerden daha yüksektir. Her üç sınıf düzeyinde hazırlanan testlere ait güvenirlik değerlerinin kabul edilebilir düzeyde olduğu söylenebilir. Bunlara ek olarak her üç testin ayırıcılık gücü indekslerinin ortalamaları 6, 7 ve 8. sınıf testleri için sırasıyla, (.50), (.47) ve (.56) olduğu için, her üç testin de ayırıcı sonuçlar verdiği ve testlerin ayırıcılık gücünün yüksek olduğu söylenebilir.

**Tablo 5 Uygulanan Test Maddelerinin Madde Güçlük ve Madde Ayrıcılık Değerleri**

	6. Sınıf		7. Sınıf		8. Sınıf	
	p	r <sub>jx</sub>	p	r <sub>jx</sub>	p	r <sub>jx</sub>
1	0.724*	0.594	0.812*	0.637	0.742*	0.595
2	0.634*	0.696	0.758*	0.701	0.776*	0.640
3	0.348*	0.350	0.483*	0.519	0.546*	0.594
4	0.423*	0.596	0.453*	0.643	0.418*	0.637
5	0.648**	0.509	0.687**	0.396	0.310	0.577
6	0.243**	0.325	0.401**	0.505	0.452	0.595
7	0.419	0.448	0.359***	0.477	0.314***	0.569
8	0.127	0.110	0.679***	0.571	0.606***	0.553
9	0.541	0.690	0.443	0.453	0.388	0.548
10	0.841	0.638	0.531	0.623	0.266	0.565
11	0.565	0.686	0.505	0.387	0.276	0.336
12	0.483	0.545	0.677	0.566	0.672	0.624
13	0.545	0.391	0.246	0.439	0.538	0.577
14	0.370	0.571	0.547	0.614	0.408	0.542
15	0.624	0.341	0.621	0.650	0.620	0.656
16	0.384	0.429	0.525	0.629	0.468	0.626
17	0.573	0.419	0.483	0.567	0.592	0.683
18	0.734	0.645	0.814	0.627	0.304	0.628
19	0.396	0.524	0.567	0.563	0.306	0.509
20	0.396	0.391	0.727	0.687	0.410	0.553
21	0.642	0.520	0.218	0.605	0.324	0.477
22	-	-	0.687	0.443	-	-
23	-	-	0.619	0.550	-	-
KR-20	0.73		0.79		0.82	

\*: Her üç testte de ortak olan maddeler

\*\* : 6 ve 7. Sınıf testlerinde ortak olan maddeler

\*\*\*: 7 ve 8. Sınıf testlerinde ortak olan maddeler

Tablo 5 incelendiğinde, 6. ve 7. sınıfta güvenirliliğin 8. sınıftan daha düşük, fakat kabul edilebilir düzeyde olduğu görülmektedir. Maddelerin ayırıcılık gücü indekslerinin yüksek ve kabul edilebilir düzeydedir.

Ortak maddelerin madde güçlükleri ve ayırıcılık gücü indeksleri incelendiği zaman üç testte de ortak olan 1. ve 2. soruların 6. sınıf seviyesinde olmalarına rağmen üç sınıf düzeyi içinde oldukça zor ancak ayırt edicilikleri de yüksek maddeler oldukları söylenebilir. Her üç testte de ortak olan 3. soru incelendiği zaman bu soru 7. sınıf seviyesinde olmasına rağmen 6. sınıflar için kolay ve ayırt ediciliği düşük bir soru olmakla birlikte 7 ve 8. sınıflar için ortalama zorlukta ve ayırt edici bir madde olduğu söylenebilir. Her üç testte de ortak ve 7. sınıf seviyesinde bir soru olan 4. soru ortalama güçlükte ve ayırt edici bir sorudur. 6 ve 7. sınıf testlerinin ortak sorusu olan 5. soru (8. Sınıf testindeki 5. soru farklı bir sorudur) her iki sınıf düzeyleri için zor ve 6. sınıf için ayırt edici olsa da 7. sınıflar için ayırt ediciliği düşüktür. 6 ve 7. sınıf testlerinin ortak olan bir diğer soru olan 6. soru (8. Sınıf testindeki 6. soru farklı bir sorudur) 6. sınıflar için kolay ve ayırt edici olmayan bir soru olsa da aynı soru 7. sınıflar için nispeten zor ve ayırt edici bir sorudur. 7 ve 8. sınıf testlerinde ortak olan 7. soru (6. sınıf testindeki 7. soru bu sorudan farklı bir sorudur) her iki sınıf düzeyi için kolay ve ayırt edici bir sorudur. Benzer şekilde 7 ve 8. sınıf testlerinde ortak olan 8. soru (6. sınıf testindeki 8. soru bu sorudan farklı bir sorudur) her iki sınıf düzeyi için zor ancak ayırt edici bir sorudur.

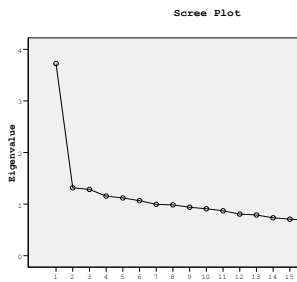
### **2.5.2 MTK Varsayımlarının Test Edilmesi**

Madde tepki kuramına göre ölçekleme yapılmadan önce MTK varsayımları test edilmiştir.

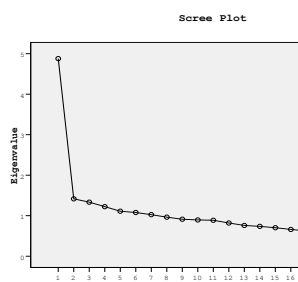
### 2.5.2.1 Tek Boyutluluk

6., 7. ve 8. sınıf verilerinin tek boyutluluğu kontrol edebilmek için temel bileşenler faktör analizi SPSS15 programı kullanılarak yapılmıştır. 6. sınıf verileriyle yapılan analiz sonucunda elde edilen ilk faktör toplam varyansın %17.741'ini açıklamaktadır ve öz değeri 3.495'dir. İkinci faktör ise varyansın %5.758'lik kısmını açıklamakta ve öz değeri 1.307'dir. Özdeğerlerin oranı 3.081 olarak bulunmuştur. Bu sonuç 6. sınıf testinin tek boyutlu olduğuna işaret eden bir kanıt olarak kullanılmıştır. 7. sınıf verilerinin analiz sonucunda elde edilen ilk faktör toplam varyansın %21.207'sini açıklamaktadır ve öz değeri 4.878'dir. İkinci faktör ise varyansın % 6.157'lik kısmını açıklamakta ve öz değeri 1.416'dır. Özdeğerlerin oranı 4.256 olarak bulunmuştur, bu sonuç da 7. sınıf testinin tek boyutlu olduğuna işaret eden bir kanıt olarak kullanılmıştır. 8. sınıf verilerinin tek boyutluluğunu test etmek için yapılan faktör analizi sonucunda elde edilen ilk faktör toplam varyansın %22.596'sını açıklamaktadır ve öz değeri 4.745'tir. İkinci faktör ise varyansın % 7.415'lik kısmını açıklamakta ve öz değeri 1.557'dir. Özdeğerlerin oranı 3.048 olarak bulunmuştur, bu sonuç 8. sınıf testinin tek boyutlu olduğuna işaret eden bir kanıt olarak kullanılmıştır. Ayrıca her üç testte de 1. boyuttan 2.boyuta özdeğerlerdeki keskin düşüş, testlerin tek boyutlu olduğuna işaret eden bir kanıt olarak kullanılmıştır.

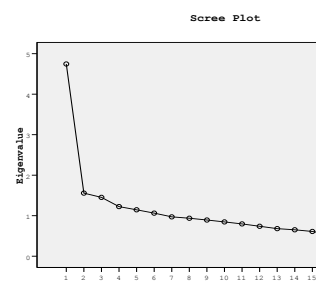
6. sınıf testi



7. sınıf testi



8. sınıf testi



Dikey ölçeklemede farklı sınıf düzeylerinden sorularında testlerde olması tek boyutluluk sayıltısını etkileyip etkilemediğini akıllara getirmektedir. Burg (2008) tarafından yapılan çalışmada bu durum incelenmiş ve bu çalışmanın sonucunda 3. ve 8. sınıflar arasında geliştirilen matematik testlerinde karmaşık ve bazı çok boyutluluklar olsa da boyutluluğun içerik için geliştirilen standartlara (NCTM standartları) veya testlerin farklı sınıf düzeylerinde soru içermesiyle ilişkili olmadığı bulunmuştur. Matematikteki öğrenme alanları doğası gereği birbiriyle ilişkilidir ve matematiksel beceriler diğer konulara veya ard arda gelen konulara dağılmaktadır. Bir diğer ifadeyle birçok konu birbiriyle ilişkili ve bir önceki veya diğer konuların bilinmesini gerektirmektedir. Matematik programları genel olarak farklı öğrenme alanlarından oluşacak şekilde kavramlaştırılmaktadır. Bu durum matematiksel başarıdaki boyutsal yapıyı ortaya koymak için değil ders programlar için prensipler organize etmek ve ders kitaplarını prensip ve standartlardan oluşacak şekilde tasarlanmak ve matematiği bu şekilde öğretmek içindir. Burg (2008)'ün çalışmasında öğrenme alanları ile boyutluluk arasında ilişki bulunamamıştır. Ancak öğrenme alanları ve matematiksel beceriler arasında yüksek bir korelasyon ve ortak yapının mevcuttur olduğunu belirtmiştir. Örneğin tam sayılarda toplama işlemi istatistik ve olasılık öğrenme alanında bir veri setinin ortalamasının hesaplanmasında, geometride verilen şeklin alan ve açı hesaplamalarında, cebirde denklem çözümlerinde gibi bir çok konuyla iç içe ve bir çoğu için gereklidir. Sonuç olarak matematiğin karmaşık ve birbiriyle ilişkili yapısı gereği karmaşık test yapısının olması sürpriz değil beklenen bir durumdur. Bu durum matematikteki öğrenme alanlarının birbiriyle karşılıklı ilişkili olmasından kaynaklanmaktadır ve verinin boyutunun sayısını değil ilişki boyutunun sayısı hakkında bazı şeyler belirtmektedir. Matematiğin karmaşık yapısı ve konuların iç içe geçmiş

olması, birbiriyle ilişkili olması ve matematik becerilerinin üst üste binmesi dolayısıyla matematik başarı testlerinin veri yapısının karmaşık olmakla birlikte temelde tek boyutlu bir yapı gösterdiği sonucuna ulaşmıştır.

Yapılan çalışmada da Burg (2008)'in çalışmasına benzer şekilde bazı karmaşık yapılar gözlemlense ( bazı soruların birden fazla faktöre yüklenmesi) de testlerin temelde tek boyutlu olduğu söylenebilir.

### 2.5.2.2 Yerel Bağımsızlık

Yerel bağımsızlık sorulara verilen cevapların istatistiksel olarak bağımsız olması demektir. Bu nedenle yetenek düzeyi sabit tutulduğunda, maddeler arasındaki korelasyonun sıfıra yaklaşması beklenmektedir. Bu çalışmada testlerin tek boyutlu olması yerel bağımsızlık için bir kanıt olarak kullanılmıştır (Hambleton, 1991; Lord, 1968).

### 2.5.3 Model Veri Uyumu

Ayrı kalibrasyonda her sınıf düzeyi için uyum istatistikleri hesaplanmış ve bulunan sonuçlar Tablo 6.'de gösterilmiştir.

**Tablo 6 Ayrı Kalibrasyona Ait Sınıflara Göre Puanların Uyum İstatistikleri**

	6. Sınıf	7. Sınıf	8. Sınıf
$\chi^2$	212.5	353.3	212.9
p	0.002	0.000	0.000

Uyum istatistiklerine bakıldığında ise 6. Sınıf testinde 2. ve 4. soruların MTK model veri uyumu istatistik değerleri ( $\chi^2$  değerleri  $p < .01$ ) düşük bulunmuştur. Ancak bu iki



soru ortak sorulardır ve 7. ve 8. sınıf testlerinde MTK model veri uyumunu sağladıkları için testlerden çıkarılmayıp dikey eşitleme sürecine dahil edilmiştir. Her üç sınıf düzeyi içinde  $\chi^2$  değerleri verinin 2PLM'ye uygun olduğunu göstermektedir.

Her üç testteki maddeler BILOG-MG (Zimowski, Muraki, Mislevy, ve Bock, 1996) programı kullanılarak ölçeklenmiştir. Maddelerin ölçek değerlerine ait özet bilgiler aşağıdaki tabloda verilmiştir.

**Tablo 7 MTK'nın 2 parametrelili modeline göre hesaplanan madde parametreleri ölçek değerleri özeti**

		Ort.	St. Sap.	Min.	Mak.
6. sınıf	b	0.670	2.176	-1.676	7.294
	a	0.994	0.882	0.268	3.485
7. sınıf	b	-0.177	0.87	-1.6	1.81
	a	1.015	0.382	0.516	1.888
8. sınıf	b	0.333	0.855	-1.122	2.075
	a	1.049	0.277	0.489	1.624

Tablo 7 incelendiği zaman 7. sınıf testinin ortalama güçlüğü'nün negatif olduğu görülmektedir. Her üç test için de ayırt edicilik değerleri incelendiğinde ise birbirine yakın değerler olduğu söylenebilir. Ayrıca altıncı sınıf testinde b değeri için maksimum değer 7.294 olması bu veride uç değerlerin varlığından söz edilebilir. Aynı zamanda model veri uyumunu da sağlamayan bu değerler ortak madde oldukları için tsetten çıkarılmamış ve analize verinin bu haliyle devam edilmiştir.

## 2.6 KULLANILAN YÖNTEMLER

Bu çalışmada iki farklı kalibrasyon yöntemiyle dikey ölçekleme yapılırken iki kategorili MTK modellerine dayanan madde cevaplarının analizleri yapılırken BILOG-MG (Zimowski ve diğerleri, 1996) kullanılmıştır.

### *Eş zamanlı Kalibrasyon*

Eş zamanlı kalibrasyon için bütün sınıf düzeylerindeki veriler bir araya getirilmiştir. Bütün tahminler ortak bir ölçek üzerine bir defa BILOG-MG programının çalıştırılmasıyla yerleştirilmektedir.

Bu çalışmada eş zamanlı kalibrasyonda şu adımlar izlenmiştir (Kim, Lee, Kim, Kelly, 2009);

- 1) Bütün sınıf düzeylerinin verileri bir araya getirilmiştir. Hangi sınıf düzeyinin temel düzey olduğuna bağlı olarak ortak kalibrasyon belirlenmiştir. Örneğin eğer 6. sınıf temel sınıf ise 6. sınıf testi referans, 8 ve 7. sınıf testi hedef testlerdir.
- 2) Bütün sınıf düzeyleri için madde parametreleri temel sınıf ile tahmin edilir. Madde parametre tahminleri ve yeterlik tahminleri, sonsal quadrature noktalar ve ağırlıklar her sınıf düzeyi için hesaplanmıştır. Ortak tahmin hedef testin ortak maddeleriyle referans testin değerini sabitlemek için uygulanmıştır. Quadrature ağırlık ve hedef testin dönüştürülmüş quadrature noktaları ön dağılımı belirlemek için kullanılmıştır. Sonuç olarak hedef testin sonsal quadrature ağırlıkları hesaplanmıştır. Bu ilk eş zamanlı tahmin çalıştırıldığında yetenek tahminleri hesaplanmaz.

- 3) 2. adımda elde edilen sonsal quadrature dağılım ikinci eş zamanlı tahminin önsel dağılımının hesaplanmasında kullanılmıştır. Bu aşamada EAP ile yetenek kestirimi yapılmıştır.
- 4) Her bir eş zamanlı kalibrasyon için madde güçlükleri yeniden ölçeklenmiştir. Madde güçlükleri başlangıç değerlerine sabitlenmek yerine değiştirilmiştir. Eş zamanlı kalibrasyon ikinci defa çalıştırılmadan önce madde güçlük değerlerinin referans sınıf düzeyine sabitlenebilmesi için sonuçlar lineer olarak dönüştürülmüş oldu. Bu dönüşüm ortak maddelerin kalibrasyon öncesi ve sonrası ortalamaları arasındaki farkları kullanılarak gerçekleştirilmiştir. MM dönüşüm sabiti ortak maddelerin ortalamaları farkı olarak hesaplanmıştır. Bu dönüşüm sabitleri ikinci eş zamanlı kalibrasyon yapılmadan önce bütün madde güçlüklerine uygulanmıştır.
- 5) Madde güçlüklerine dayanarak yetenek kestirimleri yapılmıştır.
- 6) Diğer sınıf düzeyleri için 1-5 arası adımlar tekrar edilmiştir.
- 7) Farklı sınıflar temel sınıf olarak alındığında ise 1-6 arası adımlar tekrarlanır.

#### *Ayrı Kalibrasyon*

Ayrı kalibrasyonda ise her sınıf düzeyindeki veriler BILOG-MG kullanılarak ayrı olarak kalibre edilmektedir. Daha sonra her sınıf düzeyinin parametre tahminlerini temel olarak kabul edilen ölçeğe yerleştirmek için ST programı kullanılarak SL dönüşümünde kullanılan eğim ve kesişim değerleri hesaplanmıştır. ST programı C programlama dilinde yazılmış ve MTK ölçek dönüşüm fonksiyonu sabitlerini hesaplayan bir programdır. Dört farklı MTK dönüşüm yöntemi (ortalama/ortalama, ortalama/sigma, Heabera, ve SL) içinde kullanılır (Hanson ve Zeng. Rev. Cui,2004) .

Ayrı kalibrasyonda şu adımlar izlenir (Kim, Lee, Kim, Kelly, 2009; Kim,2007 );

- 1) Her sınıf düzeyi ayrı ayrı kalibre edilir. Madde parametre tahminleri, yetenek kestirimleri (MO, EAP ve MAP) ve sonsal dağılımlar (quadrature noktalar ve ağırlıklar) elde edilmiştir.
- 2) Farklı dönüşüm sabitleri A ve B, arda arda gelen sınıf düzeylerinde MM MLve SL dönüşümlerinden biri seçilerek hesaplanır. Bu çalışmada SL dönüşümü kullanılmıştır. Dönüşüm parametreleri hesaplanırken 1. adımda hesaplanan quadrature noktalar ve ağırlıklar kullanılmıştır.
- 3) Hedef testin madde parametre tahminleri, yetenek kestirimleri (MOK, EAP ve MAP) ve quadrature noktası referans test ölçeğine koymak için dönüştürülmüştür. Quadrature ağırlık için ek bir dönüşüme ihtiyaç yoktur.
- 4) Temel olarak alınan sınıf düzeyine bağlı olarak çoklu bağlantılar geliştirilmiştir. Örneğin eğer temel sınıf 6. sınıf ise 5. sınıfı 6. sınıf ölçeğine dönüştürmek için 5. sınıf ve 6. sınıf arasındaki A ve B değerlerini kullanarak bir dönüşüm yapmak gerekmektedir. Ancak 4. sınıfı 6. sınıf ölçeğine dönüştürmek için iki dönüşüme ihtiyaç vardır. Önce 4. ve 5. sınıf arasındaki A ve B değerleri kullanılarak ve sonra 5. ve 6. sınıflar arasındaki A ve B değerleri kullanılarak yapılır.
- 5) Farklı sınıfların temel sınıf düzeyi olarak alınması durumunda 1-4 arası adımlar tekrarlanır.
- 6) Quadrature noktaların ortalaması, standart sapmalar, etki büyüklüğü ve dikey uzaklıklar 3. adımdaki dönüştürülmüş quadrature noktalar ve ağırlıklar kullanılarak hesaplanmıştır.

BILOG-MG ile yapılan ayrı kalibrasyonda SL dönüşümünü yapmak için kullanılan eğim (slope) ve kesişim (intercept) değerleri aşağıdaki Tablo 8'de verilmiştir.

Tablo 8 SL dönüşümü için Kullanılan A ve B sabitleri

Sınıf Düzeyi	A (Slope)	B (intercept)
6-7	0.470	-0.155
7-8	1.576	0.421

7. sınıfı, referans sınıf olan 6. sınıf ölçeğine dönüştürmek için sadece bir dönüşüm yeterli iken 8. sınıf için iki dönüşüme ihtiyaç vardır. Örneğin 7. sınıfı 6. sınıf ölçeğine dönüştürmek için yetenek kestirimleri  $\theta_{\text{yeni}} = \theta_{\text{eski}} \times 0.470 + (-0.155)$  eşitliğiyle yapılırken 8. sınıfı dönüştürmek için  $\theta_{\text{yeni}} = (\theta_{\text{eski}} \times 1.576 + 0.421) \times 0.470 + (-0.155)$  eşitliği kullanılmıştır. Kesişim değerleri 6 ve 7.sınıf arasında negatif iken 7 ve 8.sınıf arasında pozitifdir.

Analizler için yazılan BILOG MG kodları Ek 5'te verilmiştir.

#### *Değerlendirme Ölçütleri*

Bu çalışmada farklı yöntemlerle elde edilen dikey ölçekleme sonuçlarını değerlendirmek için ortalama farkları, standart sapmalar, etki büyüklükleri ve yatay uzaklıklar hesaplanmıştır.

Sınıf düzeyleri arasındaki büyüme örüntüsünü belirlemek için arda arda gelen sınıf düzeyleri arasındaki ortalama  $\theta$  farkları hesaplanmıştır. Öğrencilerin başarılarındaki değişimi yorumlayabilmek, nedenlerini belirleyebilmek için 7 matematik öğretmeniyle görüşülmüştür. Büyüme örüntüsü sonuçları öğretmenlerle paylaşılmış ve sonuçların nedenleri hakkındaki görüşleri sorulmuştur.

Sınıf içi çeşitliliğin yada homojenliğinin kalibrasyon yöntemleri ve yetenek kestirimlerine göre nasıl bir örüntü izlediğini belirlemek için standart sapmalar

hesaplanmıştır. Standart sapmaların küçük olması grubun daha homojen olduğunun göstergesidir.

Yetenek kestirimlerinin sınıf düzeyleri arasındaki ayrımı yada benzerliği etki büyüklüğü ve yatay uzaklık birimleri ile analiz edilmiştir. Etki büyüklüğü standardize edilmiş ortalama farklarıdır. Etki büyüklüğünün küçük olması ard arda gelen sınıf düzeyleri arasında yetenek düzeyleri arasındaki farkın az olduğunun göstergesidir. Yatay uzaklıklar ise iki dağılım arasında, seçilen belirli yüzdelerdeki noktalardaki fark olarak tanımlanmaktadır. Örneğin 6. sınıf dağılımının % 5'i ile 7. Sınıf dağılımının % 5'i arasındaki fark % 5'lik noktadaki yatay uzaklıktır. Mesafenin büyük olması ardışık sınıf düzeylerinde farkın büyük olduğunun göstergesi olarak kabul edilmektedir. Ortalama yatay uzaklıklar; %1'den % 99' kadar olan noktalarda hesaplanan yatay uzaklıkların toplanıp 99'a bölünmesiyle elde edilmiştir. Alan yazında (Kim,2007; Holland,2002; Karkee ve diğerleri, 2003) yapılan çalışmalarda %5, %10, %25, %50, %75,% %90 ve %95 noktalarda yatay uzaklıklar hesaplandığı için bu çalışmada da belirtilen yüzdelerdeki yatay uzaklıklar hesaplanmıştır.

## BÖLÜM III

### BULGULAR ve YORUM

Bu bölüm araştırma sonucunda elde edilen bulgular ve bu bulguların yorumlanmasından oluşmaktadır. Araştırma sonucunda üç farklı yetenek kestirim yöntemi ile iki farklı kalibrasyon yönteminin kombinasyonu ile elde edilen dikey ölçeklemelerin sonuçları karşılaştırılmıştır. Elde edilen farklı dikey ölçekleme sonuçlarını karşılaştırmak ve değerlendirebilmek için ortalamalar arasındaki farklar, standart sapmalar, etki büyüklükleri ve yatay uzaklıklar kullanılmıştır. Bu bağlamda bu bölüm üç alt başlıkta toplanmıştır. İlk bölümde birinci alt problemde elde edilen bulgulara yer verilirken ikinci bölümde de ikinci alt probleme ait bulgulara yer verilmiştir. Son bölümde kalibrasyon yöntemleri kendi aralarında karşılaştırılmıştır.

#### 3.1 ALT PROBLEM 1'E AİT BULGULAR VE YORUM

**Birinci Alt Problem:** Ayrı kalibrasyon yöntemiyle elde edilen maksimum olabilirlik, EAP ve MAP yetenek kestirimlerinin ard arda gelen sınıf düzeyleri arasındaki,

- a. ortalama farkları ve standart sapmaları,
- b. etki büyüklükleri,
- c. yatay uzaklıkları nasıldır?

Ortalamalar  $\theta$  ölçeğinde 6. sınıf ortalaması 0 ve standart sapması 1 olan referans grup olarak alınarak tüm sınıf düzeyleri için farklı durumlar için yapılan dikey ölçekleme sonuçlarından hesaplanmıştır.

Arda arda gelen sınıf düzeyleri arasındaki ortalama  $\theta$  değerleri, ortalamalar arasındaki fark ve standart sapma değerleri ayrı kalibrasyon yöntemiyle ve MO, EAP ve MAP ile kestirilmiş ve elde edilen sonuçlar Tablo 9’da gösterilmiştir.

Tablo 9 Ayrı Kalibrasyonda Sınıf Düzeylerine Göre  $\theta$  Ortalamaları, Ortalamalar Arası Farkları ve Standart Sapmaları

	Sınıf Düzeyleri	EAP	MAP	MO
Ortalama	6	0.007	-0.020	0.007
	7	-0.204	-0.161	-0.155
	8	0.054	-0.038	0.082
Ortalamalar arası fark	6-7	-0.211	-0.141	-0.148
	7-8	0.258	0.199	0.237
Standart sapma	6	0.893	0.844	1.191
	7	0.251	0.408	0.493
	8	0.702	0.652	0.930

Her üç yetenek kestirimi için Tablo 9 incelendiğinde ortalama  $\theta$  değerlerinin 6. sınıf düzeyinden 7. sınıf düzeyine bir düşüş göstermekte, 8. sınıf düzeyinde ise tekrar artmakta olduğu görülmektedir. Ancak alan yazında yapılan çalışmalarda ortalamaların sınıf düzeyleri arttıkça genellikle arttığı bulunmuştur (Kolen ve Tong, 2010; Kim, 2007).

6. sınıf düzeyinde en düşük ortalama  $\theta$  değeri MAP ile hesaplanırken EAP ve MO aynı sonuçları vermektedirler. 7. sınıf düzeyinde ise en düşük ortalama  $\theta$  değeri EAP ile hesaplanırken en yüksek değeri MO kestirimi vermektedir. 8. sınıf düzeyinde ise en



düşük ortalama  $\theta$  değeri MAP kestirimi ile hesaplanırken en yüksek değer MO kestirimi ile hesaplanan ortalama  $\theta$  değeridir. Ayrı kalibrasyonda 6. ve 8. sınıflarda MAP kestirimi MO ve EAP'a göre daha düşük sonuçlar verirken EAP ile MO karşılaştırıldığında EAP ile hesaplanan değerlerin daha düşük olduğu görülmektedir. 7. sınıfta ise MO en büyük değeri alırken EAP en düşük ortalama değerine sahiptir. Alan yazında yapılan çalışmalarda da benzer şekilde ortalama kriteri incelendiğinde ayrı kalibrasyonda ortalamalar açısından artan bir örüntü olduğu, yetenek kestirimleri (EAP, MAP, MO ve QD)'nin ise benzer sonuçlar verdiği bulunmuştur (Kolen ve Tong, 2007)

Sınıf düzeyleri arasındaki büyüme örüntüsünü belirlemek için ardışık sınıf düzeyleri arasındaki ortalama farkları hesaplanmış ve bulunan değerler Tablo 9'da gösterilmiştir. Ortalama farklarının gösterildiği Tablo 9 incelendiğinde ayrı kalibrasyon yöntemi ve MO, EAP ve MAP ile kestirilen 6. ve 7. sınıflar arasındaki ortalama  $\theta$  farkları negatif iken 7. ve 8. sınıf düzeylerinde ise pozitif olduğu görülmektedir. Bu durumda 6 ve 7. sınıflar arasında başarıda azalma olduğu yani 6. sınıf öğrencilerinin 7. sınıf öğrencilerine göre daha başarılı olduğu şeklinde yorumlanabilir. 8. sınıf ile 7. sınıf arasında ortalamalar arasındaki farkın pozitif olması ise 7. sınıftan 8. sınıfa geçince öğrencilerin başarılarının arttığını belirtmektedir. Bu sonuçtan farklı olarak Kolen ve Tong (2010) sınıf düzeyleri arasındaki ortalama farkların düşük sınıflarda fazla, sınıf düzeyi arttıkça ise azalmakta olduğunu bulmuşlar ve bu durumu düşük sınıf düzeylerinde daha fazla büyüme olduğu şeklinde yorumlamışlardır. Ayrı kalibrasyonda  $\theta$  ortalamaları arasındaki en yüksek fark 7 ve 8. sınıflar arasında EAP kestirimi ile elde edilen 0.258 değeridir.

Ayrı kalibrasyon yöntemi ile maksimum olabilirlik, EAP ve MAP yetenek kestirim yöntemleri kombinasyonunda standart sapmaların nasıl değiştiği Tablo 9’da gösterilmiştir. Tablo 9 ayrı kalibrasyonda EAP, MAP ve MO kestirimleriyle elde edilen standart sapmaların 7. sınıf düzeyinde en düşük değerleri alırken en yüksek standart sapma değerlerinin 6. sınıf düzeyindeki standart sapma değerleri olduğu göstermektedir. Bu durum 7. sınıfın diğer sınıflara göre daha homojen bir grup iken 6. sınıfın ise diğer sınıflara göre daha heterojen bir grup olduğu şeklinde yorumlanabilir. Bu sonuca benzer şekilde Kolen ve Tong (2007) yaptıkları çalışmalarda ayrı kalibrasyonda standart sapmaların artıp azalmakta olduğunu bulmuşlardır. Bir diğer çalışmada da ayrı kalibrasyonda  $d = 1$  ve ortak madde gücünü sınırlandırıldığı durumda varyanslar arasında azalan bir örüntü olduğu bulunmuştur (Chin, Kim, ve Nering, 2006). Tablo 9’da en düşük standart sapma değeri EAP ile 7. sınıf düzeyinde hesaplanan 0,251’dir. Bu bulguyu destekler şekilde Kim, Lee, Kim ve Kelly (2009) çalışmalarında ayrı kalibrasyonda EAP’ın en düşük standart sapma ürettiğini bulmuşlardır. Ayrıca EAP ile MO kıyaslandığında EAP’ın düşük varyans ürettiği de bu bulguyu destekleyen, alan yazında bulunan bir diğer sonuçtur (Beard, 2008). Tablo 9’dan en yüksek standart sapma değerinin 6. sınıf düzeyinde MO ile hesaplanan 1.191 olduğu görülmektedir. Ayrı kalibrasyonda 6. ve 8. sınıf düzeyinde de MAP kestirimi en düşük değerleri verirken 7. sınıf düzeyinde en düşük değeri EAP vermektedir. En yüksek değerlerin her üç sınıf düzeyinde de MO kestirimi ile hesaplanan değerler olduğu görülmektedir.

Yetenek kestirimlerinin sınıf düzeyleri arasındaki ayrımı ya da benzerliği etki büyüklüğü ve yatay uzaklık birimleri ile analiz edilmiştir. Etki büyüklüğü standardize edilmiş ortalama farklarıdır. Etki büyüklüğünün küçük olması ard arda gelen sınıf

düzeyleri arasında yetenek düzeyleri arasındaki farkın az olduğunun göstergesidir. Ayrı kalibrasyonda maksimum olabilirlik, EAP ve MAP ile kestirilen yeteneklerin etki büyüklükleri ve ortalama yatay uzaklık değerleri Tablo 10’da gösterilmiştir.

Tablo 10 Ayrı Kalibrasyonda Ard arda Gelen Sınıf Düzeyleri için Etki Büyüklükleri ve Ortalama Yatay Uzaklıklar

	Sınıf Düzeyleri	EAP	MAP	MO
Etki Büyüklükleri	6-7	-0.322	-0.213	-0.178
	7-8	1.345	0.366	0.318
Ortalama Yatay Uzaklıklar	6-7	-0.23	-0.13	-0.20
	7-8	0.29	0.21	0.22

Ard arda gelen sınıf düzeylerinde hesaplanan etki büyüklüğünün 6 ve 7. sınıflarda negatif değere sahipken 7 ve 8. sınıf düzeylerinde pozitif değerler aldığı görülmektedir. Her üç yetenek kestirim yönteminde de benzer örüntü görülmektedir. Kolen ve Tong (2007) çalışmalarında etki büyüklüğünün ayrı kalibrasyonda genellikle EAP’ın etki büyüklüğünün MO ve MAP’dan daha küçük olduğunu bulmuşlardır. Bu çalışmada da EAP ile hesaplanan etki büyüklükleri MAP ve MO ile hesaplanan etki büyüklüklerinden daha büyüktür. Bu sonuç Bear (2008)’in simülasyon verileri kullanarak yaptığı çalışma sonucu elde ettiği bulguyla örtüşmektedir.

Ard arda gelen sınıf düzeyleri arasında ayrı kalibrasyon yöntemi ve EAP, MAP ve MO yetenek kestirimleri ile yapılan dikey ölçekleme sonucu elde edilen ortalama yatay uzaklıklar hesaplanmış ve bu değerler Tablo 10’da gösterilmiştir. Ayrı kalibrasyonda maksimum olabilirlik, EAP ve MAP ile kestirilen yetenek dağılımları arasındaki

ortalama yatay uzaklık deęerleri 6. ve 7. sınıflar için negatifken 7. ve 8. sınıflar arasında ise pozitifolarak bulunmuştur. En düşük ortalama yatay uzaklık deęerini 6. ve 7. sınıflar arasında MAP verirken en yüksek deęeri ise EAP vermektedir. 7. ve 8. sınıflar arasında ise bu duruma benzer bir durum söz konusudur. En düşük ortalama yatay uzaklık deęeri MAP ile hesaplanırken en yüksek deęer ise EAP ile hesaplanan deęerdir. Ortalama yatay uzaklıklar ortalamalar arasındaki fark ile benzer sonuçlar vermektedir.

Arda arda gelen sınıf düzeyleri arasındaki farkı ölçmenin bir başka yolu ise dağılım boyunca seçilen yüzdelerdeki yatay uzaklıkları hesaplamaktır. Yatay uzaklıklar 5, 10, 25, 50, 75, 90 ve 95. yüzdeler için ayrı kalibrasyon ve EAP, MAP ve MO yetenek kestirimlerinden elde edilen dikey ölçekleme sonuçlarından hesaplanmıştır. Seçilen yüzdelerde yatay uzaklıklar hesaplanıp bulunan sonuçlar Tablo 11’de gösterilmiştir.

Tablo 11 Ayrı Kalibrasyonda Seçilen Yüzdelerdeki Yatay Uzaklıklar

Yetenek Kestirimi	Sınıf Düzeyi	% 5	% 10	% 25	% 50	% 75	%90	% 95
EAP	6-7	0.870	0.755	0.423	-0.088	-0.776	-1.401	-1.756
	7-8	-0.404	-0.402	0.755	0.142	0.566	1.099	1.744
MAP	6-7	0.561	0.449	0.180	-0.129	-0.486	-0.692	-0.822
	7-8	-0.099	-0.106	0.031	0.187	0.341	0.458	0.739
MO	6-7	0.822	0.643	0.276	-0.119	-0.631	-1.060	-1.428
	7-8	-0.166	-0.172	0.006	0.186	0.361	0.469	0.935

Tablo 11 incelendiğinde ise yatay uzaklık deęerlerinin 6 ve 7. sınıflar arasında % 50’ye kadar pozitif azalan deęerler almakta iken %50’den sonra negatif ve mutlak deęerce

artan deęerler aldıęı grlmektedir. %50'lik dilim dıřında EAP en byk deęerleri MAP ise en kk deęerleri vermektedir. 7 ve 8. sınıf arasında ise % 25'e kadar yatay uzaklıklar negatif iken bu deęerden sonra pozitif deęerler almıřtır ve yzdelik dilim arttıka yatay uzaklık artmıřtır. %25 ve % 50'lik dilimler dıřında 7 ve 8. sınıflar arasında da 6. ve 7. sınıflara benzer řekilde EAP en byk deęerleri MAP ise en kk deęerleri vermektedir. Kolen ve Tong (2007) ayrı kalibrasyonda yatay uzaklıkların dřk yzdeliklerde ve kk sınıflarda daha byk iken yksek yzdeliklerde ve byk sınıflarda arttıęını ancak bu artıřın ok ok az olduęunu belirtmektedirler. Bu da bymenin ok az olduęu řeklinde yorumlanmıřtır. Ayrıca yatay uzaklıklar yetenek kestirim yntemlerine gre incelendięinde benzer sonular verdięi gzlenmiřtir. Bir bařka alıřmanın sonularına gre yatay uzaklıklar u noktalarda daha dřk deęerler alırken orta noktalarda daha yksek deęerler almaktadır (Beard, 2008). Ayrı kalibrasyonda her u yetenek kestiriminde 6. ve 7. sınıflar arasında yatay uzaklıklar %50'ye kadar pozitif, bu deęerden sonra negatif deęer almaktadır. Bu durum dřk yzdelik diliminde fark azalmakta iken yzdelik dilim arttıka iki sınıf dzeyi arasındaki fark 6. sınıf lehine artmaktadır řeklinde yorumlanabilir. 7 ve 8. sınıflar arasında ise yzdelik dilim arttıka 8. sınıf lehine fark artmaktadır.

### **3.2 ALT PROBLEM 2'YE AİT BULGULAR VE YORUM**

**İkinci Alt Problem:** Eř zamanlı kalibrasyon yntemiyle elde edilen MO, EAP ve MAP yetenek kestirimlerinin ard arda gelen sınıf dzeyleri arasındaki,

- a. ortalama farkları ve standart sapmaları,
- b. etki byklkleri,
- c. yatay uzaklıkları nasıldır?

Arda arda gelen sınıf düzeyleri arasındaki ortalama  $\theta$  değerleri, ortalamalar arasındaki farklar ve standart sapma değerleri eş zamanlı kalibrasyon yöntemi ve maksimum olasılık, EAP ve MAP ile kestirilmiş ve elde edilen sonuçlar Tablo 12’de gösterilmiştir.

Tablo 12 Eş Zamanlı Kalibrasyonda Sınıf Düzeylerine Göre  $\theta$  Ortalamaları, Ortalamalar Arası Farkları ve Standart Sapmaları

	Sınıf Düzeyleri	EAP	MAP	MO
Ortalama	6	0.000	0.000	0.000
	7	-1.859	-2.124	-1.492
	8	0.338	0.509	0.244
Ortalamalar arası fark	6-7	-1.859	-2.124	-1.492
	7-8	2.197	2.633	1.736
Standart sapma	6	1.000	1.000	1.000
	7	0.700	0.459	0.482
	8	1.115	1.326	1.117

Eş zamanlı kalibrasyonda 6. sınıf referans sınıf olarak alındığı için her üç kalibrasyon yönteminde de ortalama  $\theta$  değerleri sıfırdır. Tablo 12 incelendiğinde her üç yetenek kestirim yönteminde de ortalama  $\theta$  değerlerinin 6. sınıf düzeyinden 7. sınıf düzeyine azalırken 8. sınıf düzeyinde ise tekrar artmakta olduğu görülmüştür. 8. sınıf düzeyinde MAP ile kestirilen 0.509 değeri en yüksek ortalama değeri olarak bulunurken 7. sınıf düzeyinde MAP ile kestirilen -2.124 en düşük değer olarak bulunmuştur.

Tablo 12’de sınıf düzeyleri arasındaki büyüme örüntüsü incelendiğinde eş zamanlı kalibrasyon yöntemi ve MO, EAP ve MAP ile kestirilen 6. ve 7. sınıflar arasındaki ortalama  $\theta$  farkları negatif iken 7. ve 8. sınıf düzeylerinde ise pozitif olarak bulunmuştur. En büyük  $\theta$  ortalama farkı 7 ve 8. sınıflar arasında MAP kestirimi ile 2.633 olarak hesaplanmıştır. 6 ve 7. sınıflar arasında en düşük değer MAP ve ayrı kalibrasyon verirken en yüksek değer ise yine MAP ve eş zamanlı kalibrasyon ile elde edilen değerdir. Ortalamalar arasındaki fark değerleri incelendiği zaman en düşük değerlerin MO ile hesaplanan değerler iken en yüksek değerler MAP ile hesaplanan değerlerdir. 7 ve 8. sınıflar arasında ise en yüksek değeri MAP; en düşük değeri ise MO vermektedir. Genel olarak eş zamanlı kalibrasyonda ortalamalar arasındaki farklar incelendiği zaman en düşük değerlerin MO kestirimi sonucu elde edilen değerler iken en büyük değerlerinde MAP kestirimi sonucu hesaplanan değerler olduğu görülmektedir. Ayrı kalibrasyona benzer şekilde eş zamanlı kalibrasyonda da 6 ve 7. sınıflar arasında başarıda azalma olduğu yani 6. sınıf öğrencilerinin 7. sınıf öğrencilerine göre daha başarılı olduğu şeklinde yorumlanabilir. 8. sınıf ile 7. sınıf arasında ortalamalar arasındaki farkın pozitif olması ise 7. sınıftan 8. sınıfa geçince öğrencilerin başarılarının arttığını belirtmektedir. Bu bulgulardan farklı olarak Chin, Kim ve Nering (2006) tarafından yapılan çalışmada eş zamanlı kalibrasyonda ortalamalar arasındaki farklar için artma ya da azalma örüntüsü gözlenmemiştir.

Eş zamanlı kalibrasyon yöntemi ve MO, EAP ve MAP yetenek kestirim yöntemleri kombinasyonunda standart sapmaların nasıl değiştiği Tablo 12’de gösterilmiştir. Eş zamanlı kalibrasyonda 6. sınıf ortalaması 0 standart sapması 1 olan referans grup olduğu için 6. sınıfın standart sapma değerleri her üç yetenek kestirim yöntemi içinde 1 olarak alınmıştır. Tablo 12 incelendiğinde MO, EAP ve MAP kestiriminden elde edilen

standart sapmalar 7. sınıf düzeyinde her üç yetenek kestirim yöntemi için de en düşük değerleri alırken en yüksek standart sapma değerlerinin 8. sınıf düzeyindeki değerler olduğu görülmektedir. Bu sonuçlar 7. sınıfın diğer sınıflara göre daha homojen bir grup olduğu şeklinde yorumlanabilir. 7. sınıf düzeyinde MAP ve MO yöntemleriyle hesaplanan standart sapma değerleri birbirine yakın sonuçlar verirken, EAP ile daha büyük standart sapma değeri elde edilmektedir. 8. Sınıfta en düşük standart sapma değerini EAP alırken en büyük değeri MAP almaktadır. Ancak Kim (2007) çalışmasında matematik testinde standart sapmaların kalibrasyon yöntemlerinde benzerlik gösterdiğini, yetenek kestirimlerinde ise MO'nun EAP'a göre daha büyük standart sapma değerleri ürettiğini bulmuştur.

Eş zamanlı kalibrasyon yöntemi ve MO, EAP ve MAP ile kestirilen yeteneklerin etki büyüklükleri ve ortalama yatay uzaklıkları ard arda gelen sınıf düzeylerinde hesaplanmış ve tablo 13'te gösterilmiştir.

Tablo 13 Eş Zamanlı Kalibrasyonda Ard arda Gelen Sınıf Düzeyleri için Etki Büyüklükleri ve Ortalama Yatay Uzaklıklar

	Sınıf Düzeyleri	EAP	MAP	MO
Etki Büyüklükleri	6-7	-2.151	-2.723	-1.900
	7-8	2.360	1.336	2.016
Ortalama Yatay Uzaklıklar	6-7	-1.82	-2.08	-1.46
	7-8	2.25	2.68	1.91

EAP, MAP

ve MO kestirimleri için etki büyüklüklerinin 6 ve 7. sınıflarda negatif değere sahipken 7 ve 8. sınıf düzeylerinde pozitif değerler aldığı Tablo 13' te görülmektedir. 6 ve 7.



sınıflar arasında MAP kestirimiyle hesaplanan etki büyüklükleri EAP ve MO kestirimlerinden elde edilen değerlere kıyasla daha büyük değerler vermektedir. 6 ve 7. sınıflar arasında en büyük etki büyüklüğü MAP ile hesaplanan değer olurken 7 ve 8. sınıflar arasında da en büyük değer EAP ile hesaplanan değer olduğu tablodan görülmektedir. Alan yazında etki büyüklükleriyle ilgili farklı bulgulara rastlanmıştır. EAP ile kestirilen etki büyüklüklerinin diğer yetenek kestirimlerine göre daha büyük olduğu bulunmuştur (Kim, 2007). Benzer şekilde bir başka çalışmada da etki büyüklüklerinin ayrı ve eş zamanlı kalibrasyon yönteminde de benzer sonuçlar verdiği görülmüştür. Ancak yetenek kestirim yöntemleri dikkate alındığında EAP ve MAP ile hesaplanan etki büyüklüklerinin MO ve test karakteristik eğrisi yöntemleriyle hesaplanan değerlere göre daha büyük olduğu belirtilmiştir (Tong ve Kolen, 2010).

MO, EAP ve MAP yetenek kestirimleri ve eş zamanlı kalibrasyon ile elde edilen yetenek dağılımları arasındaki ortalama yatay uzaklık değerleri 6 ve 7. sınıflar için negatifken 7 ve 8. sınıflar arasında ise pozitifdir. Ortalama yatay uzaklıklar MAP ile en büyük değerler alırken en küçük değerleri MO vermektedir. 7 ve 8. sınıflar arasında MAP kestirimi sonucu hesaplanan 2.68 değeri en büyük yatay uzaklık değeridir. Sınıf düzeyleri arttıkça ortalama yatay uzaklıklar artmaktadır. Kim'in (2007) çalışmasında ise sınıf düzeyi arttıkça yatay uzaklıkların azalmakta olduğunu bulunmuştur.

Etki büyüklükleri ve ortalama yatay uzaklıklar, yetenek kestirimlerinin sınıf düzeyleri arasındaki ayrımı ya da benzerliğini göstermektedirler. Etki büyüklüğünün küçük olması ard arda gelen sınıf düzeyleri arasında yetenek düzeyleri arasındaki farkın az olduğunun göstergesidir. Ortalama yatay uzaklıklar da etki büyüklüğü ile benzer

örüntüyü vermektedir; 6 ve 7. sınıflar arasında negatif, 7 ve 8. sınıflar arasında ise pozitifdir. Ayrı kalibrasyonla kıyaslandığında ise eş zamanlı kalibrasyondaki değerlerin daha büyük olduğu görülmektedir.

Eş zamanlı kalibrasyon yöntemi ve maksimum olabilirlik, EAP ve MAP kestirimleriyle elde edilen dikey ölçekleme sonuçlarından hesaplanan yatay uzaklıklar 5, 10, 25, 50, 75, 90 ve 95. yüzdeler için hesaplanıp bulunan sonuçlar Tablo 14'te gösterilmiştir.

Tablo 14 Eş Zamanlı Kalibrasyonda Seçilen Yüzdelerdeki Yatay Uzaklıklar

Yetenek	Sınıf								
	Kestirimi	Düzeyi	% 5	% 10	% 25	% 50	% 75	%90	% 95
EAP		6-7	-3.722	-3.329	-2.660	-1.875	-1.096	-0.351	0.095
		7-8	0.222	0.529	1.266	2.188	3.047	3.969	4.407
MAP		6-7	-3.861	-3.471	-2.866	-2.109	-1.412	-0.741	-0.441
		7-8	0.601	1.045	1.682	2.494	3.431	4.377	5.345
MO		6-7	-3.588	-2.901	-2.159	-1.436	-0.840	-0.216	0.110
		7-8	0.034	0.440	1.045	1.677	2.509	3.407	4.842

Eş zamanlı kalibrasyon ve EAP, MAP ve MO kestirimleri sonucu seçilen yüzdeler için hesaplanan yatay uzaklıklar incelendiğinde 6. ve 7. sınıflar arasında MAP kestirimi sonucu hesaplanan yatay uzaklık değerlerinin tamamının negatif ve mutlak değerce azalan değerler aldığı görülmüştür. Aynı sınıf düzeyleri için EAP ve MO %95'e kadar negatif değerler alırken % 95'te pozitif değer almaktadırlar. Yatay uzaklık değerlerinin negatif ve mutlak değerce azalan değerler alması yüzdeler arttıkça 6 ve 7. sınıflar arasındaki farkın azaldığı şeklinde yorumlanabilir. Bir diğer ifadeyle 6.

sınıf düzeyindeki öğrenciler 7. sınıf düzeyine göre daha başarılıdır ve yüzdeler arttıkça başarılar arasındaki fark azalmaktadır. 7 ve 8. sınıflar arasında, seçilen yüzdeler dilimlerde hesaplanan yatay uzaklık değerlerinin her üç yetenek kestirim yöntemi için pozitif ve yüzdeler arttıkça artan değerler aldığı görülmüştür. Benzer şekilde 7. ve 8. sınıflar arasında da yüzdeler arttıkça iki sınıf düzeyi arasındaki fark artmaktadır. Bir başka ifadeyle başarılı öğrenciler arasındaki fark fazla iken başarısız öğrenciler arasındaki fark daha azdır. Yetenek kestirim yöntemlerini kıyasladığımız zaman %95'lik dilime kadar en büyük değerleri MAP, en küçük değerleri MO verirken %95'te en büyük değerleri yine MAP verirken en küçük değerleri MO kestiriminin verdiği görülmektedir. Kim (2007)'in çalışmasında kalibrasyon yöntemleri ve yetenek kestirim yöntemlerine göre belirli yüzdelerde hesaplanan yatay uzaklıkların benzer sonuçlar verdiğini bulmuştur. Buna ek olarak sınıf düzeyleri arasında yüzdelerde yatay uzaklık sonuçları da benzerdir. Fakat çoğu durumda düşük yüzdelerde (%5 ve %10) sınıf düzeyleri arasında yatay uzaklıkların çok az da olsa artmakta olduğu gözlemlenmiştir.

### **3. 3 YETENEK KESTİRİM YÖNTEMLERİ VE KALİBRASYON**

#### **YÖNTEMLERİNİN KARŞILAŞTIRILMASI**

Farklı sınıf düzeyleri için yapılan dikey ölçeklemeler sonucunda elde edilen ortalama  $\theta$  değerleri kalibrasyon yöntemleri dikkate alınarak incelendiğinde; sınıf düzeyleri için farklı yetenek kestiricileri ve kalibrasyon yöntemleriyle elde edilen dikey ölçekleme sonuçlarına göre genel olarak en düşük ortalamaya sahip sınıfın 7. sınıf olduğu görülmektedir. Ortalama  $\theta$  değerleri 6. sınıf düzeyinde MAP ile ayrı kalibrasyon yöntemi kombinasyonundan elde edilen değer dışında diğer yetenek kestirim ve

kalibrasyon yöntemlerinde pozitif değerler almaktadır. 7. sınıf düzeyi için hesaplanan ortalama  $\theta$  değerlerinin hepsi her iki kalibrasyon yöntemi ve her üç yetenek kestirimi sonucunda negatif olarak bulunmuştur. 7. sınıf düzeyi için en düşük ortalama  $\theta$  değeri MAP ile eş zamanlı kalibrasyon sonucu elde edilen dikey ölçekleme sonucu hesaplanan -2.124 iken en yüksek değer MO ile ayrı kalibrasyon sonucu yapılan dikey ölçeklemeden elde edilen -0.155'tir. 8. sınıfta ise MAP kestirimiyle ayrı kalibre edilerek elde edilen ortalama -0.038 değerleri dışındaki diğer  $\theta$  değerleri pozitif değerler almaktadır. En yüksek ortalama  $\theta$  değeri MAP kestirimiyle eş zamanlı kalibre edilerek elde edilen 0.509 değeridir. Özetle iki kalibrasyon yönteminde de 6 sınıftan 7. sınıfa geçildiğinde ortalamaların düştüğü, 8. sınıfta ise tekrar arttığı görülmektedir. Benzer sonuçlar Kolen ve Tong (2008), Boughton, Lorie ve Yao (2005) ve Tong ve Kolen (2007) tarafından yapılan çalışmalarda da görülmüştür. Kolen ve Tong'un (2008) çalışmalarında İngilizce testinde genel olarak sınıf düzeyi arttıkça diğer sınıflarda ortalamalarda artmıştır. Bu artma 6. ve 7. Sınıflar arasında gözlenmemiş ve 6. sınıfın ortalaması 7. sınıfın ortalamasından daha büyük olarak bulunmuştur. Boughton, Lorie ve Yao'nun (2005) çalışmalarında ise 7. sınıfın 8 ve 9. sınıflara göre daha başarılı olduğu bulunmuştur. Ayrıca Tong ve Kolen'in (2007) çalışmasında MTK ile ölçeklemede öğrencilerin başarılarındaki dağılım sınıf düzeyleri boyunca kararsızlık göstermekte veya azalmakta olduğu bulunmuştur.

İki kalibrasyon yöntemi karşılaştırıldığında ise ortalama değerlerinin eş zamanlı kalibrasyonda ayrı kalibrasyona göre daha yüksek olduğu görülmektedir. Ortalamalar arasındaki fark incelendiği zaman da benzer bulgulara ulaşılmaktadır. Öğrencilerin sınıf düzeyleri değiştikçe başarılarındaki değişim yani 6. sınıftan 7. sınıfa geçince başarılarındaki düşüş ve 7. sınıftan 8. sınıfa geçince tekrar artmasının nedenlerinin ne

olabileceği 7 matematik öğretmenine sorulmuştur. Öğretmenler 6. sınıfta konuların çoğunlukla 4. ve 5. sınıfın tekrarı olduğunu, yani konuların çoğunu öğrencilerin daha önceden öğrendiğini ancak 7. sınıfta daha fazla yeni konu gördüklerini ve programın yoğun olduğunu bu nedenle zorlandıklarını ifade etmişlerdir. Ayrıca görüşülen öğretmenler 7. sınıfta öğrencilerin yavaş yavaş ergenlik dönemine girmeye başladığını dolayısıyla bu durumun öğrencilerin dikkatlerini ve motivasyonlarını etkilediğini belirtmişlerdir. Keskin ve Sezgin (2009)'in çalışması, ön ergenlik dönemini yaşamaya başlayan çocukların bu döneme özgü ihtiyaçların ağır basması dolayısıyla ders çalışmak, okul başarısı gibi ihtiyaçlarını daha geri plana ittiğini ve ders başarılarının düştüğünü göstermektedir. Bu araştırmada, 7. sınıfların başarısında görülen düşüşün nedeni de, öğrencilerin ön ergenlik dönemine girmiş olmasının bir sonucu olabilir. Öğretmenler, 8. sınıfta konuların öğrencileri 7. sınıftaki kadar zorlamadığını içerik olarak da 7. sınıftaki kadar yoğun olmadığını belirtmişlerdir. Buna ek olarak görüşülen öğretmenler 8. sınıfta hem öğrencilerin hem de velilerin sınavı daha çok ciddiye aldıklarını ve bu nedenle 8. sınıfta özellikle SBS'ye yönelik 6. ve 7. sınıfa göre daha fazla takviye aldıklarını belirtmişlerdir. Öğretmenlerin bu yorumlarına paralel olarak Sarier (2010) özel derse, dersanelere devam eden öğrenci sayılarının, 2002–2006 yılları arasında sürekli artış gösterdiğini belirtmektedir

Sınıf düzeyleri arasındaki büyüme örüntüsünü belirlemek için ardışık sınıf düzeyleri arasındaki ortalama farkları incelendiğinde 6. ve 7. sınıflar arasındaki ortalama  $\theta$  farkları (-2.124 ile -0.141) arasında değerler almakta iken 7. ve 8. sınıflar arasında hesaplanan ortalama  $\theta$  farkları ise (0.199 ile 2.633) arasındadır. En yüksek ortalama farkı MAP ve eş zamanlı kalibrasyon sonucunda 7. ve 8. sınıflar arasında 2.633 olarak hesaplanırken en düşük ortalama  $\theta$  farkı MAP ve eş zamanlı kalibrasyon ile 6. ve 7.

sınıflar arasında -0.141 olarak hesaplanmıştır. Ortalama  $\theta$  farkları eş zamanlı kalibrasyonda daha büyük sonuçlar vermektedir. Alan yazında büyüme örüntüsünü belirleyen ortalama farklarıyla ilgili farklı sonuçlar bulunmuştur. Yaptığı çalışmada Kim (2007) küçük sınıflarda eş zamanlı kalibrasyonun ayrı kalibrasyona göre daha düşük ortalama farkı verdiğini bulurken büyük sınıflarda eş zamanlı kalibrasyonun hesaplanamadığını belirtmiştir. Büyüme örüntüsünü tahmin etmede genellikle en kötü sonucu SL yöntemiyle ayrı kalibrasyon vermektedir. Bu bulgu ayrı kalibrasyon alan yazında en çok kullanılan yöntem olduğu için önemlidir (Meng, 2007). Bu bulgulardan farklı olarak matematik testinde 5. sınıftan 10. sınıfa kadar yapılan dikey eşitlemede ayrı kalibrasyonun daha düşük sonuçlar verdiği bulunmuştur (Karkee et al., 2003). Kim, Lee, Kim ve Kelly (2009) ortalamalar arasındaki farkın yetenek kestiriminden etkilenmezken kalibrasyon yönteminden etkilendiğini bulmuşlardır. Sınıf düzeyleri arttıkça ortalamalar artmaktadır ve yetenek kestirim yöntemlerinde benzer sonuçlar vermektedir (Kolen ve Tong, 2010).

Sınıf düzeyleri arasındaki benzerlik veya farklılıkları gösteren standart sapma değerleri incelendiğinde her iki kalibrasyon yöntemi ve her üç yetenek kestiriminde de en düşük standart sapmaların 7. sınıf düzeyinde elde edildiği görülmektedir. Bir başka ifadeyle her iki kalibrasyon yöntemi ve her üç yetenek kestiriminde en homojen sınıf 7. sınıftır. En yüksek standart sapma MAP ve eş zamanlı kalibrasyon kombinasyonu sonucu 8. sınıf düzeyinde 1.326 olarak hesaplanırken en düşük standart sapma değeri ise EAP ve ayrı kalibrasyon sonucunda 7. sınıf düzeyinde 0.251 olarak hesaplanmıştır. 8. sınıf ve MO kombinasyonu dışında diğer durumlarda eş zamanlı kalibrasyon ayrı kalibrasyona göre daha büyük standart sapma değerleri vermektedir. Kalibrasyon yöntemlerine göre karşılaştırıldığında ayrı kalibrasyonda EAP ve MAP eş zamanlı kalibrasyona kıyasla

daha düşük standart sapma deęerleri üretirken MO (8. sınıf düzeyi hariç) daha yüksek standart sapma deęeri üretmektedir. Standart sapmalar genel olarak ortalamalara benzer şekilde 6. sınıftan 7. sınıfa azalmakta, 8. Sınıfta ise tekrar artmaktadır. Bu durum Chin, Kim,ve Nering'in (2006) standart sapmaların sınıf düzeyi arttıkça azalma eğiliminde olduęu bulgusuyla çelişmektedir. Bu bulgulardan farklı olarak bir başka çalışmada standart sapmaların azalıp daha sonra düz bir çizgi izlemesi puan dağılımlarının sınıf düzeyleri arasındaki ayrımının açıkça sabit kaldığının ve bu durumun başarılı ve başarısız öğrencilerin sınıf düzeyleri arttıkça birbirleriyle neredeyse aynı olduğunun göstergesi olarak kabul edilmiştir (Tong ve Kolen, 2008). Standart sapma kriterinin kalibrasyon yönteminden deęil de yetenek kestirim yöntemlerinden daha çok etkilendięi belirtilmektedir (Kim, Lee, Kim ve Kelly, 2009). Kalibrasyon yöntemlerine göre standart sapmalar farklılık göstermezken yetenek kestirim yöntemi dikkate alındığında EAP en küçük hatayı üretmektedir (Kolen ve Tong, 2010).

Yetenek düzeyleri arasındaki farkı belirleyebilmek için etki büyüklükleri hesaplanmıştır. Etki büyüklükleri incelendiğinde 6. ve 7. sınıflar -2.151 ile -0.178 arasında deęerler almakta iken 7 ve 8. sınıflar ise 0.318 ile 2.360 arasında deęerler almaktadır. Ard arda gelen sınıf düzeyleri arasındaki etki büyüklüklerinin her iki kalibrasyon yönteminde de 6 ve 7. sınıflar arasında negatif deęerler alırken 7 ve 8. sınıflar arasında ise pozitif deęerler aldığı görülmektedir. Bu durum 6 ve 7. sınıflar arasında yetenek düzeyleri arasındaki fark 6. sınıf lehine iken 7 ve 8. sınıf düzeylerinde ise 8. sınıf lehine arttığını göstermektedir. Etki büyüklüğünün küçük olması ard arda gelen sınıf düzeyleri arasında yetenek düzeyleri arasındaki farkın az olduğunun göstergesidir. Bu nedenle yapılan çalışmada artan bir büyüme örüntüsü olduğu söylenebilir. Ayrı kalibrasyonda her üç yetenek kestirim yönteminde etki büyüklükleri

eş zamanlı kalibrasyona kıyasla daha küçüktür. Bu bulgulardan farklı şekilde Kolen ve Tong (2010) çalışmalarında etki büyüklüğünün küçük sınıflarda yüksek, büyük sınıflarda düşük olduğunu ve bunun azalan bir büyüme örüntüsünün göstergesi olduğunu belirtmektedirler. Yapılan çalışmalarda etki büyüklükleri sınıf düzeyleri arttıkça azaldığı (Tong & Kolen, 2008; Tong & Kolen, 2010) ve ikili-eş zamanlı kalibrasyonun etki büyüklükleri açısından en iyi sonucu verdiği belirtilmektedir (Meng, 2007).

Ortalama yatay uzaklıklar ise 6 ve 7. sınıflar arasında (-2.08 ile -0.13) arasında değerler alırken 7 ve 8. sınıflar arasında ise (0.22 ile 2.68) arasında değerler almaktadır. Eş zamanlı kalibrasyondaki değerler ayrı kalibrasyona göre daha büyüktür. Ortalama yatay uzaklıklar etki büyüklüğü ve ortalamalar arasındaki fark ile benzer sonuçlar vermektedir.

Belirli yüzdeler dilimlerde hesaplanan yatay uzaklık değerleri incelendiğinde ayrı kalibrasyonda 6 ve 7. sınıflar arasında % 50'ye kadar pozitif ve azalan değerler alırken bu yüzdeler dilimden sonra negatif ve mutlak değerce artan değerler almaktadır. Bu durum başarı düzeyi düşük öğrenciler arasında 7. sınıf öğrencilerinin 6. sınıf öğrencilerine göre daha başarılı olduğu ancak başarı düzeyi arttıkça bu farkın azaldığı ve % 50'lik dilimden sonra 6. sınıf lehine arttığı şeklinde yorumlanabilir. Bir başka ifadeyle başarısız öğrenciler arasındaki fark fazla iken başarılı öğrenciler arasındaki fark azalmakta ve küçük sınıflardaki başarılı öğrenciler lehine başarıda bir artma söz konusudur. Yatay uzaklıklar 7. ve 8. sınıflar arasında ise % 25'e kadar negatif, bu yüzdeler dilimden sonra pozitif ve artan değerler almaktadır. Bir başka deyişle başarı düzeyi arttıkça öğrenciler arasındaki fark da artmaktadır.



Belirli yüzdilik dilimlerde hesaplanan yatay uzaklık değerleri eş zamanlı kalibrasyonda ise her üç yetenek kestirimi için negatif ve mutlak değerce azalan değerler alması 6 ve 7. sınıflar arasındaki farkın yedinci sınıf lehine yavaş yavaş arttığı şeklinde yorumlanabilir. 7 ve 8.sınıflar arasında pozitif aratan değerler alması da yetenek düzeyleri arasındaki farkın 7 ve 8.sınıflar arasında arttığı ve başarılı öğrenciler arasında farkın daha fazla olduğu anlamına gelmektedir. 7 ve 8. sınıflar arasında eş zamanlı kalibrasyon ayrı kalibrasyona göre daha büyük değerler almaktadır.

## BÖLÜM IV

### SONUÇ VE ÖNERİLER

Bu bölümde araştırma sonucu elde edilen ve bir önceki bölümde açıklanan bulgulara dayalı sonuçlar belirtilmiş ve bu sonuçlar doğrultusunda geliştirilen bazı önerilere yer verilmiştir.

#### 4.1 SONUÇLAR

Araştırma sonucunda ulaşılan sonuçlar öncelikle alt problemlerin sırasına uygun olarak aşağıda sunulmuştur son bölümde ise kalibrasyon yöntemleri karşılaştırılmıştır.

1) Ayrı kalibrasyonda;

- a. Ortalamalar arasındaki fark:  $EAP \rangle MO \rangle MAP$  şeklinde,
- b. Standart sapma değerleri 6.ve 8. sınıflarda:  $MO \rangle EAP \rangle MAP$  şeklinde iken 7. sınıfta ise  $MO \rangle MAP \rangle EAP$  şeklinde,
- c. Etki büyüklükleri:  $EAP \rangle MAP \rangle MO$  şeklinde,
- d. Ortalama yatay uzaklıklar:  $EAP \rangle MO \rangle MAP$  şeklinde,
- e. Yatay uzaklıklar:  $EAP \rangle MO \rangle MAP$  şeklinde değerler almaktadır.

2) Eş zamanlı kalibrasyonda;

- a. Ortalamalar arasındaki fark:  $MAP \rangle EAP \rangle MO$  şeklinde,

- b. Standart sapma deęerleri 7. sınıfta: EAP)MO)MAP şeklinde iken 8. sınıfta ise MAP )MO) EAP şeklinde,
- c. Etki büyüklükleri: 6. ve 7. sınıflar arasında MAP)EAP)MO şeklinde iken 7. ve 8. sınıflar arasında EAP)MO)MAP şeklinde,
- d. Ortalama yatay uzaklıklar: MAP)EAP)MO şeklinde,
- e. Yatay uzaklıklar: EAP)MO)MAP şeklinde deęerler almaktadır.

3) Kalibrasyon yöntemlerini kıyasladığında;

- a. Ortalamalama deęerleri 7. sınıfta ayrı kalibrasyonda daha yüksek iken 8. sınıfta eş zamanlı kalibrasyonda daha yüksektir.
- b. Ortalamalar arasındaki farklar karşılaştırıldığı zaman eş zamanlı kalibrasyonun daha büyük deęerler ürettięi görülmektedir.
- c. Standart sapma deęerleri incelendiğinde 6. sınıf MO ve 7. sınıf ve MO kombinasyonu dışında eş zamanlı kalibrasyon sonucu elde edilen deęerler daha büyüktür.
- d. Etki büyüklükleri ve ortalama yatay uzaklıklar eş zamanlı kalibrasyonda daha büyük deęerler almaktadır.
- e. Belirli yüzdelerde hesaplanan yatay uzaklıklar incelendięi zaman 7. ve 8. sınıflar arasındaki deęerlerin eş zamanlı kalibrasyonda daha büyük olduęu görülmektedir.

## 4.2 ÖNERİLER

Yapılan araştırmanın sonuçlarına ilişkin öneriler aşağıda sıralanmıştır.

### 4.2.1 Araştırma Sonuçlarından Çıkan Öneriler

1) Araştırmanın sonucunda, öğrencilerin anasınıfından 12. Sınıfa kadar olan süreçte başarılarındaki değişimlerinin izlenmesi için dikey ölçekleme yönteminin kullanılması önerilebilir.

2) Dikey ölçekleme karmaşık bir süreçtir ve bu süreç boyunca yapılan seçimler ölçeklemenin sonucunu dolayısıyla öğrencilerin başarılarındaki değişimi yorumlamayı da etkilemektedir. Dolayısıyla dikey ölçekleme sonucunda öğrencilerin başarıları hakkında önemli kararlar verilecekse farklı ölçme sonuçlarından elde edilen bulguları da dikkate almak gerekmektedir.

### 4.2.2 Araştırmacılara Yönelik Öneriler

1. Araştırmada gerçek veri seti kullanılarak iki kalibrasyon yöntemi ve üç yetenek kestirim yöntemi karşılaştırılmıştır. Bu nedenle ölçekleri karşılaştıracak bir standart ya da doğru değer yoktur. Simülasyon çalışmalarında doğru madde parametreleri ve yetenek kestirimleri bilindiği için hangi yöntemin daha doğru sonuç verdiği belirlenebilir. Bu durum başka bir araştırmanın konusu olabilir.

2. Bu çalışmada ayrı ve eş zamanlı kalibrasyon karşılaştırılmıştır. Sabit kalibrasyonun da dahil edildiği benzer bir çalışma alan yazına faydalı bilgiler sunabilir.

3. Bu çalışmada sadece üç sınıf düzeyinde eşitleme yapılmıştır. Daha fazla örneğin 4. sınıftan 8. sınıfa veya 12. sınıfa kadar sınıf düzeylerinin sayısı artırılarak yapılp bağlama sayısı arttıkça dikey ölçeklemenin nasıl değiştiği başka bir çalışmanın konusu olabilir.

4. Araştırmada ayrı kalibrasyonda SL bağlama yöntemi kullanılmıştır. Ayrı kalibrasyonda Heabera, Ortalama/ Ortalama ve Ortalama/ Sigma yöntemlerinin dikey ölçekleme sürecini etkileyip etkilemediğine dair bir çalışma yararlı olabilir.

5. Araştırmada temel sınıf olarak 6. sınıf alınmıştır. Farklı sınıf düzeylerinin temel sınıf olarak alınmasının dikey ölçekme sürecine etkileri başka bir araştırmanın konusu olabilir.

## KAYNAKÇA

- Beard, J. J.(2008). An Investigation of vertical scaling with item response theory using a multistage testing framework. Yayınlanmamış Doktora Tez. University of Iowa, Iowa.
- Becker, D. F. & Forsyth, R. A. (1992). An Empirical Investigation of Thurstone and IRT Methods of Scaling Achievement Tests. *Journal of Educational Measurement, Vol. 29(4), 341-354*
- Béguin, A. A., Hanson, A. B. & Cees A. W. Glas, C. A. W. (2000). *Effect of Multidimensionality on Separate and Concurrent Estimation in IRT Equating*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, April, 2000, New Orleans.
- Boughton, K.A., Lorie, W. & Yao, L. (2005). *A Multidimensional Multigrup IRT Models for Vertical Scales with Complex Test Structure: An Emprical Evaluation of Student Growth using Real Data*. National Council on Measurement in Education: 2005. Monreal/ Quebec/ Canada.
- Briggs,D. C., Weeks, J. P. & Wiley, E. (2008). *Vertical Scaling in Value-Added Models for Student Learning*. Paper presented at the National Conference on Value-Added Modeling, April 22-24, 2008, Madison, WI.
- Briggs, D. C. & Weeks, J. P. (2009). The Impact of Vertical Scaling Decisions on Growth Interpretations. *Educational Measurement: Issues and Practice, 28( 4), 3–14*

- Burg, S. (2008). An investigation of dimensionality across grade levels and effects on vertical linking for elementary grade mathematics achievement tests. NCME: NYC, 2008.
- Chin, T.Y. , Kim,W. & Nering, M. L. (2006). *Five Statistical Factors That Influence IRT Vertical Scaling*. Paper presented at the annual meeting of National Council on Measurement in Education (NCME) at San Francisco, April 2006.
- Cohen, L., Manion, L. Ve Morrison, K. (2007). *Research Methods in Education* (6th Edition). Routledge, Canada.
- Christine A., Reid, C. A., Kolakowsky-Hayner, S.A., Lewis, A. N. & Armstrong, A. J. (2007). Modern Psychometric Methodology: Applications of Item Response Theory. **Rehabilitation Counseling Bulletin RCB 50:3 pp. 177–188**
- Custer, M., Omar, H. & Pomplun, M. (2004). Vertical Scaling With the Rasch Model Utilizing Default and Tight Convergence Settings With WINSTEPS and BILOG–MG. *Applied Measurement in Education*, 19(2), 133-149
- Çetin, E. (2009). Dikey ölçeklemede klasik test ve madde tepki kuramına dayalı yöntemlerin karşılaştırılması. (Yayınlanmamış Doktora Tezi), Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers
- Fraenkel, J. R. & Wallen,E (2003). *How to Design and Evaluate Research in Educatiin*. New York, NY: McGraw-Hils Companies.

- Hanson, B. A. & Beguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26 (1), 3-24.
- Hanson, B.& Zeng, L.(Rev. Cui,Z 2004). *ST: A Computer Program for IRT Scale Transformation*.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Ito, K., Sykes, R. C. & Yao, L. (2008). Concurrent and Separate Grade-Groups Linking Procedures for Vertical Scaling. *Applied Measurement in Education*, 21, 187-206.
- Holland, P. W. (2002). Two measures of changes in the gaps between the CDFs of test score distributions. *Journal of Educational and Behavioral Statistics*, 27, 3- 17.
- Kang, T. & Petersen, N.S. (2009). Linking Item Parameters to a Base Scale. **ATC Research Report Series 2009-2**
- Karkee, T.; Lewis, D.M.; Hoskens, M.; Yao, L. & Haug, C. (2003). *Separate vs Concurrent Calibration Methods in Vertical Scaling*. Paper presented at the annual meeting of National Council on Measurement in Education .Chicago, IL, April 22-24, 2003.
- Keskin, G. & Sezgin, B. (2009) Bir grup ergende akademik başarı durumuna etki eden etmenlerin belirlenmesi, *Fırat Sağlık Hizmetleri Dergisi*, 4 (10),3-18.
- Kenyon, D. M., Gregor, D. M., Li, D. & Cook, G. (2011). Issues in vertical scaling of a K-12 English language proficiency test. *Language Testing*, 27(4) 443–469.



- Kim, J. (2007). A comparison of calibration methods and proficiency estimators for creating IRT vertical scales. (Yayınlanmamış Doktora Tezi), University of Iowa, Iowa.
- Kim, S.H.& Cohen, A.S. (1996). *A Comparison of Linking and Concurrent Calibration under Item Response Theory*. Paper presented at the annual meeting of American Educational Research Association, New York, NY, April 8-12,
- Kim, S. & Kolen, M. J. (2007). Effects on Scale Linking of Different Definitions of Criterion Functions for the IRT Characteristic Curve Methods. *Journal of Educational and Behavioral Statistics* , 32(4), 371-397
- Kim,J., Lee, W.C., Kim,D. & Kelley, K. ( 2009). *Investigation of Vertical Scaling Using the Rasch Model*. National Council on Measurement in Education: April 2009.
- Kolen, J. M. (2004). Linking Assessments: Concept and History. *Applied Psychological Measurement*, 28(4), 219–226.
- Kolen M. J. & Tong, Y. (2010). Psychometric Properties of IRT Proficiency Estimates. *Educational Measurement: Issues and Practice*, 29(3), 8-14
- Lee, W. C. & Ban, J. C. (2010). A Comparison of IRT Linking Procedures. *Applied Measurement in Education*, 23, 23-48
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Welsley Publishing Company.

Lord, F.M & Wingersky, M.S.(1984). Comparison of IRT True Score and Equipercentaile observed Score 'Equitings' , *Applied Psychological Measuement*, 8(4), 453-461

Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 17, 179-193.

Macro, G. L. (1977). Item Characteristic Curve Solutions to Three Intractable Testing Problems. *Journal of Educational Measurement*, 14 (2), 139- 160

MEB. (2009). İlköğretim Matematik Dersi 6-8. Sınıflar Öğretim Programı. Ankara, 2009.

MEB (2008a). 6. Sınıf seviye belirleme sınavı soru kitapçığı.

MEB (2008b). 7. Sınıf seviye belirleme sınavı soru kitapçığı.

MEB (2008c). 8. Sınıf seviye belirleme sınavı soru kitapçığı.

MEB (2009a). 6. Sınıf seviye belirleme sınavı soru kitapçığı.

MEB (2009b). 7. Sınıf seviye belirleme sınavı soru kitapçığı.

MEB (2009c). 8. Sınıf seviye belirleme sınavı soru kitapçığı.

MEB (2010a). 6. Sınıf seviye belirleme sınavı soru kitapçığı.

MEB (2010b). 7. Sınıf seviye belirleme sınavı soru kitapçığı.

MEB (2010c). 8. Sınıf seviye belirleme sınavı soru kitapçığı.

Meng, H (2007). A comparison study of IRT calibration methods for mixed-format tests in vertical scaling. Unpublished Ph.D. Thesis, University of Iowa, Iowa.

McBride, J.& Wise, L. (2001) Developing the Vertical Scale for the Florida Comprehensive Assessment Test (FCAT). A Harcourt Educational Measurement, San Antonio, Texas

Polpun, M., Omar, H. & Custe, M. (2004). A comparison of WINSTEPS and BILOG-MG for vertical scaling with the Rasch model. *Educational and Psychological Measurement* , 64(4), 600-616

Rupp, A. A. ( 2003). Item Response Modeling With BILOG-MG and MULTILOG for Windows. *International Journal of Testing*, 3(4), 365–384

Sarıer, Y. (2010). Ortaöğretime Giriş Sınavları (OKS-SBS) ve PISA Sonuçları Işığında Eğitimde Fırsat Eşitliğinin Değerlendirilmesi. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi*, 11(3), 107-129.

Stocking, M. L. & Lord, F. M. (1983). Developing a Common Metric In Item Response Theory. *Applies Psychological Measurement*, 7(2), 201-210

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds), *Test Scoring* (Pp. 73-140). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Tong, T (2005). Comparison of Methodologies And Results in Vertical Scaling for Educational Achievements Tests. Yayınlanmamış Doktora tezi, University of Iowa, Iowa.
- Tong, Y. & Kolen, M. (2010) Scaling: An ITEMS Module. *Educational Measurement: Issues and Practice*, 29(4), 39-48
- Tong, Y. & Kolen (2007). Comparison of Methodologies and Results in Vertical Scaling for Educational Achievement Tests. *Applied Measurement in Education*, 20(2), 227-253
- Tong, Y. & Kolen, M. (2008). *Maintenance of Vertical Scales*. National Council on Measurement in Education: March 2008. New York City.
- Wang, & Jiao, H. (2009). Construct Equivalence Across Grades in Vertical Scale for K-12 Large-Scale Reading Assesment. *Educational and psychological Measurement*, 69(5), 760-777
- Wyse1, A. E.& Reckase, M. D. (2011). A Graphical Approach to Evaluating Equating Using Test Characteristic Curves. *Applied Psychological Measurement*, 35(3) 217–234.
- Williams, S. L., Pommerich, M. & Thissen, D. (1998). A Comparison of Developmental Scales Based on Thurstone Methods and Item Response Theory. *Journal of Educational Measurement*, 35( 2), 93-107.
- Yen, W. M. (1986). The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measurement*, 23, 299–325.

Yu, C. H. & Popp, S. E. (2005). Test Equating by Common Items and Common Subjects: Concepts and Applications. *Practical Assessment Research & Evaluation*, 10(4), 1-19

Zeng, L. (1996). An IRT Scale Transformation Method for Parameters Calibrated from Multiple Samples of Subjects. **ACT Research Report Series, 96-2.**

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). BILOG-MG 3 [computer program]. Chicago: Scientific Software Corporation.

## ÖZGEÇMİŞ

### Kişisel Bilgiler

Adı Soyadı : Ayşegül ALTUN

Doğum Yeri ve Tarihi : Çarşamba, 1980

### Eğitim Durumu

Lisans Öğrenimi : Hacettepe Üniversitesi İlköğretim Bölümü  
Matematik Eğitimi Anabilim Dalı

Yüksek Lisans Öğrenimi : ODTÜ Orta Öğretim Fen ve Matematik Alanları  
Eğitimi Bölümü, Matematik Eğitimi Anabilim Dalı

Bildiği Yabancı Diller : İngilizce

Bilimsel Faaliyetleri :

### İş Deneyimi

Stajlar :

Projeler :

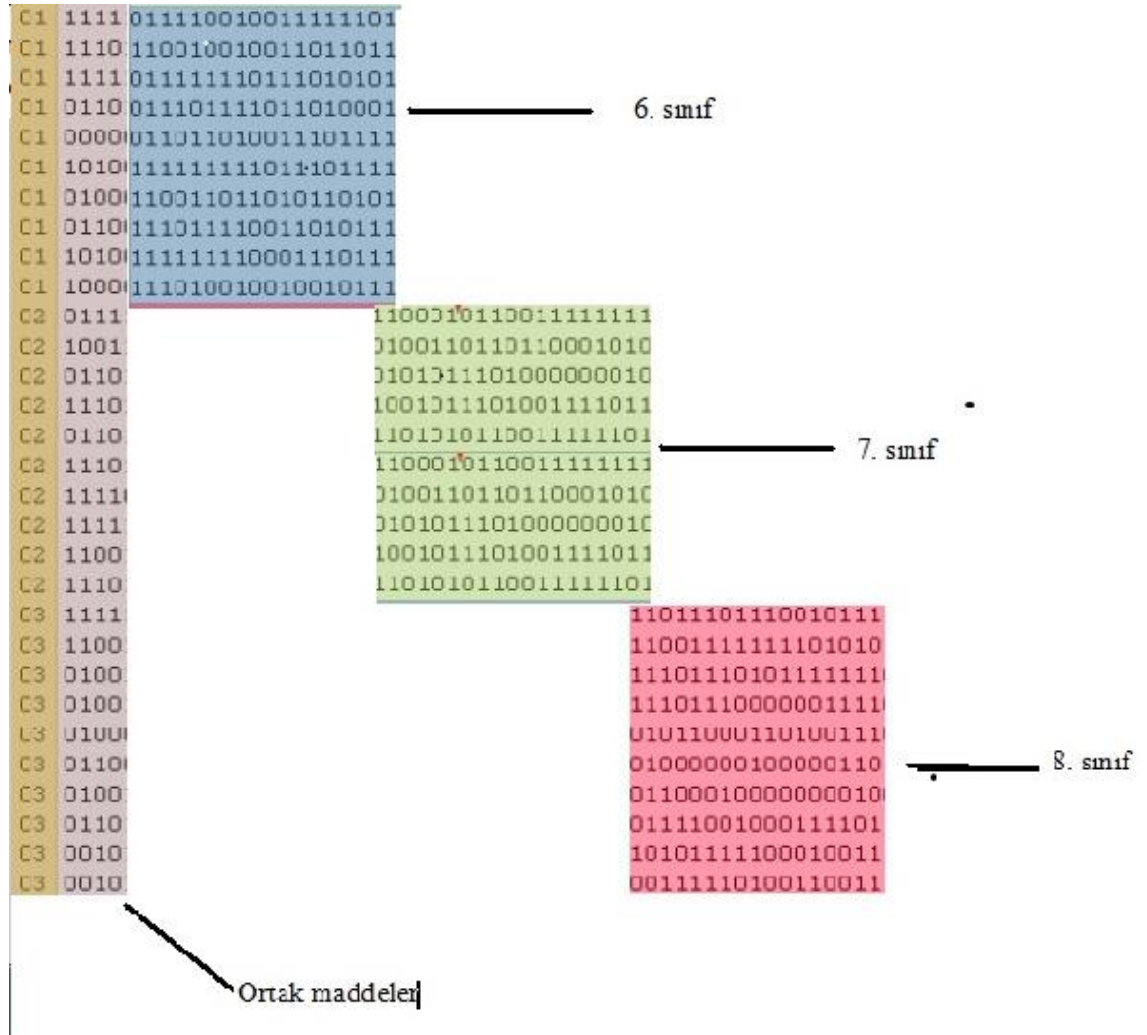
Çalıştığı Kurumlar :

### İletişim

E-Posta Adresi : aysegulaltn@gmail.com

Tarih : 04/07/2013

EK-1 : Eş zamanlı kalibrasyona ait veri giriş şeması



## EK-2 : 6. Sınıf Testi

### Soru 1:

Bir öğretmen ile 20 öğrencisi Topkapı Sarayı'nı gezmeye gidiyor. Sarayın giriş ücreti yandaki tabelada verilmiştir. Aşağıdaki işlemler-

GİRİŞ ÜCRETİ	
Öğretmen :	4 YTL
Öğrenci :	2 YTL

den hangisi öğretmen ve öğrencilerin ödeyeceği toplam para miktarını ifade eder?

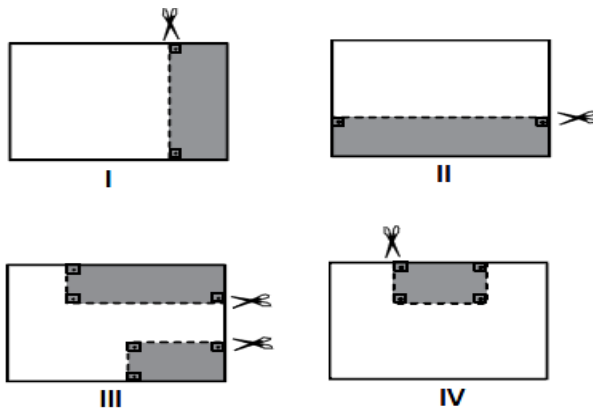
- A)  $4 + (2 \times 20)$       B)  $20 \times (2 + 4)$   
 C)  $4 \times (2 + 20)$       D)  $(4 \times 2) + 20$

### Soru 2:

Bir kitabın son iki sayfasının numaralarının toplamı 301'dir. Bu kitabın son sayfa numarası kaçtır?

- A) 150    B) 151    C) 300    D) 301

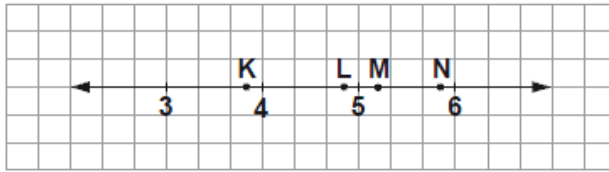
### Soru 3:



Eş dosya kâğıtları yukarıdaki gibi kesilerek boyalı parçalar atılıyor. Kalan parçanın çevre uzunluğu hangisinde en büyüktür?

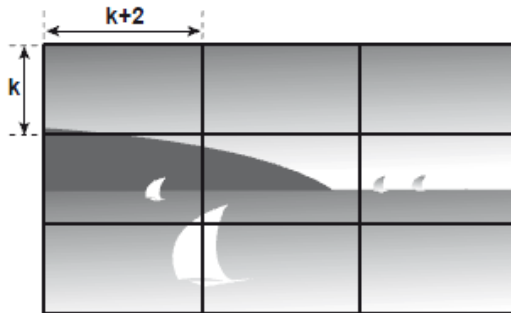
- A) I      B) II      C) III      D) IV



**Soru 4:**

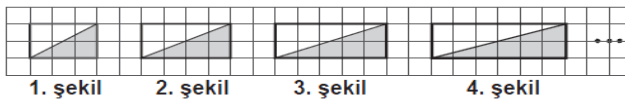
Yukarıdaki sayı doğrusunda,  $\frac{21}{4}$  sayısına karşılık gelen nokta aşağıdakilerden hangisi olabilir?

- A) K      B) L      C) M      D) N

**Soru 5:**

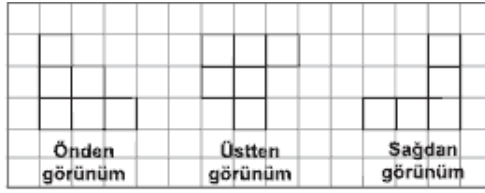
Şekildeki yapboz, kenar uzunlukları  $k$  santimetre ve  $(k + 2)$  santimetre olan eş dikdörtgen parçalardan oluşmuştur. Bu yapbozun çevresinin uzunluğunun kaç santimetre olduğunu veren cebirsel ifade aşağıdakilerden hangisidir?

- A)  $18k + 18$       B)  $12k + 12$   
C)  $9k + 9$       D)  $6k + 6$

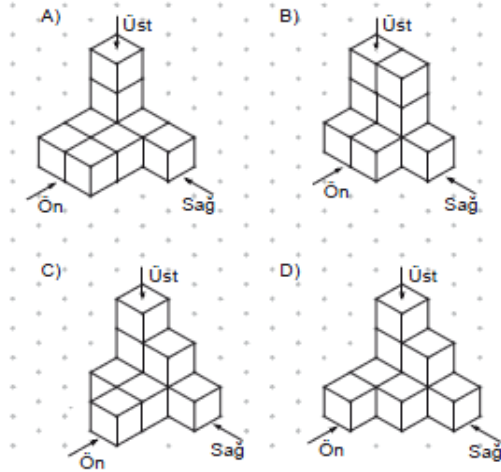
**Soru 6:**

Verilen örüntü aynı kurala göre devam ettirilirse, örüntünün 28. şeklindeki boyalı bölgenin alanı kaç birimkare olur?

- A) 26      B) 28      C) 30      D) 31

**Soru 7:**

Yukarıda farklı yönlerden görünümüleri verilen yapı, aşağıdakilerden hangisidir?

**Soru 8:**

Tanesi 4,50 TL'ye satın alınan tişörtlerin her birinin üzerine 1,25 TL'ye yazı yazdırılıyor. Bu tişörtlerin tanesi 9,50 TL'ye satıldığında,  $x$  tanesinden elde edilen kârı gösteren cebirsel ifade aşağıdakilerden hangisidir?

- A)  $(4,50 - 1,25)x - 9,50x$   
 B)  $9,50x - 4,50x + 1,25x$   
 C)  $(4,50 + 1,25)x - 9,50x$   
 D)  $9,50x - 4,50x - 1,25x$

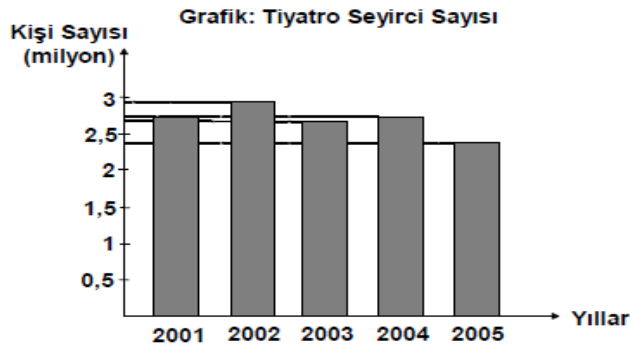
**Soru 9**

$2828 + 543 = 543 + \square$  ve  
 $\triangle \times (36 \times 3) = (28 \times 36) \times 3$  olduğuna göre,  
 $\square - \triangle$  işleminin sonucu kaçtır?

- A) 2600    B) 2792    C) 2800    D) 2856

**Soru 10:**

5. Türkiye İstatistik Kurumu verilerine göre ülkemizde 2001 – 2005 yılları arasındaki tiyatro seyirci sayıları aşağıdaki grafikte gösterilmektedir. Grafiğe göre, aşağıdaki ifadelerden hangisi yanlıştır?



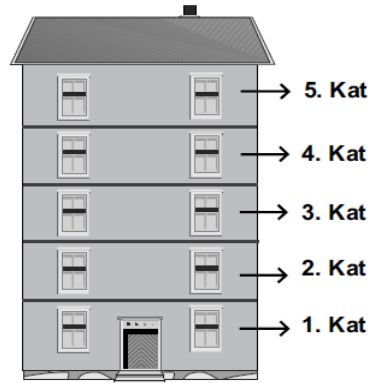
- A) 2004 yılındaki seyirci sayısı bir önceki yıla göre artmıştır.  
 B) 2002 yılındaki seyirci sayısı 2,5 milyondan fazladır.  
 C) Seyirci sayısının en fazla olduğu yıl 2002'dir.  
 D) 2003 yılındaki seyirci sayısı 2,5 milyondan azdır.

**Soru 11:**

Dengede olan bir terazinin bir kefesinde her biri 4 kg olan küp şeklinde iki cisim, diğer kefesinde ise her biri 3 kg olan küre şeklinde iki cisim ile piramit şeklinde bir cisim vardır. Buna göre, piramit şeklindeki cisim kaç kilogramdır?

- A) 4    B) 3    C) 2    D) 1

**Soru 12.**



Şekildeki 5 katlı kooperatif binasının her katında aynı sayıda daire vardır. Daireler kooperatif üyelerine çekiliş yapılarak dağıtılacaktır. Çekilen ilk dairenin üçüncü katta olma olasılığı nedir?

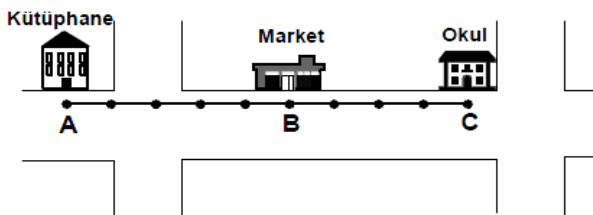
- A)  $\frac{1}{5}$       B)  $\frac{1}{3}$       C)  $\frac{3}{5}$       D)  $\frac{2}{3}$

**Soru 13:**

Bir ton kullanılmış kâğıt geri kazanıldığında 16 adet çam ağacının kesilmesi önlenmektedir. Ankara Çevre Koruma Vakfı, 2006 yılında 3700 ton kullanılmış kâğıt topladığına göre, kaç çam ağacının kesilmesi önlenmiştir?

- A) 231                                      B) 232  
C) 59 000                                    D) 59 200

**Soru 14**

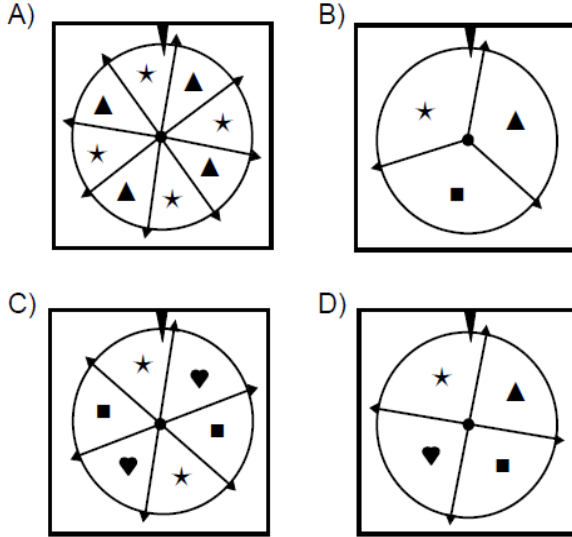


Kroki üzerinde çizilen doğru parçasında ardışık noktalar arasındaki uzaklıklar eşittir. Kütüphane, market ve okul sırasıyla A, B ve C noktalarında bulunmaktadır. Kütüphanenin markete uzaklığı 1,5 km olduğuna göre, market ile okul arası kaç kilometredir?

- A) 0,9      B) 1      C) 1,2      D) 1,4

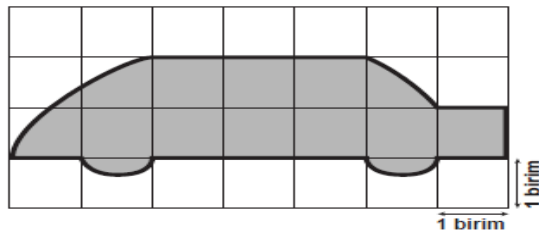
**Soru 15**

. Her biri eş dilimlere ayrılmış aşağıdaki çarklar döndürülüyor. Bu çarklar durduğunda ibrenin ★ olan dilimi gösterme olasılığı hangisinde en fazladır?

**Soru 16:**

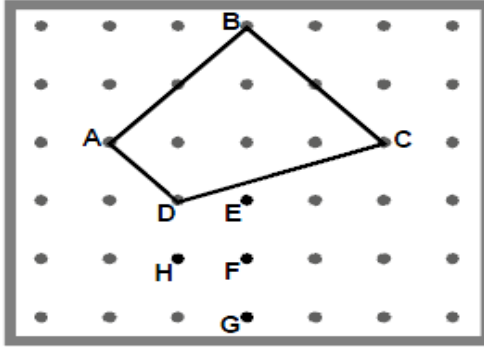
Aşağıdakilerden hangisi, birler basamağına yuvarlandığında 1, onda birler basamağına yuvarlandığında 0,5 olur?

- A) 0,45    B) 0,48    C) 0,53    D) 0,58

**Soru 17:**

Yukarıdaki şekilde boyalı alanın en yakın tahmini kaç birimkaredir?

- A) 10    B) 12    C) 14    D) 15

**Soru 18:**

Şekildeki geometri tahtasında bazı çivilerin bulunduğu noktalar harflerle adlandırılmıştır. Bu geometri tahtasına yerleştirilen şekildeki lastik, takılı olduğu D noktasındaki çividen çıkarılarak hangi noktadaki çiviye takılırsa elde edilen şekil bir düzgün çokgen olur?

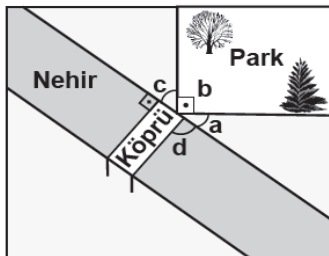
- A) E      B) F      C) G      D) H

**Soru 19:**

415\_76\_

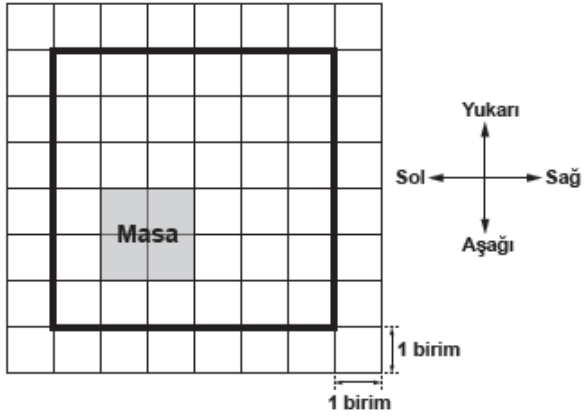
Gökçe yedi basamaklı olan yukarıdaki bilgisayar şifresinin bazı rakamlarını unutuyor. Bu şifrenin 6 ile bölünebildiğini ve rakamlarının birbirinden farklı olduğunu hatırladığına göre, Gökçe'nin bilgisayar şifresi aşağıdakilerden hangisi olabilir?

- A) 4158762      B) 4158763  
C) 4153761      D) 4152762

**Soru 20:**

Yukarıdaki planda verilenlere göre hangi açılar tümlerdir?

- A) a ve c      B) a ve b  
C) d ve c      D) b ve d

**Soru 21:**

**Şekilde bir oda ve içerisindeki masanın kroki verilmiştir. Masa, aşağıdakilerden hangisindeki gibi ötelenirse odanın ortasında yer alır?**

- |                                   |                                   |
|-----------------------------------|-----------------------------------|
| A) 2 birim yukarı<br>1 birim sağa | B) 1 birim yukarı<br>2 birim sağa |
| C) 2 birim yukarı<br>2 birim sağa | D) 1 birim yukarı<br>1 birim sağa |

### EK-3 : 7. Sınıf Testi

#### Soru 1:

Bir kitabın son iki sayfasının numaralarının toplamı 301'dir. Bu kitabın son sayfa numarası kaçtır?

- A) 150    B) 151    C) 300    D) 301

#### Soru 2:

Bir öğretmen ile 20 öğrencisi Topkapı Sarayı'nı gezmeye gidiyor. Sarayın giriş ücreti yandaki tabelada verilmiştir. Aşağıdaki işlemler-

GİRİŞ ÜCRETİ	
Öğretmen :	4 YTL
Öğrenci :	2 YTL

den hangisi öğretmen ve öğrencilerin ödeyeceği toplam para miktarını ifade eder?

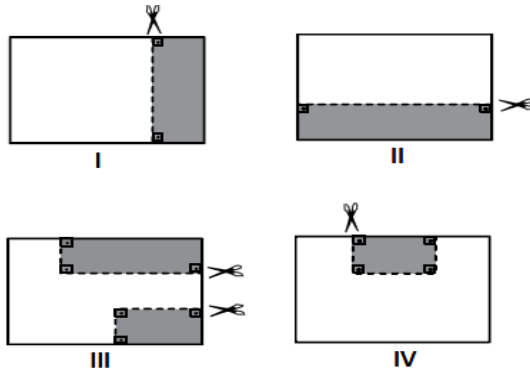
- A)  $4 + (2 \times 20)$                       B)  $20 \times (2 + 4)$   
 C)  $4 \times (2 + 20)$                       D)  $(4 \times 2) + 20$

#### Soru 3:

Bir çiftlikteki hayvanların  $\frac{3}{10}$ 'ü koyun,  $\frac{1}{8}$ 'i inek, geriye kalanı tavuktur. Bu çiftlikteki hayvanların sayısının 150 ile 200 arasında olduğu bilindiğine göre, çiftlikte kaç tavuk vardır?

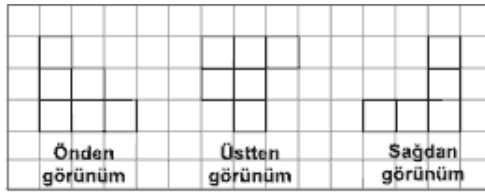
- A) 68    B) 80    C) 92    D) 138



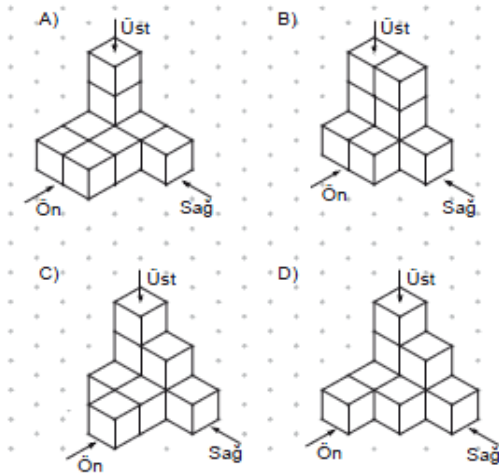
**Soru 4:**

Eş dosya kâğıtları yukarıdaki gibi kesilerek boyalı parçalar atılıyor. Kalan parçanın çevre uzunluğu hangisinde en büyüktür?

- A) I      B) II      C) III      D) IV

**Soru 5:**

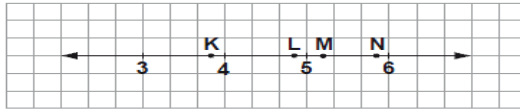
Yukarıda farklı yönlerden görünümü verilen yapı, aşağıdakilerden hangisidir?



**Soru 6:**

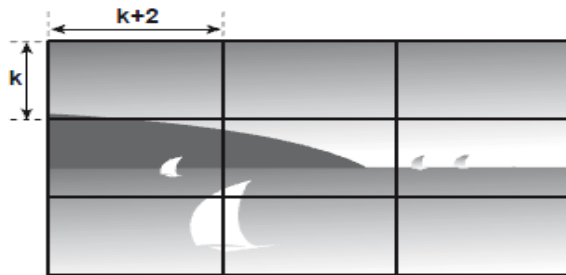
Yukarıdaki haritada, Karadeniz Bölgesi'nin doğu, orta ve batı bölümleri gösterilmiştir. Bölümlerden biri mavi, biri turuncu, biri yeşil olmak üzere harita kaç farklı şekilde boyanabilir?

- A) 2      B) 3      C) 4      D) 6

**Soru 7:**

Yukarıdaki sayı doğrusunda,  $\frac{21}{4}$  sayısına karşılık gelen nokta aşağıdakilerden hangisi olabilir?

- A) K      B) L      C) M      D) N

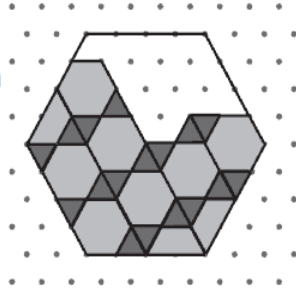
**Soru 8:**

Şekildeki yapboz, kenar uzunlukları  $k$  santimetre ve  $(k + 2)$  santimetre olan eş dikdörtgensel parçalardan oluşmuştur. Bu yapbozun çevresinin uzunluğunun kaç santimetre olduğunu veren cebirsel ifade aşağıdakilerden hangisidir?

- A)  $18k + 18$       B)  $12k + 12$   
C)  $9k + 9$       D)  $6k + 6$

**Soru 9:**

Şekilde verilen altıgensel bölge içindeki süslemenin tamamlanması için üçgensel bölgelerden kaç tane daha gerekir?



- A) 7      B) 6      C) 5      D) 4

**Soru 10:**

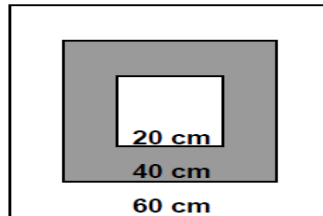
$$\blacksquare \times \blacktriangle + \bullet$$

-8, +3, -2 sayıları yukarıdaki semboller yerine hangisindeki gibi yerleştirilirse elde edilen işlemin sonucu en büyük olur?

- |    | $\blacksquare$ | $\blacktriangle$ | $\bullet$ |
|----|----------------|------------------|-----------|
| A) | -8             | +3               | -2        |
| B) | +3             | -8               | -2        |
| C) | -2             | +3               | -8        |
| D) | -8             | -2               | +3        |

**Soru11:**

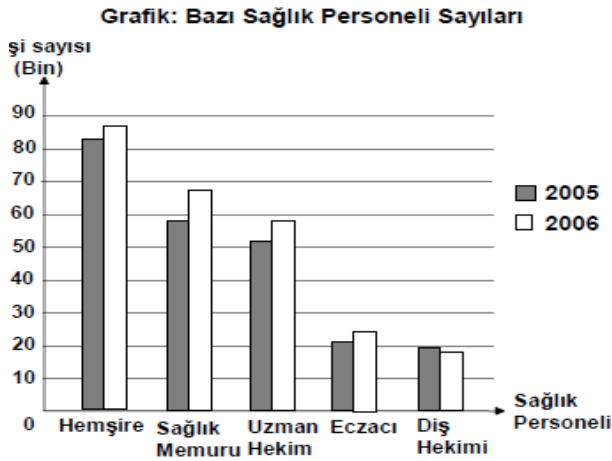
Bir kenar uzunluğu 60 cm olan karesel bölge şeklindeki hedef tahtası iç içe iki kare ile bölgelere ayrılmıştır. Bu karelerin kenar uzunlukları 20 cm ve 40 cm olduğuna göre, hedef tahtasına isabet eden bir okun boyalı bölgeye isabet etme olasılığı nedir?



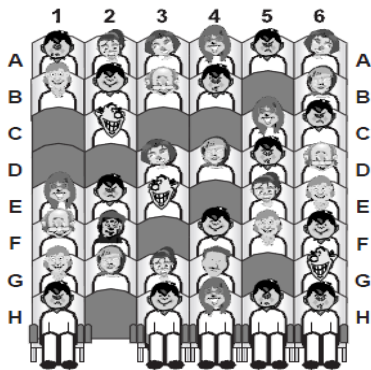
- A)  $\frac{1}{3}$       B)  $\frac{4}{9}$       C)  $\frac{5}{9}$       D)  $\frac{2}{3}$

**Soru12:**

- . Türkiye İstatistik Kurumunun verilerine göre ülkemizde 2005 – 2006 yıllarındaki bazı sağlık personeli sayıları aşağıdaki grafikte verilmiştir. Grafığe göre, aşağıdaki ifadelerden hangisi **yanlıştır**?



- A) En az değişim eczacı sayısında olmuştur.  
 B) Hemşire sayısı, sağlık memuru sayısına göre daha az artmıştır.  
 C) En fazla değişim sağlık memuru sayısında olmuştur.  
 D) Sadece diş hekimlerinin sayısı azalmıştır.

**Soru 13:**

Şekilde bir konferans salonundaki boş ve dolu koltukların yerleri belirtilmiştir. Bu salona sonradan gelen bir kişinin C sırasında veya çift numaralı bir koltuğa oturma olasılığı nedir?

- A)  $\frac{1}{10}$       B)  $\frac{1}{8}$       C)  $\frac{3}{10}$       D)  $\frac{3}{5}$

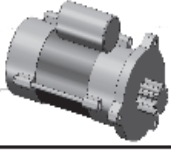


**Soru 14:**

16 üyesi bulunan Sağlıklı Yaşam Derneğine haftada 2 üye, 4 üyesi bulunan Kitap Sevenler Derneğine ise haftada 6 üye kaydedilmektedir. Bu iki derneğin üye sayıları kaç hafta sonra eşit olur?

- A) 2      B) 3      C) 6      D) 8

**Soru 15:**

Tablo: Bir Makinenin Farklı Modellerinin Özellikleri

MODELLER			
ÖZELLİKLER			
Boy (cm)	60	30	20
Maliyet (bin TL)	1	2	3
Kütle (kg)	6	3	2
Dayanıklılık (yıl)	2	4	6

Bir fabrikada üretilen bir makinenin farklı modellerine ait veriler yukarıdaki tabloda gösterilmektedir. Buna göre, aşağıdakilerden hangisindeki özellikler doğru orantılıdır?

- A) boy ve maliyet  
B) boy ve dayanıklılık  
C) maliyet ve dayanıklılık  
D) dayanıklılık ve kütle

**Soru 16:**

Aşağıdaki doğrusal denklemlerden hangisi, yandaki tabloda verilen x ve y değerleri arasındaki ilişkiyi açıklar?

x	y
3	10
4	12
5	14
6	16

- A)  $y = 2x + 4$       B)  $y = 3x + 1$   
C)  $y = x + 7$       D)  $y = 3x - 2$

**Soru 17:**

$(-7) + 5 + (-3)$  işleminin yapılırken, aşağıdakilerden hangisinde hata yapılmamıştır?

A)  $(-7)+5+(-3) = (-7)+(-2) = -9$

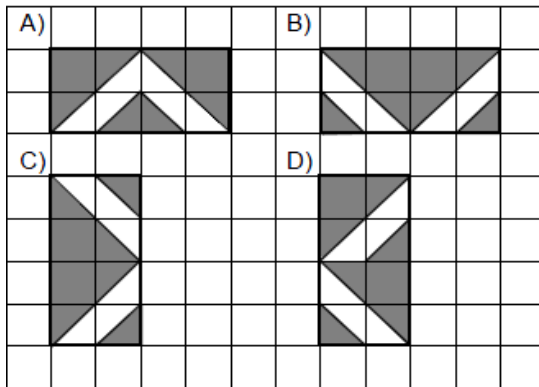
B)  $(-7)+5+(-3) = (-12)+(-3) = -9$

C)  $(-7)+5+(-3) = (-7)+(-2) = -5$

D)  $(-7)+5+(-3) = (-2)+(-3) = -5$

**Soru 18:**

Kareli kâğıt üzerinde verilen şekillerden hangisinde yansıma simetrisi yoktur?



**Soru 19:**

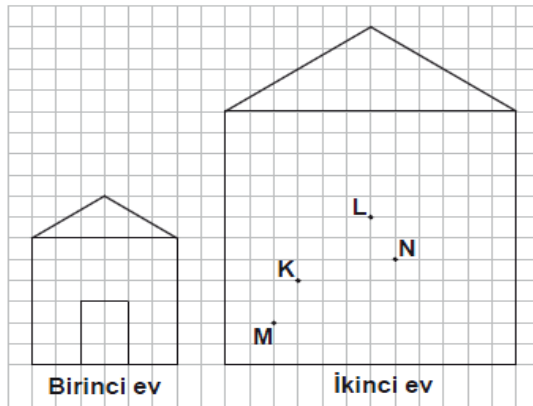
Yandaki saatin farklı mağazalardaki etiket fiyatları ve yapılacak indirim yüzdeleri aşağıda verilmiştir. Yapılacak indirimler sonrasında bu saat hangi mağazadan en ucuza satın alınır?



- A) Sevgi Mağazası  
100 YTL  
%25 indirim
- B) Dost Mağazası  
90 YTL  
%10 indirim
- C) Yıldız Mağazası  
90 YTL  
%20 indirim
- D) Gül Mağazası  
100 YTL  
%30 indirim

**Soru 20:**

Selen, birinci eve benzer olacak şekilde ikinci evi çiziyor. İkinci evin kapısını çizdiğinde, hangi nokta kapının içerisinde yer alır?



- A) K      B) L      C) M      D) N

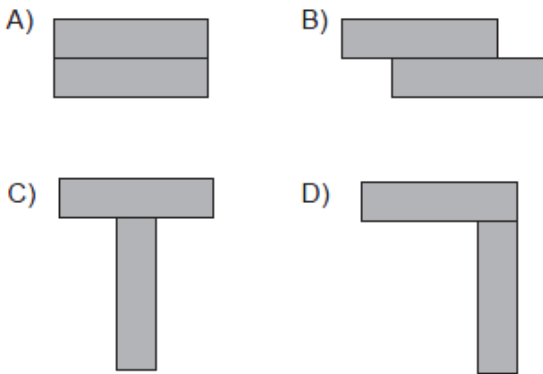
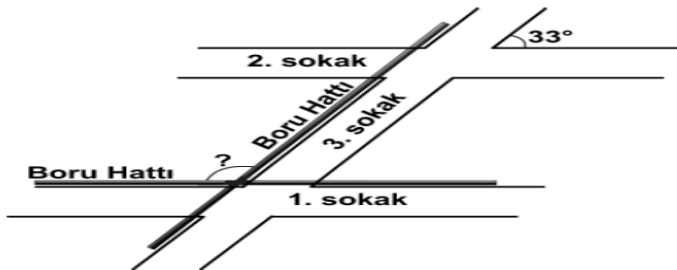
**Soru 21:**

Bir arařtırmaya gre Trkiye'de bir gnde retilen 123 000 000 ekmeęin % 5'i pe atılarak israf ediliyor. 3 gnlk ekmeę israfı ile bir okul yaptırılabilir. Bir ekmeę 50 Kr olduęuna gre bir okulun maliyeti ka TL'dir?

- A) 3 075 000                      B) 6 150 000  
C) 9 225 000                      D) 18 450 000

**Soru 22:**

İki eř dikdrtgensel blge, ařaęıdakilerden hangisindeki gibi birleřtirilirse oluřan Őeklin evresinin uzunluęu en az olur?

**Soru 23:**

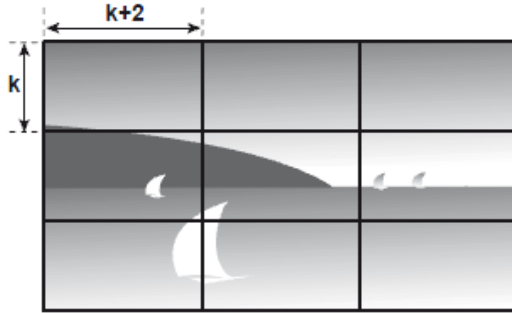
Bir mahalledeki 3. sokak, birbirine paralel olan 1. ve 2. sokakları kesmektedir. Bu mahalleye, 1. ve 3. sokaklara paralel olacak Őekilde krokideki gibi borular dřenecektir. 2. ve 3. sokakların oluřturduęu dar aı 33° olduęuna gre, boru hatlarının oluřturduęu geniř aı ka derecedir?

- A) 114                      B) 123                      C) 147                      D) 153



## EK-4 : 8. Sınıf Testi

### Soru 1:



Şekildeki yapboz, kenar uzunlukları  $k$  santimetre ve  $(k + 2)$  santimetre olan eş dikdörtgen parçalardan oluşmuştur. Bu yapbozun çevresinin uzunluğunun kaç santimetre olduğunu veren cebirsel ifade aşağıdakilerden hangisidir?

- A)  $18k + 18$                       B)  $12k + 12$   
C)  $9k + 9$                               D)  $6k + 6$

### Soru 2:

Bir öğretmen ile 20 öğrencisi Topkapı Sarayı'nı gezmeye gidiyor. Sarayın giriş ücreti yandaki tabelada verilmiştir. Aşağıdaki işlemler-

GİRİŞ ÜCRETİ	
Öğretmen :	4 YTL
Öğrenci :	2 YTL

den hangisi öğretmen ve öğrencilerin ödeyeceği toplam para miktarını ifade eder?

- A)  $4 + (2 \times 20)$                       B)  $20 \times (2 + 4)$   
C)  $4 \times (2 + 20)$                       D)  $(4 \times 2) + 20$

**Soru 3:**

Bir kitabın son iki sayfasının numaralarının toplamı 301'dir. Bu kitabın son sayfa numarası kaçtır?

- A) 150    B) 151    C) 300    D) 301

**Soru 4:**

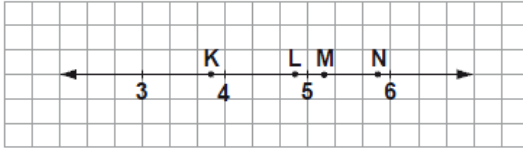
Bir çiftlikteki hayvanların  $\frac{3}{10}$ 'ü koyun,  $\frac{1}{8}$ 'i inek, geriye kalanı tavuktur. Bu çiftlikteki hayvanların sayısının 150 ile 200 arasında olduğu bilindiğine göre, çiftlikte kaç tavuk vardır?

- A) 68    B) 80    C) 92    D) 138

**Soru 5:**

Yukarıdaki haritada, Karadeniz Bölgesi'nin doğu, orta ve batı bölümleri gösterilmiştir. Bölümlerden biri mavi, biri turuncu, biri yeşil olmak üzere harita kaç farklı şekilde boyanabilir?

- A) 2    B) 3    C) 4    D) 6

**Soru 6:**

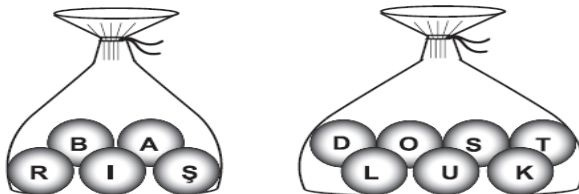
Yukarıdaki sayı doğrusunda,  $\frac{21}{4}$  sayısına karşılık gelen nokta aşağıdakilerden hangisi olabilir?

- A) K      B) L      C) M      D) N

**Soru 7:**

$\left(\frac{1}{4} + \frac{1}{2}\right) - \left(\frac{1}{4} - \frac{1}{2}\right)$  işleminin sonucu kaçtır?

- A) 0      B)  $\frac{1}{4}$       C)  $\frac{1}{2}$       D) 1

**Soru 8:**

“BARIŞ” ve “DOSTLUK” sözcüklerini oluşturan harfler, eş topların üzerlerine yazılarak, şekildeki gibi torbalara atılıyor. Torbalardan rastgele birer top çekildiğinde, üzerlerindeki harflerin R ve K olma olasılığı nedir?

- A)  $\frac{32}{35}$       B)  $\frac{12}{35}$       C)  $\frac{2}{35}$       D)  $\frac{1}{35}$

**Soru 9:**

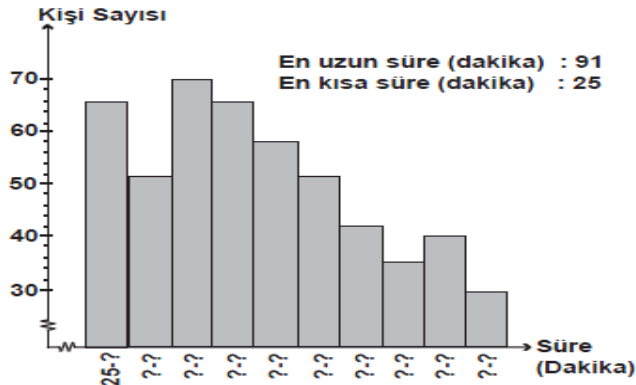
Bir torbada, renkleri dışında aynı özelliklere sahip siyah, beyaz ve mavi toplar vardır. Bu torbadan rastgele çekilen bir topun siyah olma olasılığı  $\frac{1}{3}$ , mavi olma olasılığı  $\frac{2}{9}$ 'dur. Torbada 36 tane top olduğuna göre, bu toplardan kaç tanesi beyazdır?

- A) 8      B) 12      C) 14      D) 16

**Soru 10:**

Bir pastanede bir günde satılan börek ve poğaçaların toplam sayısı 144 ve bu satıştan elde edilen gelir 88 TL'dir. Börek 50 Kr, poğaça 75 Kr olduğuna göre kaç adet poğaça satılmıştır?

- A) 80      B) 64      C) 58      D) 44

**Soru 11:**

Bir okuldaki öğrencilerin günlük kitap okuma süreleriyle ilgili bazı bilgileri içeren histogram yukarıda verilmiştir. Grafiğe göre, kişi sayısı **en fazla** olan grubun günlük okuma süresi hangi dakika aralığındadır?

- A) 39 - 45      B) 41 - 46  
C) 34 - 41      D) 41 - 49

**Soru 12:**

Aşağıdakilerden hangisi 64 doğal sayısının üslü sayı olarak yazılışlarından biri değildir?

- A)  $2^6$       B)  $4^3$       C)  $6^3$       D)  $8^2$

**Soru 13:**

$$1 \times 1 = 1$$

$$11 \times 11 = 121$$

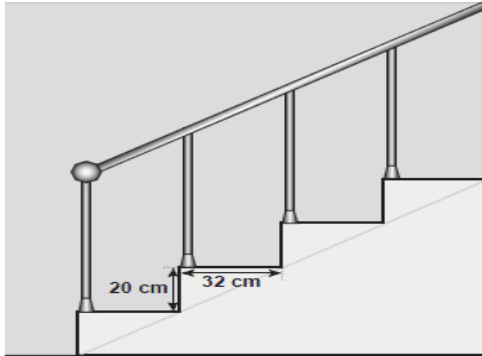
$$111 \times 111 = 12321$$

$$1111 \times 1111 = 1234321$$

$$\vdots \quad \quad \quad \vdots$$

Yukarıdaki örüntüye göre,  
 $11111111 \times 11111111$  işleminin sonucu kaç basamaklı bir sayıdır?

- A) 13      B) 14      C) 15      D) 16

**Soru 14:**

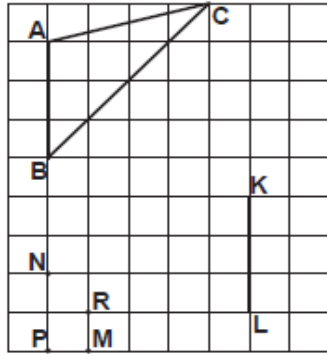
Şekildeki merdivenin basamaklarının yüksekliği 20 cm, derinliği 32 cm'dir. Bu merdivenin eğimi kaçtır?

- A)  $\frac{5}{8}$       B)  $\frac{5}{6}$       C)  $\frac{6}{5}$       D)  $\frac{8}{5}$

**Soru 15:**

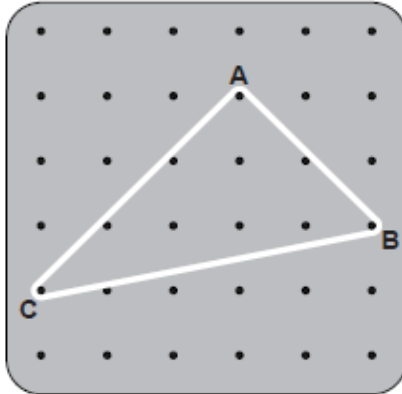
Alanı  $39 \text{ m}^2$  olan kare şeklindeki bir bahçenin bir kenar uzunluğu, hangi metreler arasındadır?

- A) 4 - 5    B) 5 - 6    C) 6 - 7    D) 7 - 8

**Soru 16:**

Şekilde adlandırılan noktalardan hangisi, KL doğru parçasının uç noktaları ile birleştirilirse ABC üçgenine eş bir üçgen elde edilir?

- A) M    B) N    C) P    D) R

**Soru 17:**

Yukarıdaki geometri tahtasında, bir lastik A, B ve C noktalarındaki çivilere takılarak üçgen oluşturulmuştur. Bu üçgenin iç açılarının ölçülerinin büyükten küçüğe doğru sıralanışı aşağıdakilerden hangisinde verilmiştir?

- A)  $m(\widehat{A}) > m(\widehat{C}) > m(\widehat{B})$
- B)  $m(\widehat{C}) > m(\widehat{A}) > m(\widehat{B})$
- C)  $m(\widehat{B}) > m(\widehat{A}) > m(\widehat{C})$
- D)  $m(\widehat{A}) > m(\widehat{B}) > m(\widehat{C})$

**Soru 18:**

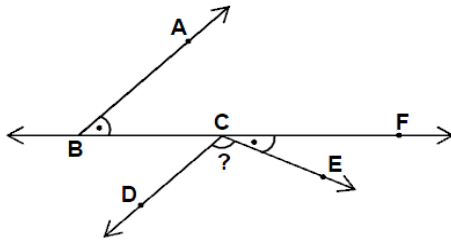
Bir sınıftaki kızların sayısının, erkeklerin sayısına oranı  $\frac{4}{5}$ 'tir. Bu sınıfın mevcudu aşağıdakilerden hangisi olamaz?

- A) 18
- B) 24
- C) 27
- D) 36

**Soru 19:**

Bir çiftlikteki hayvanların  $\frac{3}{10}$ 'ü koyun,  $\frac{1}{8}$ 'i inek, geriye kalanı tavuktur. Bu çiftlikteki hayvanların sayısının 150 ile 200 arasında olduğu bilindiğine göre, çiftlikte kaç tavuk vardır?

- A) 68      B) 80      C) 92      D) 138

**Soru 20:**

Şekilde  $[BA \parallel [CD$  'dir.  $s(\hat{A}BC) = s(\hat{F}CE)$  ve  $s(\hat{D}CE) = 3.s(\hat{A}BC)$  olduğuna göre,  $s(\hat{D}CE)$  kaç derecedir?

- A) 72      B) 90      C) 108      D) 126

**Soru 21:**

Aşağıdaki geometrik şeritler, uçlarından tutturularak üçgenler oluşturulacaktır. Hangi şeritlerle oluşturulacak üçgenin çevresi **en büyük** olur?

K

L

M

N

P

- A) K, N ve P      B) K, L ve M  
C) L, N ve P      D) L, M ve N



## EK-5 : 6-8 Eş Zamanlı Kalibrasyona Ait Örnek BILOG-MG Komutları

```

>GLOBAL DFName = 'C:\6-8.DAT',
      NParm = 2,
      SAVE;
>SAVE CALib = 'EXAMPL01.CAL',
      PARM = 'EXAMPL01.PAR',
      SCORE = 'EXAMPL01.SCO',
      TStat = 'EXAMPL01.TST',
      EXPECTed = 'EXAMPL01.EXP',
      IStat = 'EXAMPL01.IST';
>LENGTH NITems = (53);
>INPUT NTotal = 53,
      NALT = 2,
      NIDchar = 4,
      NGroup = 3;
>ITEMS ;
>TEST1 TName = 'TEST0001',
      INumber = (1(1)53);
>GROUP1 GName = 'GROUP001',
      LENGTH = 21,
      INumbers = (1(1)21);
>GROUP2 GName = 'GROUP002',
      LENGTH = 23,
      INumbers = (1(1)4, 20(1)38);
>GROUP3 GName = 'GROUP003',
      LENGTH = 21,
      INumbers = (1(1)4, 37(1)53);
(4A1, 1X, I1, 1X, 53A1)
>CALIB NQPt = 15,
      REFerence = 3;
>SCORE METHod = 3,
      RSctype = 3;

```