ÖNSÖZ

15/11/2015 tarihinde başlayan, TÜBİTAK SOBAG 3501 programı tarafından desteklenen 115K531 No.lu "Uluslararası Geniş Ölçekli Sınavlarda Türkiye'nin Matematik Başarısını Arttırabilmek İçin Bir Model Önerisi: Bilişsel Tanıya Dayalı İzleme Modelinin Etkililiği" isimli proje, sonuç raporu ile tamamlanmış bulunmaktadır. Proje, Ege Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü tarafından yürütülmüştür.

Proje genel hatları ile planlandığı gibi yürütülmüş ve sonuçlandırılmıştır. Bununla birlikte süreç içinde temel araştırma probleminden sapmadan yeni sorular ortaya çıkmış ve bunların yanıtları aranmıştır. Bu yönüyle de proje oldukça dinamik, yaratıcı ve bilgilendirici bir araştırma sürecinin yaşama geçirilmesini sağlamıştır. Rapor içinde detaylı olarak aktarılacak bulgular sadece ülkemiz için değil genel olarak uluslararası nitelikte bir etki yaratma potansiyeli taşımaktadır. Bu noktada en sevindirici olan konu, henüz sonuçlanmadan bile eğitim sistemi içinde projenin işaret ettiği sorunların önem düzeyinin giderek arttığının gözlenmesidir. Henüz ortada çok örneği yokken proje tarafından gerçekleştirilen "üst düzey düşünme becerilerine yönelik soru yazma eğitimi" son iki yılda gerek MEB, gerekse özel kurumlar tarafından giderek yaygınlaşmıştır. Aynı zamanda projenin temel sorunlarından biri olan Bilişsel Tanı Modellerinin bu süreç içerisinde MEB uygulamalarında gündeme alındığı görülmektedir. Ülkemizin PISA başarısını arttırmaya yönelik ölçme değerlendirme etkinliklerine yönelik ilgi ise giderek artmaktadır. Bütün bu gelişmeler projenin hazırlık aşamasından itibaren eğitim sistemimiz açısından önemli sorunları dikkate aldığının en önemli göstergesidir.

Çalışmanın sonuçları göz önüne alındığında eğitim sistemi açısından sadece öğrencilerin PISA ve TIMSS gibi sınavlarda başarılarını arttırmak değil, aynı zamanda önemi giderek anlaşılan üst düzey düşünme becerilerinin geliştirilmesi açısından da etkin olarak kullanılabilecek bir yöntemi sunduğu görülmektedir. Rapor kısmında da ayrıntılı olarak açıklandığı gibi araştırma bulgularının iç ve dış geçerliğinin sağlanması ve sonuçların genellenebilirliği konusunda bütün süreçlerde oldukça hassas davranılmıştır. Bu yönden de araştırma sonuçlarının deneysel süreçle elde edilmiş birer laboratuvar çıktısı olmadığı, gerçek yaşam için etkili olabilecek yöntem ve yaklaşımlara işaret ettiği de söylenebilir.

Araştırma süresince Türkiye genelinde çok sayıda öğretmen, idareci ve MEB personeli ile etkileşim içinde bulunulmuştur. Aynı zamanda Ege Üniversitesi Eğitim Fakültesi öğrencileri de öncelikle toplam 24000 testin dijital ortama geçirilmesi ve bir çok diğer konuda proje ekibine yardımcı olmuşlardır. Gerek soru yazarı, gerek hakem, gerekse puanlayıcı olarak projeye destek olan bu değerli insanlara genel bir teşekkürün yeterli olmayacağı düşünüldüğünden emeklerinin hatırlanması için rapor sonunda isimleri eklenmiştir.

Aynı zamanda İzmir İl Milli Eğitim personeli, Ankara MEB Ölçme Değerlendirme Genel Müdürlüğü ve İzmir İli Aliağa, Balçova, Bayraklı, Bornova, Buca, Çiğli, Gaziemir, Karabağlar, Karşıyaka, Konak İlçe Milli Eğitim Müdürlüklerine desteklerinden dolayı teşekkür ederiz. Bununla birlikte kendilerine sunduğumuz bütün kitapçıkların içindeki soruları çoğunlukla büyük bir istek ve samimiyetle cevaplayan değerli öğrencilerimize de teşekkür ederiz.

Son olarak bu çalışmanın gerçekleşmesini sağlayan TÜBİTAK'a desteklerinden dolayı teşekkürü bir borç biliriz.

İÇİNDEKİLER

1 GİRİŞ 1

1.1 Amaç ve Önem 3

2 ALANYAZIN TARAMASI 4

2.1 Bilişsel Tanı Modelleri 5
2.1.1 Q Matris 5

2.2 Proje Kapsamında Kullanılacak Bilişsel Tanı Modelleri 6
2.2.1 DINA Model 6
2.2.2 G-DINA Model 7
2.2.3 Fusion Model 8

2.3 Bilişsel Tanı Modellerinde Örtük Sınıflar ve Yetenek Kestirimi 9

2.4 Yapılandırılmış Yanıtlı Madde Formatı 10

2.5 Üst Düzey Düşünme Becerileri 12
2.5.1 Üst Düzey Düşünme Becerilerin Ölçülmesi ve Değerlendirilmesi 13
2.5.2 Matematikte Üst Düzey Düşünme ve Gerçek Hayat 14
2.5.3 Açık Uçlu Sorularda Üst Düzey Düşünme Becerisi Ölçümleri 15

2.6 Üstbilişsel İzleme 17
2.6.1 Sinyal Tepki Kuramı ve Üst Bilişsel İzleme Ölçümleri 18

2.7 PISA Testi 22
2.7.1 Pisa Örneklemi 22
2.7.2 Pisa'da yetenek kestirimi RASCH 23
2.7.3 Rasch Modelde öğrenci puanının hesaplanması 25
2.7.4 Olası (Plausible) Değerler 25
2.7.5 PISA Yeterlik Düzeyleri 26

3 GEREÇ ve YÖNTEM 28

3.1 Test Geliştirme Süreci 28
3.1.1 Hazırlık aşaması 28
3.1.2 Matematiksel yeterlik matrislerinin belirlenmesi 29
3.1.2.1 PISA'da Matematiksel Yeterlikler 29
3.1.2.2 Araştırmada Matematiksel Yeterliklerin belirlenmesi 31
3.1.2.3 Matematiksel Hata Matrislerinin Belirlenmesi 32
3.1.3 Soru yazama ekibinin kurulması 33
3.1.4 Madde havuzu oluşturma süreci 33
3.1.5 Soru değerlendirme süreci 34

3.2 Deneme Uygulaması ve Ön-Son Testin Hazırlanması 35
3.2.1 Deneme uygulamasının puanlanması 36
3.2.2 Proje soruları ile PISA, TIMSS testlerinin eşdeğerliği çalışması 36
3.2.3 Deneme uygulaması yapılan grupların denkliklerinin araştırılması 40
3.2.4 Proje Sorularının Psikometrik Özelliklerinin İncelenmesi 42
3.2.5 Madde havuzunun değerlendirilmesi 45
3.2.6 Ön-test sınavının belirlenmesi 46

3.3 Araştırma Örneklemi ve Deney-Kontrol Grupları 46

3.3.1 Araştırma Örnekleminin Belirlenmesi 46
3.3.2 Deney ve Kontrol Gruplarının Belirlenmesi 48

3.4 Deneysel İşlem Süreci 49
3.4.1 Ön-test uygulaması 50
3.4.2 Ön-test Psikometrik Özellikleri 52
3.4.3 Geri bildirim sistemi 53
3.4.4 İzleme Testleri 53
3.4.4.1 İzleme-1 Uygulaması 54
3.4.4.2 İzleme-2 Uygulaması 55
3.4.4.3 İzleme-3 Uygulaması 57
3.4.4.4 İzleme-4 Uygulaması 59
3.4.5 Üstbilişsel İzleme Ölçümleri Uygulaması 61
3.4.6 Son Test Uygulaması 63
3.4.7 Son-test Psikometrik Özellikleri 64

4 BULGULAR 66

4.1 Deneysel Sürece İlişkin Bulgular 66
4.1.1 Ön-Son Testler için Varyans Analizleri 66
4.1.2 Ön-Son Test Uygulamaları Arasında Tekrarlı Ölçümler Analizi 67
4.1.3 Deneysel Sürece ilişkin Kovaryans Analizleri 69
4.1.4 Deneysel Sürecin Kalıcılığına İlişkin Tekrarlı Ölçümler Varyans Analizi 69

4.2 Öğrenci Yetenek Düzeyleri ve BTM Analizlerine İlişkin Bulgular 71
4.2.1 MTK Analizleri ve Öğrenci Yetenek Düzeylerindeki Değişim 72
4.2.1.1 Matematiksel Yeterlik Düzeylerindeki Değişimlere İlişkin Bulgular 73
4.2.2 BTM Analizleri ve Öğrenci Örtük Sınıflarındaki Değişim 75

4.3 Araştırma Sürecinde Ulaşılan Diğer Bulgular 77
4.3.1 Üst Bilişsel İzleme Ölçümlerine İlişkin Bulgular 77
4.3.1.1 BTM Örtük Sınıflarına Göre Üstbilişsel İzleme Performansındaki Değişim 77
4.3.1.2 Öğrenci Başarı Düzeylerine Göre Üstbilişsel İzleme Performansındaki Değişim 78
4.3.2 Öğrencilerin Test Performans Farkındalıklarına İlişkin Bulgular 79
4.3.3 Açık Uçlu Sorulara İlişkin Bulgular 82
4.3.4 Uygulanan Testlere Yönelik Tutumlar ile İlgili Bulgular 85

5 SONUÇ ve TARTIŞMA 92

5.1 Deneysel Süreç Sonuçları 92

5.2 Öğrenci Yetenek Düzeyleri ve BTM Analizleri ile İlgili Sonuçlar 93

5.3 Proje Sürecinde Ulaşılan Diğer Sonuçlar 94

5.4 Öneriler 96

6 KAYNAKÇA 98

TABLOLAR LİSTESİ

Tablo 1. Örnek Q Matris 5
Tablo 2. Proje Matematiksel Yeterlik Kodları ve Tanımları 31
Tablo 3. Proje Hata Kodları ve Tanımları 32
Tablo 4. Pilot Uygulama 00 Deseni 35
Tablo 5. PISA 2012 Matematiksel Yeterliklerin Dağılımı 37
Tablo 6.PISA 2012 Matematik Puanları Ortalama ve Standart Sapmaları 38
Tablo 7.PISA ve Proje ölçümleri için yüzdelik düzeylerin kesme puanlarının karşılaştırılması 38
Tablo 8. PISA ve Proje ölçümleri için matematiksel yeterlik seviyeleri yüzdelik dağılımları ... 39
Tablo 9. Deneme Uygulaması Gruplarının Denkliklerine İlişkin ANOVA Sonuçları 41
Tablo 10. Deneme Uygulaması Ankor Maddelerine İlişkin Madde Parametreleri 41
Tablo 11. Deneme Uygulaması Maddelerine İlişkin KTK Parametreleri 43
Tablo 12. Deneme Uygulaması Maddelerine İlişkin MTK Parametreleri 44
Tablo 13. Deneme Uygulaması Maddelerine İlişkin BTM Parametreleri 45
Tablo 14. Araştırma Örnekleminde Yer Alan Okullar 47
Tablo 15. Araştırma Örneklemi Deney ve Kontrol Grupları 48
Tablo 16. Deneysel Desen Basamakları 49
Tablo 17. Deney ve Kontrol Grupları Betimsel İstatistikleri 50
Tablo 18. Deney ve Kontrol Gruplarına ilişkin Varyans Analizi Sonuçları 51
Tablo 19. Ön-Test KTK Madde Analizleri 52
Tablo 20. Deney Süreci Sınav Takvimi 53
Tablo 21. İzleme-1(21) Testi Betimsel İstatistikleri 54
Tablo 22. İzleme-1(21) Testi Madde Parametreleri 54
Tablo 23. İzleme-1(21) BTM Analizlerine Göre Madde Parametreleri 55
Tablo 24. İzleme-2(31) Testi Betimsel İstatistikleri 56
Tablo 25. İzleme-3(31) Testi Madde Parametreleri 56
Tablo 26. İzleme-2(31) BTM Analizlerine Göre Madde Parametreleri 57
Tablo 27. İzleme-3(41) Testi Betimsel İstatistikleri 57
Tablo 28. İzleme-3(41) Testi Madde Parametreleri 58
Tablo 29. İzleme-3(41) BTM Analizlerine Göre Madde Parametreleri 58
Tablo 30. İzleme-4(51) Testi Betimsel İstatistikleri 59
Tablo 31. İzleme-4(51) Testi Madde Parametreleri 59
Tablo 32. İzleme-4(51) BTM Analizlerine Göre Madde Parametreleri 60
Tablo 33. Deney ve Kontrol Grupları Betimsel İstatistikleri 63
Tablo 34. Son-Test KTK Madde Analizleri 65
Tablo 35. Ön-Test Toplam Puanlarına ait Betimsel İstatistikleri 66
Tablo 36. Grupların Ön-Test Puanlarına ait ANOVA Analizi 66
Tablo 37. Son-Test Toplam Puanlarına ait Betimsel İstatistikleri 67
Tablo 38. Grupların Son-Test Puanlarına ait ANOVA Analizi 67
Tablo 39. Ön-Son Test Puanlarının ANOVA Sonuçları 68
Tablo 40. Son Test Düzeltilmiş Puanlarının Gruba Göre ANCOVA Sonuçları 69
Tablo 41. Kalıcılık Testi Örneklemine ait Ön-Son ve Kalıcılık Test Puanları Betimsel İstatistikleri 69
Tablo 42. Ön-Son-Kalıcılık Test Puanlarının ANOVA Sonuçları 70

Tablo 43. 2015 PISA ve Proje Matematiksel Yeterlik Seviyeleri 72
Tablo 44. Ön-Son Test Yetenek Puanları için Gruplara Göre ANOVA Sonuçları 72
Tablo 45. Son Test Düzeltilmiş Puanlarının Gruba Göre ANCOVA Sonuçları 73
Tablo 46. Ön Test Puanlarına Göre Öğrencilerin Matematiksel Yeterlik Dağılımları 73
Tablo 47. Son Test Puanlarına Göre Öğrencilerin Matematiksel Yeterlik Dağılımları 74
Tablo 48. Deney-1 Grubu Ön-Son Test Yeterlik Düzeyleri Çapraz Tablosu..... 74
Tablo 49. Deney-2 Grubu Ön-Son Test Yeterlik Düzeyleri Çapraz Tablosu..... 75
Tablo 50. Kontrol Grubu Ön-Son Test Yeterlik Düzeyleri Çapraz Tablosu 75
Tablo 51. Gruplara Göre Ön ve Son Test Örtük Sınıf Frekansları..... 76
Tablo 52. Gruplara Göre Ön-Son Test Arasında Örtük Sınıf Değişim Yüzdeleri..... 76
Tablo 53. Örtük Sınıflara ait Sonsal Olasılıklara göre Ortalamalar ve AUC Değerleri 77
Tablo 54. Matematiksel Yeterlikler ve Üstbilişsel İzleme Performanslarının ait İlişkisiz Örneklemler t-testi Sonuçları 78
Tablo 55. Başarı Gruplarına İlişkin Betimsel İstatistikler..... 78
Tablo 56. Grupların Son-Test Puanlarına ait ANOVA Analizi 79
Tablo 57. Gruplara Göre Ön Tahmin ve Gerçek Puanlarının ANOVA Sonuçları 80
Tablo 58. Gruplara Göre Son Tahmin ve Gerçek Puanlarının ANOVA Sonuçları..... 81
Tablo 59. İzleme Testlerine Ait Madde Dağılımı..... 83
Tablo 60. Çoktan Seçmeli Test Madde Parametreleri 83
Tablo 61. Yapılandırılmış Yanıtlı Maddeleri Ait Parametre Değişmezliğinin İncelenmesi 84
Tablo 62. Testle İlgili Görüşlerin İzleme Testlerine göre Betimsel İstatistikleri 85
Tablo 63. Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum ifadesi için ANOVA Sonuçları..... 86
Tablo 64. “Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum” ifadesi için T Testi Sonuçları..... 86
Tablo 65.“Bu testteki sorulara benzer sorularla daha sık karşılaşmak isterim” ifadesi için ANOVA Sonuçları..... 87
Tablo 66.“Bu testteki sorulara benzer sorularla daha sık karşılaşmak isterim” ifadesi için T Testi Sonuçları..... 88
Tablo 67. “Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi" ifadesi için ANOVA Sonuçları..... 89
Tablo 68. "Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi" ifadesi için T Testi Sonuçları..... 89
Tablo 69. “Bu testte yer alan soruların öğrenme isteğimi artırdığını düşünüyorum" ifadesi için ANOVA Sonuçları..... 90
Tablo 70. "Bu testte yer alan soruların öğrenme isteğimi artırdığını düşünüyorum" ifadesi için T Testi Sonuçları 91

ŞEKİLER LİSTESİ

Şekil 1. Tip-1 Sinyal Tespit Kuramı' na göre elde edilebilecek olasılıklar 20
Şekil 2. Tip-2 Sinyal Tespit Kuramı' na göre elde edilebilecek olasılıklar 20
Şekil 3. Eminlik Düzeyi için ROC Grafiği 21
Şekil 4. Madde güçlük ve öğrenci puanlarının Rash sürekliliğindeki dağılımı 24
Şekil 5. PISA 2012 Matematik Yeterlik Düzeylerinin Dağılım Grafiği 38
Şekil 6. PISA 2015 ve Proje Kesme Puanların Çizgi Grafiği 39
Şekil 7. PISA ve Proje ölçümleri için matematiksel yeterlik seviyeleri çizgi grafiği 40
Şekil 8. Deney-1Ön-Test puanlarına Ait Histogram Grafiği 50
Şekil 9. Deney-2 Ön-Test puanlarına Ait Histogram Grafiği 51
Şekil 10. Kontrol Ön-Test puanlarına Ait Histogram Grafiği 51
Şekil 11.Üst Bilişsel İzleme Çoktan Seçmeli Soru 61
Şekil 12. Üst Bilişsel İzleme Eminlik Düzeyi Ölçümü 61
Şekil 13. Üst Bilişsel İzleme Doğru Yanlış Soru 61
Şekil 14. Öğrenci Performans Tahmin Formu 62
Şekil 15. Üst Bilişsel İzleme Öğretmen Soru Formu 62
Şekil 16. Üst Bilişsel İzleme Öğretmen Test Formu 62
Şekil 17. Deney-1 Son-Test puanlarına Ait Histogram Grafiği 63
Şekil 18. Deney-2 Son-Test puanlarına Ait Histogram Grafiği 64
Şekil 19. Kontrol Son-Test puanlarına Ait Histogram Grafiği 64
Şekil 20. Gruplara Ait Ön-Son Test Profil Grafiği 68
Şekil 21. Gruplara Ait Ön-Son-Kalıcılık Testi Profil Grafiği 71
Şekil 22. Öğrencilerin Gruplara Göre Performans Tahminleri 80
Şekil 23.Gruplara Ait Ön Tahmin Test Performans Puanları Profil Grafiği 81
Şekil 24.Gruplara Ait Son Tahmin Test Performans Puanları Profil Grafiği 82
Şekil 25. Uygulamaya İlişkin Görüş Formu 85
Şekil 26. Deney-1 ve Deney-2 Gruplarının "Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum" ifadesi için Profil Grafiği 87
Şekil 27.Deney-1 ve Deney-2 Gruplarının "Bu testteki sorulara benzer sorularla daha sık karşılaşmak isterim" ifadesi için Profil Grafiği 88
Şekil 28. Deney-1 ve Deney-2 Gruplarının "Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi" ifadesi için Profil Grafiği 90
Şekil 29. Deney-1 ve Deney-2 Gruplarının "Bu testte yer alan soruların öğrenme isteğimi artırdığını düşünüyorum" ifadesi için Profil Grafiği 91

ÖZET

Bu projenin temel amacı, Uluslararası geniş ölçekli sınavlarda (PISA, TIMSS vb.) Türkiye'nin matematik başarısını arttırabilmek için ortaöğretim 6. sınıf matematik dersindeki alan bilgisi ve bilişsel süreç becerilerinin bir arada ölçülebilmesine dayalı ölçme araçları tasarlamak, öğrencilerin bu bilgi ve becerilere sahip olma düzeylerini etkili istatistiksel yöntemlerle belirlemek, belirlenen öğrenme yetersizlikleri ve kaynakları konusunda öğretmen ve öğrencilere geribildirimler sunarak başarının gelişimini izlemektir.

Proje, deneysel desen kullanılarak yürütülmüştür. Tabakalı, tesadüfi küme örnekleme yöntemi kullanılarak İzmir'in 10 ilçesindeki 20 okul ve 148 şubesinden seçilen toplam 4562 öğrenci araştırmanın örneklemi olarak belirlenmiştir. Desen gereği, belirlenen örneklem Deney-1, Deney-2 ve Kontrol gruplarına ayrılmıştır. Deney-1 grubu 8 okulun, 51 şubesindeki toplam 1481 öğrenciden; Deney-2, 5 okulun 42 şubesindeki toplam 1433 öğrenciden; Kontrol grubu ise 7 okulun 55 şubesindeki toplam 1678 öğrenciden oluşmuştur. Deney-1 öğrencilerine izleme testleri yanında BTM analizlerine göre kazanım, bilişsel süreç ve hata konularında detaylı geribildirim verilmiştir. Deney-2, test etkisinin belirlenebilmesi için oluşturulduğundan bu gruptaki öğrencilere izleme testleri uygulanmış ancak sadece toplam puan üzerinden geribildirim verilmiştir. Kontrol grubu öğrencilerine ise sadece ön ve son-testler uygulanmıştır. Deney-1 ve Deney-2, BTM'ye göre geribildirim vermenin; Kontrol grubu hem izleme testlerinin hem de geribildirimin etkisini belirlemek için oluşturulmuştur. Tüm gruplara PISA testi ile eşdeğerliği kanıtlanan ön test, deney süreci sonunda son test olarak tekrar uygulanmıştır.

Sonuçlar, araştırmanın temel hipotezinin kanıtlandığını göstermiş; Deney-1 ile Deney-2 ve Deney-2 ile Kontrol grubu arasında son test ortalamalarında anlamlı farklar saptanmıştır. PISA, TIMSS gibi uluslararası geniş ölçekli sınavlarda ölçülen özelliklere sahip olma düzeyinin üst düzey düşünme becerilerini ölçen testlerin kullanılarak arttırılabileceği görülmüştür. Yanı sıra, BTM ile belirlenen öğrenci profilleri kullanılarak sunulan geribildirimlerin başarı düzeyini daha da yükselttiği ortaya çıkmıştır.

Anahtar Sözcükler: PISA, TIMSS, Bilişsel Tanı Modelleri (BTM), Üst Düzey Düşünme Becerileri, Üstbilişsel İzleme Becerisi

ABSTRACT

The main objective of this project is designing measurement instruments based on the ability to measure major knowledge and cognitive process skills in 6th grade mathematics course, determining the level of students who had acquired these knowledge and skills by using effective statistical methods and monitoring the development process of success by providing feedback to teachers and students about the determined learning inefficiencies and sources.

The project was carried out using the experimental design. total of 4562 students selected from 20 schools and 148 classes in 10 districts of İzmir were determined as the sample of the study by using stratified, randomized cluster sampling method. According to the pattern, the determined sample was divided into Experiment-1, Experiment-2 and Control groups. Experiment-1 group consisted of 1481 students in 51 classes of 8 schools, Experiment-2 group consisted of 1433 students in 42 classes of 5 schools and the control group consisted of 1678 students in 55 classes of 7 schools. Not only detailed feedback about the learning outcome, cognitive progress and error issues according to BTM analysis, but also follow-up tests were provided to Experiment-1 students. As Experiment-2 was created to determine the test effect, follow-up tests were conducted to the students in this group but only feedback in terms of total score was provided. Only pre- and post-tests were conducted to the control group students. Experiment-1 and Experiment-2 provide feedback based on BTM; The control group was established to determine the effect of both monitoring tests and feedback. The pre-test, which is equivalent to the PISA test for all groups, was re-conducted as a final test at the end of the experimental process.

The results showed that the basic hypothesis of the study was proven; Significant differences were found between Experiment-1 and Experiment-2, Experiment-2 and Control group. It has been observed that the level of having the abilities that are measured in international large scale exams such as PISA, TIMMS can be improved by using the tests which measure high-level thinking skills. In addition, it was found that the feedback provided by the student profiles determined by BTM increased the level of success.

Keywords: PISA, TIMMS, Cognitive Diagnostic Models (CDM), High level thinking skills, metacognitive monitoring

1 GİRİŞ

Bu bölümde; problem durumuna, projenin amacına, alt amaçlarına ve önemine ilişkin açıklamalara yer verilmiştir.

Eğitim ve öğretim sürecinde yaygın olarak kullanılan testler aracılığı ile çoğunlukla öğrencilerin konu alanı bilgisine yönelik başarı düzeyleri belirlenebilmektedir. Eğitimde amaçlanan ise öğrencinin hem alan bilgisini hem de bilişsel süreç becerini geliştirmektir. Bu durumda, mevcut ölçme değerlendirme sistemi sadece alan bilgisine yönelik dönüt vermektedir. Diğer bir anlatımla, eğitim öğretim süreçlerinde kullanılan ölçme araçları öğrencilerin bilişsel süreç becerilerini ölçmede yetersiz kalmaktadır. Bilişsel süreç beceri ile ilgili dönüt verilememesi eğitim ihtiyacının tam olarak karşılanamamasına sebep olmaktadır. Bu sorunu ortadan kaldırabilmek için Bilişsel Tanı Modelleri (BTM) geliştirilmiştir.

Bilişsel Tanı Modelleri, geleneksel ölçmenin, her maddenin tek bir özelliği ölçtüğü varsayımını kabul etmeyerek bir maddeyi birden fazla özellikle ilişkilendirme olanağı sağlamaktadır. Böylece hem alan bilgisi hem de bilişsel süreç becerileri detaylandırılarak aynı anda tek bir madde ile ölçülebilmektedir. Aynı zamanda öğrencilerin bir maddeyi yanlış cevaplamasının o madde ile ilişkilendirilen özelliklerden hangisine veya hangilerine sahip olmadığı bilgisini vermektedir. Bu modelleme anlayışı ile öğrencinin eğitim çıktıları doğrultusunda hangi bilgi ve becerilere sahip olup olmadığını gösteren bireysel profilleri belirlenebilmektedir. Bu anlamda BTM, öğrenme eksikliklerini geleneksel izleme testlerinden daha detaylı ve çok boyutlu istatiksel yöntemler kullanarak belirleme olanağı sağlamaktadır.

BTM'ye olan ihtiyaç ABD'de gerçekleştiren "Hiçbir Çocuk Geride Kalmasın Reformu" (No Child Left Behind Act) ile artmıştır. Hiçbir Çocuk Geride Kalmasın Reformu ABD'nin eğitim sistemindeki problemlerin giderek artması ile ortaya çıkmıştır. Ülkede 4. sınıf öğrencilerinin %70'i ulusal okuma testlerinde basit seviyede bile okuma başarısı gösterememiştir. Lise son sınıf öğrencileri de, uluslararası matematik testlerinde son sıralarda yer almaktadır. Kolejleri kazanan birinci sınıf öğrencilerinin yaklaşık üçte birinin ise derslere başlamadan bir destek eğitim almaları gerekmektedir. Reform kapsamında öğrencilerin bilişsel süreç gelişiminde etkili faktörler üzerinde durulmuştur. Öğrencilerin sadece bir test puanı ile değerlendirilmesi değil, başarılı ve başarısız oldukları yönleri ile raporlaştırılması istenmiştir. Her öğrenci için hazırlanan bilişsel tanı raporları ile öğretmenler, aileler ve öğrencilerin kendileri bilgilendirilmiştir(The White House, 2003). Reformun başlaması ile birlikte bilişsel modeller, bilgisayar destekli değerlendirme sistemleri ve puanlama tekniklerindeki gelişimler de hızlanmıştır. Öğrencilerin yeterliliklerine ilişkin vurgu yapılması ve bilişsel dönüt verilmesi

BTM'nin gelişimini ve kullanımını desteklemiştir (Council, 2001; Embretson, 1999; J. Leighton & Gierl, 2007; Mislevy & Levy, 2006; Nichols, 1994; Stout, 2002).

ABD Ulusal Araştırma Konseyi'nin (Council, 2001), yayınladığı raporda bilişsel tanı modellerine olan ihtiyaç vurgulanmış ve yayınlanan raporda değerlendirmenin *bilişsel unsuru, gözlemleme unsuru* ve *yorumlama unsuru* olmak üzere üç temel yapısının bulunduğu belirtilmiştir. *Bilişsel unsurun* bilişsel modellerini işaret ettiği ve bu modellerin öğrencilerin bilişsel süreçlerinin ve mevcut yeterliğinin nasıl geliştiğini açıklamaya çalışan bilişsel tanı modelleri olduğu belirtilmiştir. Bilişsel unsurun etkilediği *gözlemleme unsuru* ise bilişsel yapı ile ilgili öğrenci davranışlarının ortaya çıkarılmasını ifade etmektedir. Ölçülecek bilişsel yapının daha detaylı tanımlanmasıyla öğrenciden beklenen performansın içeriği ve şekli ile ilgili ayrıntılı bilgi edinilebilmektedir. Aynı zamanda bilişsel unsur, yorumlama unsurunun da temelini oluşturmaktadır. *Yorumlama unsuru*, öğrencilerin davranışlarına ilişkin bilişsel sürecin nasıl yorumlanması gerektiğini vurgulamaktadır

Benzer bir durum Almanya'nın 2000 yılındaki PISA sonuçları sonrasında da yaşanmıştır. PISA 2000 sınavında yaşanan başarısızlık "PISA-şok" olarak isimlendirilmiş ve bu durumun telafi edilebilmesi için eğitim politikası ve öğretim programında reform yapılması gerekliliği ülkenin gündemine girmiştir. Program yenileme çalışmaları, konu alanı yeterlikleri ve bilişsel değerlendirme modeli üzerine kurulmuştur. Bu model kapsamında ulusal performans standartları değiştirilerek girdi temelli program yerine yeterlikleri ve bilişsel özellikleri dikkate alan standartlar geliştirilmiştir. Berlin'deki Ulusal Eğitim Süreçleri Enstitüsü tarafından Matematik, Fen bilimleri, Almanca, İngilizce ve Fransızca öğretim programlarına ve ölçme değerlendirme sistemlerine yeni standartlar geliştirilmiştir. Matematik alanı için ulusal standartlar PISA'daki çerçeve genişletilerek 6 adet genel matematik yeterlik alanı belirlenmiştir. Bu yeterlikler, (1) matematiksel tartışma, (2) matematiksel problem çözme, (3) matematiksel modelleme, (4) matematiksel simgelerden yararlanma, (5) sembolik, formal ve teknik açılardan matematiksel çalışma, (6) matematik hakkında iletişim olarak belirlenmiştir. Genel matematik yeterliklerinin yanı sıra Matematik alanına özgü yeterlilik olanları da belirlenmiştir. Bu yeterlik alanları (1) sayılar, (2) ölçme, (3) objeler ve grafikler, (4) fonksiyonel ilişkiler, (5) olasılık ve istatistik olarak belirlenmiş ve genel yeterlik alanları ile alana özgü yeterlik alanları birlikte değerlendirilerek bilişsel model tamamlanmıştır (Rupp, A. A. & Mislevy, 2007).

Yukarıda yapılan açıklamalar çerçevesinde, çeşitli ülkelerin deneyimleri değerlendirme sistemlerinin öğrenmeye ve bireysel gelişime katkıda bulunması gerektiğini göstermektedir. Bilişsel tanı modelleri ile öğrencilerin bilişsel profillerinin çıkarılabilmesi öğrencinin kendisine, öğretmene, eğitim programlarına ve ölçme değerlendirme sistemlerine katkıda bulunmakta ve

yön vermektedir (Basokcu, 2010). Pek çok ülkede eğitim reformları ile gündeme gelen bilişsel tanı modellerinin Türk Eğitim Sisteminde kullanılması ve eğitim programları ile ölçme değerlendirme sistemlerine katkıda bulunması projenin temel konu alanı oluşturmuştur.

Türkiye'nin de katıldığı PISA, TIMSS, PIRLS gibi öğrenci başarısını belirlemeye yönelik testlerde elde edilen başarı düzeyi dikkate alındığında çeşitli ülkelerde olduğu gibi öğrenci başarısını artırmak, öğrencilerin bireysel profillerini dikkate almak, bireye özgü eğitim ihtiyacını karşılamak amacıyla BTM'den faydalanılabileceği görülmektedir.

1.1 Amaç ve Önem

Genel olarak TIMSS ve PISA testlerinin amacı Matematik ve Fen alanlarında öğretim ve eğitimi geliştirmek için ülkelerin eğitim sistemleri hakkında karşılaştırmalı veri sağlamaktır. Bu amaç doğrultusunda öğrencilerin Fen ve Matematik alanındaki performansları, eğitim sistemleri, öğretim programları, öğrenci, öğretmen ve okulların özellikleriyle ilgili bilgiler toplanmaktadır. Bu testlerde Türkiye'nin başarı düzeyi çoğunlukla katılımcı ülkelerin ortalamasının altında kalmaktadır. Bu nedenle araştırma ortaöğretim matematik öğretim programının BTM çerçevesinde nasıl geliştirilebileceğine odaklanmaktadır. Diğer bir anlatımla, BTM kapsamında Matematik öğretiminde öğrencilerin konu alanı bilgisi ve bilişsel profillerini belirlemeyi sağlayacak bir izleme modeli kullanarak ortaöğretim Matematik öğretim programının geliştirmesi için kaynak teşkil edebilecek ampirik kanıtlar sağlamak amaçlanmıştır.

Uluslararası geniş ölçekli sınavlarda (PISA, TIMSS, vb.) Türkiye'nin matematik başarısını arttırabilmek için ortaöğretim 6 sınıf matematik dersindeki alan bilgisi, üst düzey düşünme becerileri ve bilişsel süreç becerilerinin bir arada ölçülebilmesine dayalı ölçme araçları tasarlamak, öğrencilerin bu bilgi ve becerilere sahip olma düzeylerini etkili istatistiksel yöntemlerle belirlemek, belirlenen öğrenme yetersizlikleri ve kaynakları konusunda öğretmen ve öğrencilere geribildirimler sunarak başarının gelişimini izlemek bu çalışmanın temel amacıdır. Bu temel amaca ulaşmak için 6. Sınıf Matematik öğretim programının kazanımlarında yer alan bilişsel süreçleri tanımlamak, her bir süreç becerisinde gerekli konu alanı bilgisini tanımlamak, konu alanı bilgisi, üst düzey düşünme becerileri ve bilişsel süreçleri birlikte ölçmeye yönelik ölçme araçları geliştirmek, geliştirilen ölçme araçları için Q-matrisler hazırlamak, ölçme araçlarının okul uygulamalarını gerçekleştirmek, bilişsel süreç becerileri ve konu alanı bilgisini aynı anda ölçebilecek farklı BTM kullanarak analiz etmek ve böylelikle hem öğrencilerin bilişsel profillerini belirlemek hem de model uyumunu test etmek, bireye özgü öğrenme eksiklikleri ve sahip oldukları bilişsel süreç becerileri içeren bilişsel profilleri hazırlayarak öğrencileri bilgilendirmek, geliştirilen izleme modelinin geçerliliğini deneysel yöntemlerle test etmek araştırmanın alt amaçları olarak belirlenmiştir.

Eğitimde öğrenme eksikliklerinin belirlenmesi; biçimlendirmeye yönelik (formative) değerlendirme aşamasında gerçekleştirilmektedir. Bu süreçte öğrencilerin öğretim hedefleri doğrultusunda belirlenmiş kazanımlara sahip olup olmadıklarına yönelik geribildirimler veren ölçme araçları kullanılmaktadır. Fakat kullanılan ölçme araçları, eğitimin önemli çıktılarından biri olan bilişsel süreç becerilerindeki eksikleri belirleyememektedir. Aynı zamanda halihazırda kullanılan bu ölçme yöntemiyle bir soru ile yalnızca tek bir özellik ölçülebilmektedir. Araştırmada kullanılan BTM ile bir soru hem birden fazla kazanımla ilişkilendirilebilmekte hem de öğrencinin bilişsel süreç becerilerine sahip olma düzeyi belirlenebilmektedir. Ayrıca BTM, açık uçlu soruları analiz etmek için de uygun bir yöntem olduğundan araştırmada açık uçlu soru formatı da kullanılarak öğrenci başarısını belirlemek için şans başarısından bağımsız, daha üst düzey özellikleri ölçmeye yönelik testler geliştirilmiştir. BTM'ye göre geliştirilen izleme modelinin işlerliği, Milli Eğitim Bakanlığı (MEB) ile işbirliği içinde deneysel yöntemlerle test edilmiş ve Türk Eğitim Sisteminde BTM'nin kullanılabilirliğine ilişkin çıkarımlarda bulunulmuştur. Sonuçlar, Türk Eğitim Sisteminde matematik başarısının arttırılmasına katkı getirebilecek niteliktedir. Araştırmada, Türk Eğitim Sisteminde BTM uygulaması ilk kez gerçekleştirilmiştir. Böylelikle 6. Sınıf Matematik öğretim programının hem kazanımlarından konu alanı bilgisine, eğitim durumlarından ölçme ve değerlendirme yaklaşımına, hem de programın uygulama süreçleri ve ürünleri konusunda çok yönlü bilgiler elde edilmiştir. Araştırmada aynı zamanda madde parametreleri, madde güçlüğü, model uyumu, q-matris geçerliği, bilişsel sınıflara ait sonsal olasılık dağılımları gibi BTM ile ilgili kuramsal bilgiler de elde edilebilmiştir. Yukarıda listelenen hedeflere ulaşıldığında Türk Eğitim Sisteminde BTM uygulaması gerçekleşmiş olacaktır.

Eğitimde test geliştirme teknikleri aynı zamanda karar süreçlerine yön vermektedir. Öğretim süreçlerinde öğrencilerin başarısı hakkında verilen kararlar, uygulanan testler sonucunda alınan toplam puanlar üzerinden verilmektedir. Ancak günümüzde bireyin başarısı hakkında yalnızca toplam puan üzerinden karar vermeye dayalı yöntemlerin yerine seçenek ya da tamamlayıcı olabilecek yeni yöntemler geliştirilmeye başlanmıştır. Bu yöntemlerden biri olan BTM'ye dayalı yürütülecek bu çalışma; literatür, karar alıcılar ve program paydaşları açısından özgün bir değere sahiptir.

2 ALANYAZIN TARAMASI

Bu bölümde; projenin kuramsal alt yapısını oluşturan BTM modelleri, BTM modellerinde örtük sınıflar ve yetenek kestirimi, üst düzey düşünme becerileri, yapılandırılmış yanıtlı maddeler ve üstbilişsel izleme ile ilgili alanyazın özet olarak sunulmuştur.

2.1 Bilişsel Tanı Modelleri

BTM'nin amacı, iki kategorili niteliklerin sıralanışına dayanan örtük kategorilere göre öğrencileri sınıflamaktır. Bu modellerde uzmanlığı belirleyen bir vektör olan örtük değişkenler, öğrenciye yönelik teşhisin altında yatan özellik setini tanımlamaktadır. Burada "özellik" kavramı nitelik, vasıf, görev, alt görev, bilişsel süreç veya beceri olarak tanımlanabilir (Tatsuoka, 1983, 1995).

BTM ile test geliştirme sürecinde, testteki her bir madde ölçülmek istenen özelliklerin ve bilişsel yapıların temsilcisidir. Testteki her madde bu özelliklerden birini veya birkaçını ölçecek şekilde düzenlenir. BTM ile test geliştirme sürecinde maddelerin ölçülecek özelliklerle eşleştirilmesi için Q-matris kullanılır. Q-matris terimi ilk defa Tatsuoka (1995) tarafından kullanılmıştır ve maddeler ile maddelere ait özelliklerin dağılımını göstermektedir (Kato, 2009). Proje kapsamında testlere ait Q matrisler belirlenerek farklı bilişsel tanı modelleri ile analiz edilmiştir. Q-matris ve proje kapsamındaki bilişsel tanı modelleri aşağıda açıklanmıştır.

2.1.1 Q Matris

Q matris, bir test için nitelik ya da beceri vektörlerinin sütunlarda, maddelerin ise satırlarda temsil edildiği bir örüntüdür. Nitelikler, alan uzmanları tarafından belirlenen özellikler, prosedürler, buluş yöntemleri, stratejiler, beceriler ve başka bilgi bileşenleri olabilir. Q matris 1-0 şekilde ikili numaralar kullanarak özelliğin maddede bulunup bulunmadığını gösterir. Bu numaralandırma ilk önce Fischer (1973) tarafından "ağırlıklandırılma" şeklinde tanımlanmış ve bir k niteliğinin bir i maddesinde bulunduğu durumda 1, bulunmadığı durumda ise 0 şeklinde kodlandığını belirtmiştir.

J × K matrisini Q =[q jk] 0 ve 1 için şu şekilde yazılabilir:

$$q_{jk} \begin{cases} 1 \text{ eğer k özelliği , j maddesinde bulunuyorsa} \\ 0 \text{ eğer k özelliği , j maddesinde bulunmuyorsa} \end{cases}$$

Tablo 1'de, 3 nitelik ve 5 madde için hazırlanmış örnek bir Q matrisi görülmektedir. Q matrisinde görüldüğü gibi 1. maddeyi doğru cevaplamak için ilk niteliğe sahip olmak gerekmektedir. İkinci madde için ise 1. ve 2. niteliklere öğrencilerin sahip olması gerekmektedir.

Tablo 1. Örnek Q Matris

Maddeler	α1	α2	α3
1	1	0	0
2	1	1	0
3	1	1	1
4	0	1	1
5	0	0	1

Ölçülen niteliklerle birlikte düzenlenen Q matrisi, madde güçlüğünün psikometrik model ve bilişsel süreçle arasındaki bağlantısını kurmaktadır. Modelin seçimi ve Q matrisi birbirine eşit sınırlardan oluşan, örtük sınıf cevaplama olasılıklarını belirlemeyi sağlamaktadır.

2.2 Proje Kapsamında Kullanılacak Bilişsel Tanı Modelleri

2.2.1 DINA Model

Haertel (1989) tarafından geliştirilen DINA model, diğer modeller gibi ikili yetenek modellerine benzeyen bir örtük sınıf analizidir. DINA model çoğu BTM'de görüldüğü gibi madde özellik ilişkisini temel alır. Modelin iyi işleyebilmesi, bir maddenin doğru cevaplanması için gerekli olan özelliklerin doğru belirlenmesine bağlıdır. Bununla birlikte DINA modelin geliştirilen G-DINA ve HO-DINA gibi modifikasyonlarıyla bir maddeyle ilişkili birden çok özelliğin gerek ağırlıklandırılarak gerekse hiyerarşik olarak analize katılabilme imkanı bulunmaktadır (de la Torre & Douglas, 2004; de la Torre & Lee, 2010).

DINA model, cevaplayıcının gözlenen yeteneğinin altında yatan örtük özelliği ortaya çıkartmayı amaçlamaktadır. Bu anlamda model, örtük özellikle gözlenen özellik arasındaki ilişkiyi olasılıkla temellendirmekte ve her madde için iki madde parametresinin sınıflanmasını sağlamaktadır. Bunlar s "kaydırma" (slip) ve g "tahmin" (guess) parametreleridir.

$s_j = P\left[Y_{ij} = 0 | \eta_{ij} = 1\right]$ ve

$g_j = P\left[Y_{ij} = 1 | \eta_{ij} = 0\right]$,

Yukarıdaki formülde s_j örtük özelliğe sahip bireyin j maddesine yanlış cevap verme olasılığını gösteren durumu (yanlış pozitif olasılık), ve g_j ise örtük sınıfa sahip olmayan bireyin doğru cevap verme olasılığı durumunu (doğru pozitif olasılık) ifade etmektedir. s_j parametresi ne kadar düşük olursa aranan özelliklere sahip bireylerin doğru cevap verme olasılığı o kadar artmaktadır(de la Torre, 2009). Bununla birlikte g_j parametresine tahmin parametresi de denir. Maris (1999), alternatif olarak g_j parametresini zihinsel tahmin yeteneğine başarıyla kullanmak olarak da açıklar. Tahmin parametresi bu anlamda Madde Tepki Kuramındaki (IRT) şans parametresinden farklı bir yapıdadır. Modelde g parametresi sadece maddeyi doğru cevaplamak için gerekli olan özelliğe sahip olmayan bireyin maddeyi doğru cevaplaması anlamına gelmemektedir. Aynı zamanda cevaplayıcının maddeyi doğru cevaplamak için gerekli olduğu düşünülen özellikler dışında başka özellikleri kullanarak maddeyi doğru cevaplaması anlamına da gelmektedir. Bu nedenle Q matris tarafından tanımlanmayan farklı özelliklerin de bir maddenin doğru cevaplanması için yeterli olabileceğini göstermektedir. Bir madde için "1" değerine yakın çıkan g parametresi sadece o maddeyi özelliğe sahip olmayan bireylerin cevapladığı anlamına gelmez, aynı zamanda maddeyi doğru cevaplamak için gerekli

olan bazı özelliklerin belirlenemediği şeklinde de yorumlanabilir. DINA modelde madde tepki fonksiyonu aşağıdaki biçimdedir:

$$P(\alpha_{ij}^{*}) = \begin{cases} g_j & \text{if } \alpha_{ij}^{*} \prec 1_{K_j^{*}} \\ 1 - s_j & \text{otherwise} \end{cases}$$

DINA modelde koşullu dağılan madde cevap değişkeni Y_{ij} aynı zamanda α_{ij} den η_{ij} ye de bağlıdır. Bu DINA modelin olasılık fonksiyonunun bir uzantısıdır, konular arasındaki bağımsızlığı gösteren koşullu bağımsızlık şu şekilde yazılabilir:

$$L(s, g; \alpha) = \prod_{i=1}^{N} \prod_{j=1}^{J} \left[s_j^{1-y_{ij}} (1 - s_j)^{y_{ij}} \right]^{\eta_{ij}} \left[g_j^{y_{ij}} (1 - g_j)^{1-y_{ij}} \right]^{1-\eta_{ij}}$$

DINA modelde s ve g parametreleri madde düzeyinde ortaya çıkmaktadır. Her madde popülasyonu iki sınıfa bölmekte ve aynı sınıfa düşen öğrencilerin o maddeye doğru cevap verme olasılıkları eşit olmaktadır. Öğrencilerin sınav performansı testte yoklanan niteliğin vektörünün tam bir göstergesi değildir. Bu nedenle olasılık temelinde bir model ancak s ve g'nin olasılığını görmeye izin verir. "Kaydırma" durumu öğrencinin maddede aranan niteliğe sahip olmasına rağmen alt görevi doğru cevaplayamadığı ya da maddeyi doğru cevaplayamadığında gerçekleşir. "Tahmin" ise öğrencinin maddede aranan özelliklerden birine ya da birkaçına sahip olmamasına rağmen alt görevleri tamamlaması veya maddeyi doğru cevaplaması durumudur.

DINA modelin bir cevaplayıcının bir özelliğe sahip olup olmadığını belirleme aşaması olasılık temelli bir süreçtir. Öğrencinin özellik bakımından 0 sınıfında mı yoksa 1 sınıfında mı olacağı bir olasılık değeridir. Bu değer araştırmacı tarafından değiştirilebileceği gibi genellikle kullanımda 0,50 eşiği temel alınır. Öğrencinin özelliğe sahip olma olasılığı 0,50 değerinin altında kalırsa 0 sınıfına, üstünde ya da aynı değeri alırsa 1 sınıfına dahil olur. DINA model öğrenciye ilişkin α kestirimini yaparken o özelliği temsil eden maddelerin doğru cevaplanma oranını temel almaz. Öğrencinin özelliğe sahip olma olasılığı özelliği temsil eden maddelerin güçleriyle ilgilidir (de la Torre & Douglas, 2008).

2.2.2 G-DINA Model

G-DINA model DINA modelin genelleştirilmiş halidir. Bu model de birçok bilişsel tanı modeli gibi JxK Q matrisine dayanır. G-DINA model de örtük sınıfları $2^{K_j^{*}}$ sayıda örtük gruba ayırır. Her örtük grup $\propto_{lj}^{*}$ ile gösterilen bir beceri vektörüne indirgenir. Her örtük grubun kendisine ait $P(\propto_{lj}^{*})$ ile ifade edilen başarı olasılığı vardır (de la Torre, 2011).

G-DINA için $P(\propto_{lj}^{*})$ temelli olan orijinal formül, her bir spesifik beceri ve bu becerilerin birbirleriyle etkileşimi toplam etkilerine göre parçalara ayrılabilmektedir. Aşağıda G-DINA model için olasılık formülü verilmiştir:

$$P(\alpha_{lj}^{*}) = \delta_{j0} + \sum_{k=1}^{K_j^{*}} \delta_{jk}\alpha_{lk} + \sum_{k'=k+1}^{K_j^{*}} \sum_{k=1}^{K_j^{*}-1} \delta_{jkk'}\, \alpha_{lk}\alpha_{lk'} \ldots + \delta_{j12\ldots K_j^{*}} \prod_{k=1}^{K_j^{*}} \alpha_{lk}$$

δ_{j0} = j maddesinin kesişimi

δ_{jk} = αk üstündeki temel etki

$\delta_{jkk'}$ = αk ve αk' nın birbirleriyle etkileşiminin etkisi

$\delta_{j12\ldots K_j^{*}}$ = α1,…….,$\alpha_{K_j^{*}}$ bağlı karşılıklı etkileşimin etkisi anlamına gelmektedir.

G-DINA modelin tahminleme kodları EM algoritmasının bir uygulamasıdır. Analiz sürecinde öncelikle $P(\propto_{lj}^{*})$ değerleri standart hatalarıyla birlikte hesaplanır. Daha sonra yeteneklerin sonsal (posterior) olasılıkları belirlenir ve bu olasılıklara göre öğrencilerin örtük sınıfları ile madde ve testte ilişkin uyum istatistikleri hesaplanır.

2.2.3 Fusion Model

Fusion Model, Rule Space Model(RUM)'un Hartz (2002) tarafından geliştirilmiş ve kestirimler için hiyerarşik yapıda Bayesian yöntemi kullanan bir uzantısıdır. Fusion Modele ait MTF aşağıdaki gibidir:

$$P(X_{ij} = 1|\alpha_j, \theta_j) = \pi_i^{*} \prod_{k=1}^{K} r_{ik}^{*(1-\alpha_{jk}).q_{ik}} P_{c_i}(\theta_j)$$

Formülde Xij =, j cevaplayıcısının i maddesine yanıtı, αj, j cevaplayıcısının Q matrisi tarafından belirlenmiş K becerisine ilişkin bilişsel beceri düzeyinin vektörünü, θj = , j cevaplayıcısının Q matrisi tarafından belirlenemeyen testle ilişkili beceri veya yeteneğini, πi* = i. madde için gerekli bütün becerilere sahip olan cevaplayıcının i maddesini doğru cevaplama olasılığı, ri*= i maddesi için k becerisine ilişkin (Q matrisi tarafından tanımlanan) ayırt edicilik parametresini, qik = i maddesi için k becerisinin gerekli olup olmamasını, ci= Q matrisi tarafından belirlenen becerilere ek olarak maddeyi doğru cevaplamak için gerekli olan θ_j düzeyini, $P_{ci}(\theta_j)$ = Q matrisi tarafından tanımlamayan becerilerin doğru uygulanma olasılığını ifade eder (Roussos et al., 2007).

Modelde genel olarak Hartz (2002), $\pi_{jk}^{*} = \prod_{k=1}^{k} \pi_{jk}$ ve $r_{jk}^{*} = \frac{r_{j,k}}{\pi_{j,k}}$ yi Bayesian Markov Monte Carlo Zinciri yöntemi kullanarak reparametrize etmiştir. RUM'da π_{jk}^{*} , Q temelli şartlı madde

güçlüğüyken r_{jk}^{*} ise j maddesi tarafından temsil edilen beceriler hakkında elde edilen bilgiyi göstermektedir. Bunun yanında bu kestirim için "Arpeggio" isimli bilgisayar programını geliştirmiştir (Hartz, 2002). Diğer IRT modelleri gibi Fusion Model de cevaplayıcının bir maddeyi cevaplama olasılığını yetenek ve madde parametrelerini kullanarak belirlemeye çalışır. Model madde cevap olasılıkları ile test geliştirenler tarafından belirlenen özellik setleri arasında bağlar kurarak cevaplayıcıların özelliklere sahip olup olmadıklarını belirlemektedir. Özellikler ve maddeler arasındaki bilişsel ilişki 1-0 şeklinde belirlenen Q matris ile tanımlanır (Tatsuoka & Linn, 1983).

Beceri ve madde arasındaki ilişkinin doğası gereği Fusion model, maddenin doğru cevaplanması için bireyin sahip olması gereken özelliklerle tanımlanan özellikler arasında bir bağ kurar. Bu durumda model doğru cevaplamanın maddeyle ilişkili bütün özelliklere sahip olmakla ilişkili olduğunu varsayar. Bu bir madde için yüksek doğru cevaplama olasılığının o maddeyle ilişkili özelliklere yüksek düzeyde sahip olma gerekliliğidir.

2.3 Bilişsel Tanı Modellerinde Örtük Sınıflar ve Yetenek Kestirimi

BTM'de öğrenciler her madde için temel iki sınıfta yer alır. Bu sınıflardan ilki yokluk sınıfı (null class) yani beklenen hiçbir beceriye sahip olmayan öğrencilerin oluşturduğu grup ve diğeri de tam sınıf (full class) yani bütün becerilere sahip olan öğrencilerin sınıfıdır. DINA modelde tek bir beceriyi bile kaçıran biri yokluk sınıfı (null class) içinde yer alır. Bir maddenin doğru cevaplanması için gerekli bütün becerilere sahip olan bireyin maddeyi doğru cevaplama olasılığını gösteren fonksiyon aşağıdaki gibidir:

$$P\left[Y_{ij} = 1 \middle| \eta_{ij}, s_j, g_j\right] = \left(1 - s_j\right)^{\eta_{ij}} g_j^{1-\eta_{ij}}$$

P aranan bütün becerilere sahip olan öğrencinin maddeyi doğru cevaplama olasılığıdır. η_{ij} , α tarafından belirlenen örtük cevaplamadır ve i inci konunun niteliği ve qj nin vektörüdür. Q matrisinin j inci maddesine tekabül eden sıra şu şekilde gösterilebilir:

$$\eta_{ij} = \prod_{k=1}^{K} \propto_{ik}^{q_{jk}}$$

Tatsuoka (1983) $\alpha_i = (\alpha_{i1}, \dots, \alpha_{iK})$ yi "knowledge states" –bilgi durumları– olarak tanımlar. Burada $\alpha_{iK} = 0\ veya\ 1$ olması i öğrencisinin k; $\eta_i = \left(\eta_{i1}, \dots, \eta_{ij}\right)$ niteliğine sahip olup olmamasına bağlıdır. j = toplam madde sayısını göstermektedir, bu maddeler i öğrencisinin istenilen niteliklere sahip olup olmadığını ve Y_{ij} gözlenen puanının belirlenmesini sağlar. Örtük yanıt öğrenci yeteneği ve maddenin gerekliliklerinin bir fonksiyonudur. η_{ij} belirlendiğinde i öğrencisinin j maddesine doğru cevap verme olasılığı; η_{ij}= 0 olduğunda g_j ve η_{ij}=1 olduğunda 1-sj olur (de la Torre, 2009).

DINA model geliştirilen ya da uygulanan bir test için belirlenen k sayıdaki özelliği temel alarak 2k sayıda örtük sınıf belirlemektedir. Örneğin bir testte sadece 3 özelliğin ölçüldüğü düşünülürse bu durumda testi alan bireyler toplam 8 örtük sınıf içinde sınıflanır. 3 özellik için olası sınıflar “000”, “100”, “010”, “001”, “110”, “011”, “101” ve “111” şeklinde sıralanır. Hiçbir özelliğe sahip olmayanlar ilk sınıfa dâhil olurken sadece birinci ve üçüncü özelliğe sahip olan bireyler 7. sınıfa yerleşirler.

Maddenin olasılık belirleyen temel parametresi ise s dir. Daha önce değinildiği gibi s parametresi özelliğe sahip bireylerin maddeyi yanlış cevaplama oranıdır. Bu durumda eğer bir madde, doğru cevaplanması için gerekli özelliklere sahip bireylerin maddeyi büyük oranda doğru cevaplandığı yönünde bir parametre üretiyorsa, o maddeyi doğru cevaplayan bireylerin de o özelliklere sahip olduğu söylenebilir. De la Torre (2010b) örtük sınıfların belirlenmesi ve α parametresi kestirimi konusunda maddenin gücünün önemli bir etkisi olduğunu söylemiştir. Bununla birlikte modelin çok boyutlu yapısı da öğrencinin örtük sınıfını belirlemeyi etkilemektedir. BTM’de bir madde birden çok özellikle ilişkilendirilebildiği için her madde her öğrenci için aynı değeri taşımamaktadır. Bazı durumlarda aynı maddelere doğru cevap vermesine rağmen bir öğrencinin 0 diğer öğrencinin 1 sınıfına atanması mümkün olmaktadır. Başokçu’nun (2012) DINA model kullanarak yaptığı çalışmada, 40 madde ve 8 özellik için Q matrisi belirlenmiş bir testte A ve B cevaplayıcılarının her ikisi de toplamda 20 doğru yanıt vermiştir. Buna rağmen A cevaplayıcısının atandığı örtük sınıf “11101010” iken B cevaplayıcısının “00000000” sınıfına dâhil olduğu görülmüştür. Aynı zamanda Başokçu ve Arlı’nın (2014) araştırmasında, bu özellikler olasılık değeri olarak da hesaplanmış ve öğrencilerin gelişimleri daha detaylı izlenebilmiştir.

2.4 Yapılandırılmış Yanıtlı Madde Formatı

Doğru çıktılara ulaşabilmek için model ve analizlerin önemi kadar maddelerin kalitesi ve türü de oldukça önemlidir. Mevcut testlerin çoğu, doğası gereği tanı amaçlı değildir ve öğretime olan katkıları tartışmalıdır (Nickerson, 1989). Fakat daha genel olarak, standartlaştırılmış, çoktan seçmeli sorular, pek çok araştırmacıya göre; eleştirel düşünmeyi, fikirleri sentezlemeyi, formüle etmeyi ve planlamayı içeren çeşitli bilişsel süreçleri ortaya çıkarmakta başarısız olarak algılanmaktadır. (Frederiksen, 1984; Haney & Madaus, 1989). Çoktan seçmeli maddeler genel olarak bunun gibi eleştirilerin hedefi olup, kısmen de bilginin çağrılmasını gerektirmedikleri düşünülmektedir. Bu sebeple çoktan seçmeli maddelere alternatifler geliştirilmeye çalışılmıştır. Yapılandırılmış yanıtlı maddeler bunlardan biridir.

Bilişsel psikolojide bilginin kullanılması için iki ayrı isleyiş söz konusudur. Bunlardan biri tanıma (recognition) diğeri ise geri çağırmadır (recall). Eğer madde tanımayı gerektiriyorsa, içerisinde bir uyaran bulundurur. Bu uyaran, bilginin çağırılmasını tetikler ve doğru yanıtın tanınmasını

sağlar. Çoktan seçmeli maddeleri, tanıma gerektiren maddelere örnek olarak gösterebiliriz. Geri çağırma ise uyaranla karşılaştıktan sonra bellekte doğru bilgiyi aramayı gerektirir. Doğru bilginin bellekte bulunması için herhangi bir hatırlatıcıya gerek yoktur ve sonrasında bilgi organize bir şekilde yapılandırılarak yanıta dönüştürülür. Geri çağırma, daha çok açık uçlu soruları yanıtlarken kullanılır. Ancak çoktan seçmeli maddelerle de geri çağırma gerektiren bilişsel süreçler sağlanabilir. Bu şekilde bellekte çeşitli öğrenme bölümlerine erişmek için karmaşık düşünme süreçleri oluşur (Nuthall & Alton-Lee, 1995). Çoktan seçmeli sorular karmaşık düşünme süreçlerini ortaya çıkarmak için tasarlansa bile, yapılandırılmış yanıtlı maddeler ile oluşturulabilecek düşünce dizisinin tümünün çoktan seçmeli maddelerle yakalanabileceği anlamına gelmemektedir (Messick, 1995). Birçok çalışma, öğrencilerin açık uçlu ve çoktan seçmeli maddelerden aldıkları puanların birbiriyle yakından ilişki gösterdiğini ve her iki türün de benzer şeyler ölçtüğünü bulmuştur (Bridgeman, 1992; Lukhele, Thissen, & Wainer, 1994; Walstad & Becker, 1994). Açık uçlu maddelerin oluşturulması, uygulanması ve değerlendirmesi gibi zorlulukların olması elde edilen sonuçların bu çabaya değmeyeceği iddiasını ortaya çıkarsa da ölçme ve değerlendirmenin amacı düşünüldüğünde bu iddianın doğru olmadığına yönelik görüşte yaygındır. Popham (2003), amaç bilgi basamağındaki öğrenme çıktılarını ölçmek ise çoktan seçmeli maddelerin daha kullanışlı olduğunu ve bu tür testlerin, konu tekrarı yapılıp yapılmaması kararı için uygun bir ölçme aracı olduğuna değinmektedir. Ancak, öğrencinin yaratıcılığını ölçmek, fikirleri sentezleme becerisini, etkili ifade ya da problem çözme becerisini ölçmek için yapılandırılmış yanıtlı maddelerin daha doğru bir tercih olacağını savunmaktadır.

Öğrencilerin sınav türüne bağlı olarak farklı çalışmalar yaptıkları ve bu durumun da öğrenci öğreniminin niteliğini değiştirdiğine dair kanıtlar bulunmaktadır. Yapılandırılmış yanıtlı maddelerden oluşan sınav kavrama ve üst basamaklardaki öğrenmeyi desteklerken, çoktan seçmeli maddeler ise genellikle hatırlamayı gerektiren becerileri geliştirmeyi desteklemektedir(Martinez, 1999; Traub & Rowley, 2018).

California State Department of Educations'ın açık uçlu sorularla ilgi araştırmalarına göre öğrencilerin çoğunun matematiksel fikirleri yazılı olarak ifade etmekten yoksun olduklarına dair bulgular elde edilmiştir (Stenmark, EQUALS., & Committee, 1989). NCTM (National Council Of Teachers Of Mathematics) (1991),'ye göre öğrencilerin matematiksel becerilerini günlük hayat becerileri ile birleştirmeleri için yanıtlarını kendilerinin oluşturmaları gerekmektedir. Sadece hatırlamayı gerektiren maddeler okul dışındaki hayatta matematiksel düşünme, iletişim kurma, ilişkilendirme becerilerini talep eden dünyaya etkili bir şekilde hazırlamayacaktır. SAT, GRE, Key Stage gibi sınavlarda sadece çoktan seçmeli ya da sadece yapılandırılmış yanıtlı maddeler kullanmak yerine ölçülmek istenilen özelliğe ve amaca göre iki

tür birlikte kullanılmaktadır. PISA VE TIMSS gibi uluslararası sınavlarda da amaç ve kapsam düşünüldüğünde hem çoktan seçmeli maddeler hem de yapılandırılmış yanıtlı maddeler test içerisinde birlikte yer almaktadır.

2.5 Üst Düzey Düşünme Becerileri

Zoller, (Miri, David, & Uri, 2007; Zoller, 1993, 1999; Zoller & Pushkin, 2007; Zoller & Tsaparlis, 1997) tarafından yapılan üst ve alt düzey bilişsel becerilerin tanımlanmasına ilişkin çalışmalarda Bloom Taksonomisi esas alınmıştır. Zoller'e göre üst düzey bilişsel beceriler (HOCS: Higher Order Cognitive Skills) olarak kastedilen özellikler; soru sorma, eleştirel ve sistemli düşünme, problem çözme, analiz etme, değerlendirme, yeni bilgiler sentezleme ve karar verme yeteneklerini içermektedir.

Üst düzey düşünme becerileri hakkında Bloom Taksonomisi ve Revize Edilmiş Taksonomiye (the Revised Taxonomy) ek olarak, literatürde çeşitli tanımlar yapılmakta ve değişik isimlerle de ifade edilmektedir (Lewis & Smith, 1993). Bunlar eleştirel düşünme, üst düzey bilişsel süreçler, problem çözme, sebeplendirici düşünme veya yaratıcı düşünme gibi isimlerdir. Üst düzey düşünme terimi bu ifadelerin tamamını içeren kapsamlı bir terim olarak ortaya konmuştur (Lewis & Smith, 1993). Maiere'e göre(1937; 1933), üst düzey düşünme her zaman sayısal ifadeler taşıyan bir muhakeme değil, öğrenmeyi içeren ve sürecin bir parçası olmayı gerektiren durumdur.

Brookhart'a (2010) göre yapılan üst düzey düşünme tanımlamaları üç kategoriye ayrılır. Bunlar: transfer açısından üst düzey düşünmeyi tanımlayanlar, eleştirel düşünme açısından üst düzey düşünmeyi tanımlayanlar, problem çözümü açısından üst düzey düşünmeyi tanımlayanlardır.

En önemli eğitimsel hedeflerden ikisi geri çağırma ve transferi arttırmaktır. Bellek, öğrencilerin ne öğrendiklerini hatırlaması, *transfer* ise öğrencilerin sadece öğrendiklerini hatırlaması değil, aynı zamanda öğrendiklerini kullanabilme becerisi anlamına gelmektedir. Geri çağırma için öğrenme düşünmeyi gerektirirken, yüksek üst düzey düşünme "transfer" edilmektedir. Yani, öğrenciler sadece bilgi ve becerileri kazanmakla kalmaz, aynı zamanda bunları yeni durumlara da uygulayabilirler. Brookhart'a (2010) göre düşüncenin bir dizi geri çağırma görevi yerine bir dizi transfer fırsatı ile karakterize edildiği okul dışı yaşam için geçerli olan bu tür bir düşünce vardır.

Eleştirel düşünme kategorisi mantığa, neye inanacağına ya da seçeceğine karar veren düşünceye (Ennis & Norris, 1990) ve akıl yürütmeyi, sorgulamayı ve araştırmayı, gözlemlemeyi ve tanımlamayı, karşılaştırmayı ve bağlantı kurmayı, karmaşıklığı anlamayı ve

bakış açılarını keşfetmeyi içeren sanatsal düşünmeye atıfta bulunan tanımları içermektedir (Barahal, 2008).

Problem çözme kategorisinde Brookhart (2010) aşağıdaki tanımı yapmıştır:

Öğrenci belirli bir sonuca veya hedefe ulaşmak istediğinde ancak buna ulaşmak için kullanılacak uygun yolu veya çözümü otomatik olarak tanımazsa bir problem çıkarır. Çözülmesi gereken problem, istenen hedefe nasıl ulaşılacağıdır. Bir öğrenci istenen hedefe ulaşmak için uygun yolu otomatik olarak tanıyamadığı için, bir veya daha fazla üst düzey düşünme sürecini kullanmalıdır. Bu düşünme süreçlerine problem çözme denir (Nitko & Brookhart, 2007). Bilgiyi hatırlamayı, anlayışla öğrenmeyi, fikirleri eleştirel bir şekilde değerlendirmeyi, yaratıcı alternatifleri formüle etmeyi ve etkili bir şekilde iletişim kurmayı içerebilir.

Problem çözmenin geniş tanımı, bir kişinin sadece ezberle çözülemeyecek bir problem için bir çözüm bulmasını sağlayan beceri olmasıdır. Öğrencilerin belirli bir algoritmayı tekrar tekrar yapmak için hafızayı kullanmasını gerektiren birçok kapalı problem varken, örneğin matematikte birçok problem açık uçludur ve yalnızca hafızadan çözülemez. Ya da birden fazla çözüme sahip olabilirler. Ya da bir cevabın henüz bilinmediği gerçek sorunlar olabilir. Koşullar değiştikçe çözümlerin değişmesiyle de problem açık uçlu olabilir (R. Collins, 2014).

Problem çözme, çıkarım yapma, hesaplama, tahmin yürütme, genelleme ve yaratıcı düşünmenin daha üst düzey düşünme becerileri olduğu düşünülmektedir (Miri et al., 2007). Birçok araştırmacı, her yaştan öğrenciler arasında daha yüksek düzeyli düşünmeyi teşvik etmenin önemli bir eğitimsel hedef olduğuna ve daha üst düzey düşünmenin yaşamsal başarının önemli bir unsuru olduğuna inanmaktadır (Gough, 1991; Marzano, 1998; Zohar, Degani, & Vaaknin, 2001; Sousa, 2008; T. Lord & Baviskar, 2007). Miri ve arkadaşları (2007), sürekli değişen ve zorlayıcı dünyamızın, gelecekteki vatandaşlarımızın, bilgi kapasitelerinin inşasının ötesine geçmelerini ve daha yüksek mertebeden düşünme becerilerini geliştirmeleri gerektiğini şart koşmaktadır.

2.5.1 Üst Düzey Düşünme Becerilerin Ölçülmesi ve Değerlendirilmesi

Bloom'un taksonomisi düşünme becerileri ve hedefleri altı bilişsel karmaşıklık düzeyine göre kategorize etmiş ve sıralamıştır (Bloom, 1956). Buradaki mantık, bazı öğrenme türlerinin diğerlerinden daha fazla bilişsel işlem gerektirmesidir. Bloom'un taksonomisindeki en düşük üç biliş seviyesi bilgi, kavrama ve uygulamadır ve düşük dereceli düşünme becerileridir. En yüksek üç seviye analiz, sentez ve değerlendirmedir ve bir üst düzeye sahip olduğu düşünülmektedir; bunlar için olguların ve kavramların öğrenilmesinden farklı öğrenme ve öğretim yöntemlerine gereksinim vardır (Krathwohl, 2002).

Düşünme, bireyin doğuştan getirdiği hayati bir aktivitedir. Ancak üst düzey düşünme, deneyimin çeşitliliği ve bireyin yaşının ve bilişsel düzeyinin çeşitliliği dahilinde, üst düzey düşünmenin gelişimi ve tecrübenin çeşitliliğine yol açacak yüksek seviyeli sorularla geliştirilebilir. En üst düzey sorular, bir bireyin önceki bilgilere nasıl bakabileceği, yeni bilgilere nasıl bağlandığı, yorum yapabileceği, yeni fikirler yaratabileceği ve bilgiyi ön bilgi ile nasıl değerlendirdiği sorularıdır (Ersözlü & Kazu, 2011). Bununla birlikte, "Neden", "Ne", "Nasıl" ve "Eğer" soruları hem öğrencilerin bilincini oluşturmak hem de üst düzey düşünmeye ulaşmak için yaratıcı ve eleştirel düşünme becerileri geliştirmek için kullanılabilir. (Akt. Bümen, 2006).

Üst düzey düşünmeyi değerlendirmek neredeyse her zaman üç ek ilkeyi içerir(S M Brookhart, 2010):

• Öğrencilere genellikle giriş metni, görseller, senaryolar, kaynak materyalleri veya bir tür problemler şeklinde, düşünmeleri için bir şeyler sunmak.

• Yeni materyal kullanmak - öğrenciye yeni, sınıfta yer almayan ve dolayısıyla geri çağırmaya maruz bıraktıracak materyaller.

• Zorluk seviyesi (kolaya karşı zor) ve düşünme derecesi (alt düzey düşünmeye karşı üst düzey düşünmeyi ya da geri çağırma) ayrımı yapmak ve her birini ayrı ayrı kontrol etmek.

Kolayca erişilebilen tek bir doğru cevabı olan sorular, öğrenilen bilginin değerlendirilmesinde yararlıdır, ancak düşünme yeteneğini çok küçük bir oranda gerektirir. Daha zor sorular, öğrencilerin bilgi erişim becerilerini geliştirmeleri, bilgilerini test etmeleri, sorunları tanımlamaları ve kendilerine çözüm sunmaları için çok yararlıdır. Ayrıca öğrencilerin tüm bilişsel becerilerini geliştirmelerine yardımcı olur (Koray, Altunçekiç, & Yaman, 2005). Değerlendirme süreci sırasında öğrenciler bilişsel düşük seviyeli sorularla karşılaştıklarında, basit bir düzeyde düşünmeye teşvik edilirler, ancak yüksek seviyeli sorularla karşılaştıklarında, daha fazla zihinsel aktivite sergilerler ve bu düzeydeki sorunlar öğrencileri daha yaratıcı olmaya zorlar.

2.5.2 Matematikte Üst Düzey Düşünme ve Gerçek Hayat

Matematik eğitimcileri arasında, birçok öğrencinin 11 yıl veya daha fazla bir süre için matematik öğrendiği, ancak matematik bilgisini sınıf bağlamı dışındaki durumlarda tamamen kullanamadığı yönünde giderek artan bir endişe vardır. Çeşitli araştırma projelerinde yetişkin ve öğrencilere, okulda öğrendikleri matematikten yararlanmalarını gerektiren görevler verilmiştir (Boaler, 1998). Bu projeler, yetişkin ve öğrencilerin okuldan öğrendikleri matematiksel yöntem ve formülleri gerçek dünyadaki matematiksel durumlarda kullanmadıklarını göstermiştir (Lave, Murtaugh, & de la Rocha, 1984; Nunes, Schliemann, & Carraher, 1993).

Prawat, (1996), Henningsen & Stein (1997), Jaworski (2006) ve Libman (2010), yazılarında sınıf aktivitelerinin temellerinin günlük yaşamdan alınması gerektiğine, böylece öğrencilerin günlük yaşamlarında ve gelecekteki çalışma hayatlarında kullanacakları matematik uygulamalarından zevk alacaklarına işaret etmektedir.

"Okul Matematiğinde Problem Çözme" temasının Amerika'da NCTM 1980 "Eylem gündemi" ve İngiltere'de "Cockroft Report" olarak kabul edildiği seksenli yılların başlarından beri, matematiksel tartışmayı teşvik etmek için sınıfta açık uçlu problemleri kullanma yöntemi süregelmektedir. O yıllardan itibaren sınıfta açık uçlu problem çeşitlerinin kullanılması tüm dünyaya yayılmıştır. Bununla birlikte, açık uçlu problemler olarak ortaya konan pek çok problem türü olduğu için "açık uçlu" kavramı iyi tanımlanmamıştır (Foong, 2000).

2.5.3 Açık Uçlu Sorularda Üst Düzey Düşünme Becerisi Ölçümleri

Açık uçlu sorular, cevaplayıcıların kendi ifadelerinde bir cevap formüle etmelerini ve sözlü ya da yazılı olarak ifade etmelerini gerektirir. Yanıtlayıcılar, önceden tanımlanmış yanıt kategorileriyle belirli bir yönde yönlendirilmez. Aşağıdaki durumlarda bilgi toplama amaçlı kullanılırlar (Züll, 2016):

• Bilgi ölçümü

• Bilinmeyen olası cevap aralığı

• Çok fazla doğru cevap ihtimali olan sorulardan kaçınmak

• Yönlendirici sorulardan kaçınmak

• Bilişsel ön test

Açık uçlu sorular, bilgiyi ölçmek için kapalı uçlu sorulardan daha uygundur, çünkü hem cevaplayanları doğru cevabı tahmin etmeye çalışmakla sınırlamazlar, hem de daha güvenilir ve geçerli bilgi getirisi de sağlarlar. Ancak, cevaplayanlar, doğru cevabı hemen hatırlayamadıklarında, açık uçlu sorular, kapalı sorulardan daha çok yanıtsız bırakılırlar (bkz. Krosnick & Presser, 2010).

Açık uçlu soruların kullanımı, anket konusunu net bir şekilde sınırlamak henüz mümkün değilse veya biri ortaya çıkacak yeni konular bekliyorsa tavsiye edilir.

Açık uçlu sorular için başka bir kullanım alanı, sorunun olası cevaplarının önceden belirtilmiş cevap kategorileri ile sorulamayacağı kadar uzun olduğu durumlardır.

Açık uçlu sorular, cevaplayanlar belli bir yöne yönlendirmekten kaçınmak için kullanılabilir. Porst (2011) bunu bir soru örneğiyle bunu açıklamıştır, "Sizce tüm zamanların en iyi rock grubu kimdir?" Cevap kategorileri en iyi bilinen gruplardan oluşmaktadır ve araştırmacının kendi

düşüncelerini yansıtır. Bu tür seçenekler cevaplayanı belli bir yöne yönlendirir ve verilen cevaplayanın gerçek seçimini ya da görüşünü yansıttığı için değil, sunulan tek seçenek oldukları için seçilmiş olma tehlikesini doğururlar.

Açık uçlu sorular, belirli bir ölçek değerinin neden seçildiğini veya bir sorunun anlamının amaçlandığı şekilde anlaşılıp anlaşılmadığını kontrol etmek için bilişsel ön testlerde sıkça kullanılır (ayrıca bkz. GESIS Anket Rehberi katkıları "Bilişsel Ön Değerlendirme"; Lenzner, Neuert, & Otto, 2016).

Açık uçlu soruların öğrencilerin eleştirel düşünme becerilerini kazanmaları için önemli olduğu anlaşılmaktadır. Hirose (1992)'ye göre, öğrencilere problemleri çözmek, soru sormak ve kendi düşüncelerini ve inançlarını bir sünger gibi bilgi emmek yerine diğer insanların düşünceleri ve inançlarıyla karşılaştırmaları öğretilmelidir. Ayrıca, öğrencilere yöneltilen sorulara cevap vermeyi beklemenin süresi de önemlidir. Örneğin, cevaplamak için 3-5 saniye beklemek öğrencilerin daha yüksek düzeyli cevap vermesine izin vermektedir (Carrol, 1989).

Açık uçlu sorulara verilen cevaplarla, öğrencilerin nasıl analiz ettiğini, sentezlediğini ve değerlendirdiğini (McKeachie, 1987b); verilen problemleri hangi stratejilerle çözdüklerini ve nasıl hipotez oluşturduklarını görmek mümkündür. Stodolsky, (1985), matematiğe dair inançların, öğrencilerin (ve öğretmenlerin), özellikle sosyal bilgiler sınıflarına kıyasla, ilkokul matematik derslerinde nasıl performans gösterdiğini açıklar. Sosyal çalışmalarda, öğrencilerin gruplar halinde çalışma, araştırma becerilerini geliştirme ve genel olarak, üst düzey düşünme becerilerinin gelişmesiyle uyumlu görevler üzerinde çalışma olasılıkları daha yüksektir. Aksine, matematik dershanelerinde öğrenciler "oturmak" için çok zaman harcarlar (McLeod, 1992). Diğer yazarlar, öğrencilerin matematiği nasıl yetenek odaklı bir konu olarak gördüklerini ve matematiğe karşı bu gibi sınırlı görüşlerin, matematiğe ilişkin kaygıya nasıl yol açtığını (Greenwood, 1984) ve daha genel olarak matematikte üst düzey düşünmeyi engellediğini belirtmiştir (Garofalo, 2018).

Matematiğe karşı güçlü duygusal tepkilerin raporları, araştırma literatüründe sıklıkla görülmemektedir. Buxton'nun (1981) çalışması bu konuda önemli bir istisnadır. Bu araştırma matematiksel görevlere duygusal tepkilerini panik olarak rapor eden yetişkinlerle ilgilidir. Panik raporlarına yüksek derecede fizyolojik uyarılma eşlik etmektedir; bu uyarılma, görevlere konsantre olma yeteneklerini bozduğu ve kontrol edilemediği görülmüştür. Katılımcılar duygusal tepkilerini korku, endişe, utanma ve panik olarak tanımlamaktadır (McLeod, 1992).

Çalışmalar, öğrencilere entelektüel çalışma ve eleştirel düşünme gerektiren ödev ve değerlendirmeler kullanarak üst düzey düşünmeden sorumlu tutmanın, öğrenci motivasyonunu ve başarıyı artırdığını göstermiştir. Öğrencilerin soyut kavramlar ile yaptıkları

çalışmalara çok ilgi göstermedikleri ve motive olmadıkları görülmektedir. Daha çok, bir şeyler düşünmeye ve belirli şeyleri öğrenmeye motive olmaktadırlar. Bu noktada üst düzey düşünme, öğrencilerin fikir üzerinde kontrol duygusunu artırmaktadır. Özetle düşünmenin ezberlemekten daha eğlenceli olduğu söylenebilir.

Açık uçlu sorular, eleştirel düşünme, yaratıcı düşünme ve problem çözme gibi üst düzey becerilerin değerlendirilmesinde yararlıdır. Öğrenciler genellikle açık uçlu sorulara verdikleri cevapların yüksek kaliteye sahip olduğunu düşünürler çünkü bu sorulara verilen cevaplar daha fazla düşünme yeteneği gerektirmektedir (Herman, Klein, & Wakai, 1997). Ancak, daha fazla zihinsel aktivite gerekli olduğu için, bu formattaki sorular öğrencilerin kaygısının da artmasına neden olmaktadır (O'Neil Jr. & Brown, 1998).

Öğretmenler genellikle bu tür soruların incelenmesi sırasında öğrencilerin çabalarını ve davranışlarını dikkate alma eğilimindedir (Susan M Brookhart, 1994). Böyle bir karar, bu sınavları objektif olarak değerlendirmeyi zorlaştırmaktadır. Öte yandan, devamsızlık ve çaba gibi ders dışı faktörlerin sınavda öğretmen tarafından değerlendirilebileceğini düşünen bir öğrencinin, derse motivasyonunu artabilir, böylece ders başarısı artarak kaygı düzeyi azalabilir.

McKeachie (McKeachie, 1987a; Naveh-Benjamin, McKeachie, Lin, & Tucker, 1986), öğrencilerin aynı ders için uzun cevaplar gerektiren açık uçlu sorular içeren sınavlara çoktan seçmeli soruları olan sınavlardan çok daha fazla çalıştıklarını belirtmektedir. Bu durumda, ne tür bir testin uygulandığı öğrencinin derse karşı tutumunu değiştireceği söylenebilir. Açık uçlu soruları olan bir sınav için çalışırken, daha fazla kavramsal temel çalışmanın gerekliliği ve öğrenmeye daha fazla çaba ve zaman harcanması, muhtemelen daha verimli bir çalışma dönemini sağlayacaktır (Aydın & Önder, 2010).

2.6 Üstbilişsel İzleme

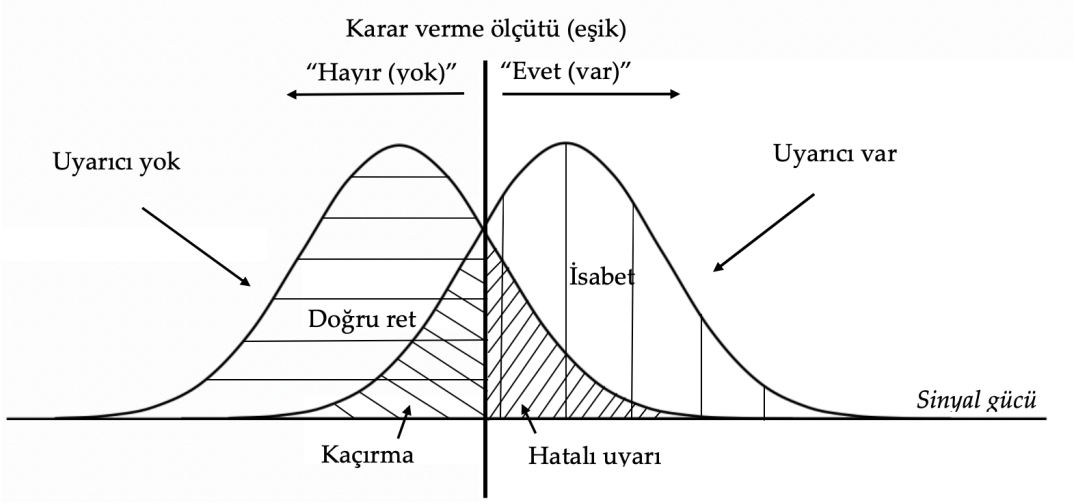
Kısaca "bilginin bilgisi" olarak ilk kez Flavell (1979) tarafından tanımlanan üstbiliş, giderek artan sayıda araştırmacının ilgisini çekmiş ve bu çalışmalar, örneğin, öğrenme kolaylığı (ease of learning [EOL]), öğrenme kararı (judgment of learning [JOL]) veya bilme hissi (feeling of knowing [FOK]) gibi becerilere ait standart tekniklerle çalışabilmiştir. Bütün bu çalışmaların ortak özelliği ve sonuçlarının bize sunduğu en temel çıktı ise, özellikle bilişsel becerilerin araştırıldığı çalışmalarda, bireylerin bilgileri hakkında ne düzeyde farkındalığa sahip olduklarını görgül olarak ölçümleyebilmektir. Bu anlamda, bilişsel becerilerin incelendiği çalışmalarda, üstbilişe ait etkilerin ortaya çıkardığı üstbilişsel performansların araştırılması, bireylerin örneğin, yalnızca hangi bilgileri doğru bilip bilmediğini araştırmaktan öte (ör., testten elde edilen toplam doğru yanıtlama sayısı), sahip oldukları doğru ya da yanlış yanıtlar veya bilgiler

hakkında ne düzeyde farkındalığa sahip olduklarını da belirleyebilmek için oldukça önemlidir. Diğer bir deyişle, bilgiye sahip olmaktan ya da yalnızca doğru bilmemekten öte, raporlanan bilgilerin doğru ya da yanlış olduklarına dair farkındalığın ne düzeyde olduğunu ölçmek, araştırmacılara bireylerin doğru yanıt sayılarını hesaplamanın ötesinde veriler sağlayabilmektedir. Üstbilişin ölçülmesi, çeşitli tekniklerle yapılabilmesine rağmen (ör., EOL, JOL, FOK becerilerine ait ölçümler), günümüzde yaygın olarak Green ve Swets'in (1966) önerdiği Sinyal Tespit Kuramı ile, özellikle Tip-2 olarak adlandırılan STK ile de standart şekilde ölçülebilmektedir. STK'nın önerdiği temel varsayımlar temelinde, katılımcıların "isabet oranları (hit rate)" ve "hatalı uyarı (false alarm)" hesaplanarak bu katılımcılara ait üstbilişsel izleme performansları ve yanıtlarını verirken ne düzeyde katı ya da esnek davrandıklarını işaret eden "raporlama kriteri (report criterion)" ölçülebilmektedir.

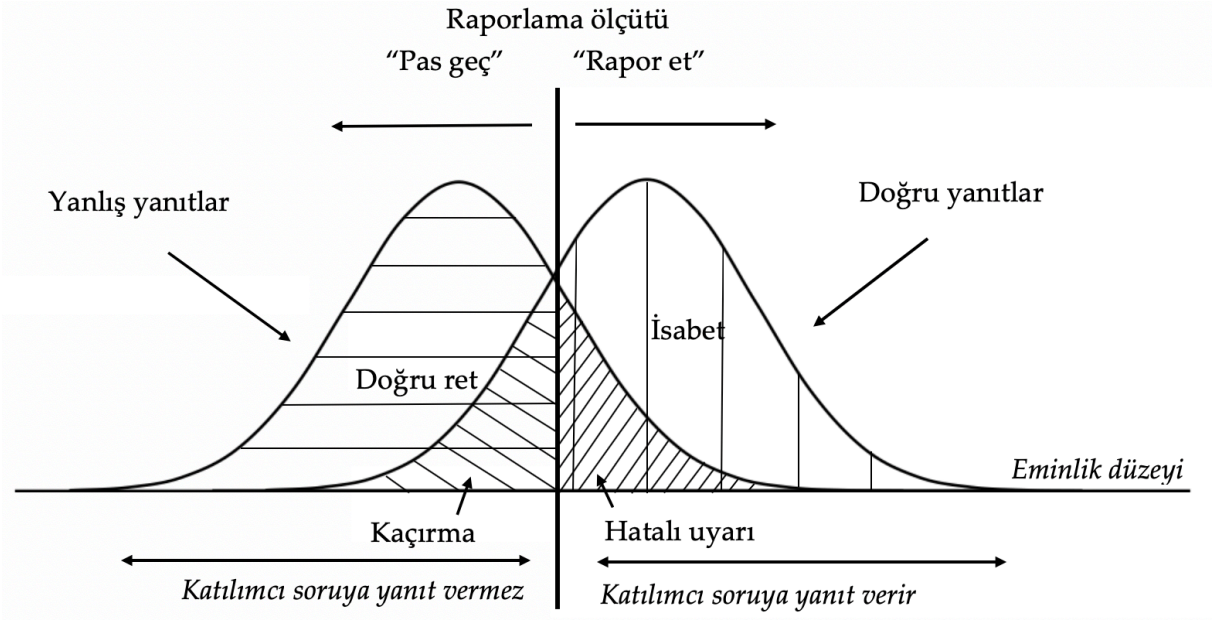
2.6.1 Sinyal Tepki Kuramı ve Üst Bilişsel İzleme Ölçümleri

Fiziksel uyarıcı ve bunun duyum ve algı ile olan ilişkisine odaklanan "psikofizik" (Luce & Krumhansl, 1988) bakış açısına dayanan Sinyal Tespit Kuramı (Green & Swets, 1966) Tip-1 ve Tip-2 olarak iki türde değerlendirilebilir (ör., Higham, 2002; Higham & Tam, 2005). Standart olarak ölçülen Tip-1 STK'na göre, katılımcıdan, örneğin, beyaz ses (white noise) olarak tanımlanan gürültülü bir ses içine gizlenen sinyali tespit edip edemediğini "evet-hayır" olarak yanıtlaması istenir. Sinyal, bazı denemelerde gürültünün içinde gerçekten de bulunurken, bazı denemelerde böyle bir sinyal bulunmaz. Kurama göre, sinyalin bulunduğu ve bulunmadığı denemelerin iki normal dağılım oluşturdukları varsayılır ve katılımcı sahip olduğu raporlama kriterinin (eşiğin) üzerinde kalan denemeler için sinyal varlığına 'evet' yanıtı verirken, bu eşiğin altında kalan denemelerde için sinyal varlığına ise 'hayır' yanıtı verdiği varsayılır. Sinyalin gerçekte bulunduğu ve bulunmadığı durumlar (sinyal: var-yok) ile katılımcının bu iki durum için verdiği "evet-hayır" yanıtlarına ait olasılıkların türü belirlenebilir. Buna göre, isabet oranı, katılımcının "var" yanıtı verdiği denemelerdeki sayının, gerçekten bir sinyalin bulunduğu bu denemelerin sayısına oranıdır. Hatalı uyarı oranı ise, katılımcının yine "var" yanıtı verdiği denemedeki sayının, gerçekte böyle bir sinyalin bulunmadığı denemelerin sayısına oranıdır. Ancak katılımcılar, bir sinyal bulunmazken uygun şekilde beklendiği üzere "yok" yanıtını da verebilirler. Bu denemelerde ölçülen oran ise, doğru ret (correct rejection) olarak tanımlanırken, katılımcılar yine gerçekte bir sinyal bulunmasına rağmen yanıtlarını 'yok' olarak verebilirler. Son duruma ait olan oran, bu türdeki denemeler için kaçırma (miss) olarak tanımlanır. Özetle, Tip-1 olarak adlandırılan bu ölçümde, katılımcının aslında bir makina gibi pasif bir rol edinerek sinyalleri doğru şekilde tespit edip etmediğine odaklanılır ve bu sinyal tespit yöntemi, uyarıcı-bağımlı (stimulus-contingent) olarak nitelendirilir (Higham, 2007).

Tip-2 sinyal tespit kuramına göre ise, yukarıda bahsedilen oranların isimleri aynı kalmakla birlikte, ölçülen değerlerin özellikleri farklıdır. Buna göre, katılımcı örneğin yöneltilen bir soru için yanıtlarını oluştururken zihninde yanlış ve doğru yanıt olabilecek listeler üretir. Diğer bir deyişle, varsayımsal olarak bir bilginin doğrudan, örneğin bir vakum gibi, depolama alanından çağrılmasından ziyade, yanıt vermeden önce yanlış ve doğru yanıtları içeren yanıtlarının bilince aktığı varsayılır (ör. bkz., hatırlamada üret-tanı (generate-recognize) modeli; Bahrick, 1970; Kintsch, 1970; gözden geçirme için ör., Watkins & Gardiner, 1979). Tip-2 STK, yanlış ve doğru yanıtların sinyal güçleri ve dolayısıyla eminlik düzeyleri bakımından normal dağıldıklarını varsayarak, bu dağılımların ortalama değerleri arasındaki farkın hesaplanmasıyla, bireyin doğru ve yanlış yanıtları ne düzeyde ayırt edebildiklerine dair farkındalıklarını ölçmeyi hedeflemektedir. Tip-2 STK'daki raporlama kriteri ise, katılımcının ürettiği yanıtları, bu kriterin üzerindeyse rapor etmek istemesini belirler. Diğer bir deyişle, üretilen yanıtlar eğer bu kriterin üzerindeyse, katılımcı yanıtının doğru olduğunu düşünmekte ve rapor etme (soruya yanıt verme) eğilimindedir. Bu varsayımdan da çıkarsanabileceği gibi, standart olarak katılımcıya doğru yanıtını vermeden önce, örneğin soruyu "yanıtlayacağım"-"yanıtlamayacağım" (çalışmalarda sıklıkla kullanıldığı gibi "raporlayacağım" ya da "pas geçeceğim" [report/pass]) şeklinde bir yanıt vermesi istenir. Katılımcı "yanıtlamayacağım" seçeneğini vermiş olsa da katılımcıdan sorunun yanıtına ait en iyi tahmininin ne olduğunu belirtmesi istenir. Bu durumda, örneğin katılımcı bazen yüksek bir raporlama kriteri koymuş ve gerçekte doğru olabilecek yanıtı için "yanıtlamayacağım" seçeneğini belirtmiş olabilir. Yani, katılımcının yanıtından yüksek derecede emin olduğu sorulara yanıt verme eğilimde olduğu düşünülebilir. Son olarak, katılımcıdan verdiği yanıtın doğru olduğundan ne kadar emin olduğunu Likert-tipi bir ölçekte belirlemesi istenir. Yanlış üretilen yanıtlara ait dağılım tarafında (Şekil 1 soldaki dağılım) düşük eminlik derecesinden başlamak üzere, doğru üretilen yanıtlara ait dağılım tarafındaki (Şekil 1 sağdaki dağılım) eminlik düzeyleri yüksek değerlere sahiptir. Özetle, bireyin doğru ve yanlış yanıtlarını ayırt edebilme becerisi, doğru ve yanlış olarak üretilen yanıtlara ait dağılımların birbirinden ne kadar uzak şekilde dağıldığına ve raporlama kriterinin katılımcı tarafından hangi düzeye yerleştirildiğine bağlıdır (katı veya esnek raporlama kriteri). Tüm bu varsayımlar temelinde, Tip-1 STK uyarıcı-bağımlı iken, Tip-2 STK ise tepki-bağımlıdır (response-contingent; (Higham, 2002)).



Şekil 1. Tip-1 Sinyal Tespit Kuramı'na göre elde edilebilecek olasılıklar

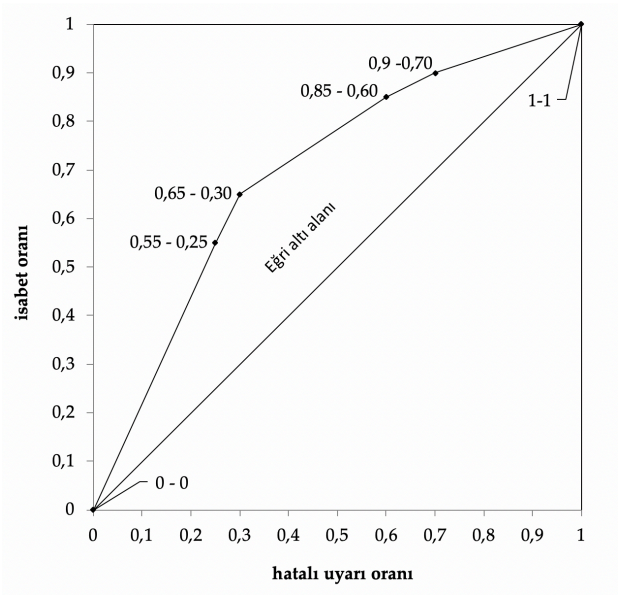


Şekil 2. Tip-2 Sinyal Tespit Kuramı'na göre elde edilebilecek olasılıklar

Şekil 1. Tip-1 ve Tip-2 Sinyal Tespit Kuramı'na göre elde edilebilecek olasılıklar.

Tip-1'de bireyin uyarıcı var veya yok demesindeki eşik, karar verme ölçütü (decision criterion) olarak adlandırılırken, Tip-2'de bu eşik raporlama ölçütü (report criterion) olarak adlandırılır. Yine Tip-1'de uyarıcının bulunmadığı (diğer bir ifadeyle, gürültü) ve uyarıcının bulunduğu (uyarıcı var) denemeler sinyalin gücü üzerinde dağılırken, Tip-2'de yanlış ve doğru yanıtlar eminlik düzeyine göre dağılırlar(Higham & Arnold, 2007).

Üstbilişsel izleme becerisinin düzeyini belirlemek, zihinde üretilen doğru ve yanlış yanıtlara ait iki dağılım arasındaki uzaklığın isabet ve hatalı uyarı oranları belirlendikten sonra hesaplanmasıyla mümkündür. Bu amaçla, araştırmacılar yanlış ve doğru olarak üretilen yanıtlara ait iki dağılımın tepe noktaları arasındaki uzaklığı, örneğin "*d*'" (d-prime) adıyla hesaplayabilmektedir (*d*'=Z(isabet oranı)-Z(hatalı uyarı)). Bu değer, karar vericinin etkinlik özellikleri (receiver operating characterictics [ROC]) yöntemiyle de gösterilebilir. Bu amaçla, katılımcının her bir yığılmalı eminlik değerindeki ve "0.00" ve "1.00" arasında değişen isabet ve hatalı uyarım oranları hesaplanarak, bunların kesişim değerleri saçılım grafiğine işaretlenir; bkz., Şekil 3. İşaretlenen noktalar ile, hiçbir bilinçli ayrımı işaret etmeyen (diğer bir ifadeyle %50 olasılıkla da zaten doğru cevaplanabilecek raporlamayı ifade eden) "0.50" değerindeki bütün isabet ve hatalı uyarım kesişimlerinden elde edebilecek varsayımsal çizginin üzerinde kalan alan hesaplanabilir. Bu alan, ne kadar büyük ve pozitif bir değere sahipse, katılımcının, alan altı eğrisi değeri (dolayısıyla *d*' değeri de) -yani doğru yanıtlarını doğru, yanlış yanıtlarını da gerçekten yanlış yanıtlar olarak fark ettiğine dair puanı- o kadar yüksektir. Eksi bir değer ise, katılımcının doğru yanıtlarını yanlış, yanlış yanıtlarını ise doğru olarak düşündüğünü ifade eder. Şekil 3 farazi bir katılımcının eğri altı alanının hesaplanmasında kullanılabilecek bir ROC eğrisini göstermektedir.



Şekil 3. Eminlik Düzeyi için ROC Grafiği

Hesaplanan alan, kırık çizgilerle birleştirilen eğrinin köşegenel çizgiyle arasında kalan alandır. Likert-tipi ölçeğinde 1 (doğru olduğundan hiç emin değilim) ve 5 (doğru olduğundan kesinlikle eminim) arasında değişen eşit aralıklı bir ölçüme göre, birikimli eminlik değerlerindeki oranlara ait kesişimler, en yukarıdaki kesişim noktasında "1 ve üzeri" eminlik değerlerine ait olup (ör., 1.00 isabet oranı ve 1.00 hatalı uyarı oranı), diyagramın yukarıdan aşağı sayıldığındaki ortaya

çıkan 5. kesişim noktası ise “yalnızca 5” eminlik değerlerinde verilen yanıtlar göz önünde bulundurularak hesaplanan oranların kesişimidir (ör., 0.55 isabet oranı ve 0.25 hatalı uyarım oranı). Diyagramdan da çıkarsanabileceği üzere, eminlik değerlerinde kullanılacak Likert-tipi ölçekte ne kadar fazla değer bulunursa, alan hesaplamasında ortaya çıkan eğri o denli daha az kırık şekilde olacak ve izleme performansı daha kesin olarak ölçümlenebilecektir.

2.7 PISA Testi

Uluslararası Öğrenci Değerlendirme Programı (PISA), sanayileşmiş ülkelerin hükümetler arası bir organizasyonu olan Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından gerçekleştirilmiş ve ilk olarak 2000 yılında uygulanan, 15 yaşındaki öğrencilerin matematik okuryazarlığı, fen okuryazarlığı ve okuma becerilerini 3 yılda bir ölçen uluslararası bir değerlendirme sistemidir. 2015 verilerine göre 35 OECD ülkesi, 37 ortak ülke ve ekonomiyi kapsamaktadır.

Dünya genelinde tüm politikacıların eğitim adına birinci gayesi, vatandaşlarının potansiyellerini tam anlamıyla gerçekleştirmelerine ve onların yeteneklerini, değişen Dünya şartlarına uygun olarak geliştirmelerine olanak sağlamaktır. Bu bağlamda PISA sonuçları, puanlardan veya sıralamalardan çok daha fazlasını ifade eder. Aynı zamanda okul sistemlerinin kalitesini, eşitliğini ve verimliliğini değerlendirmek amacıyla kullanılan dünyanın önde gelen eğitim ölçütlerindendir. Bu durum, hükümetlere ve eğitimcilere, yüksek performanslı eğitim sistemlerinin özelliklerini belirleyerek, kendi öz değerlerini adapte edebilecekleri etkili eğitim politikaları tanımlamalarını sağlamaktadır. (OECD, 2015a)

PISA, Türkiye sonuçlarına bakıldığında beklenilenin de altında ve devamlı azalan bir başarı grafiği söz konusudur. Bu yüzden bu araştırmada, PISA uygulamalarında Türkiye örnekleminin yetenek dağılımının, daha erken eğitim seviyelerindeki benzerlikleri ve farklılıkları incelemiştir. Araştırma kapsamında yapılan ölçümler ve elde edilen bulgular, PISA uygulamalarında üst sıralarda yer alan ülkeler ve OECD ortalamaları ile karşılaştırılmıştır. Bu şekilde Türkiye’nin PISA uygulamalarında var olan başarı düzeyini diğer eğitim kademeleri için genelleştirmek ve beklenenin altındaki başarının gerekçeleri ile ilgili bulgulara ulaşılmak amaçlanmaktadır.

2.7.1 Pisa Örneklemi

PISA iki aşamalı oluşturulmuş tabakalı ve seçkisiz bir örneklem tasarımına sahiptir. İlk tabakada okullar, ikinci tabakada ise okullardaki öğrenciler örneklenir. Örneklemede ilk adım, PISA öğrencilerinin hedef popülasyonunun tanımlanması ihtiyacıdır. Genellikle 15 yaşındakiler olarak kabul edilse de daha da ayrıntılı ifade edilirse 15 yaş ve +3,-3 arasında tamamlanmış aylar ve 16 yaşında ve -2,+2 tamamlanmış aylar arasına giren yaş grubu örneklemi

oluşturmaktadır. Her ülkeden alınan örneklem büyüklüğü ise en az 150 okul örneği ve en az 4500 öğrenci olarak belirlenmiştir.

2.7.2 Pisa'da yetenek kestirimi RASCH

PISA öğrenci yetenek düzeylerini belirlerken istatistiksel olarak IRT modellerinde Rasch modeli kullanmaktadır. Madde tepki kuramı(Item Response Theory) , klasik test kuramının zayıf yönlerini giderebilecek matematiksel modeller sunan bir yaklaşımdır. Özellikle "örneklemden bağımsız madde parametre" kestirimi ve " test/madde güçlüğünden bağımsız yetenek kestirimi" yapabilme iddiasından dolayı psikometristler tarafında kullanılma eğilimi artan bir kuramdır. Diğer yandan çok sayıda model sunması nedeniyle farklı ölçme sonuçlarına MTK analizleri uygulama olanağı bulunabilmektedir. Aynı zamanda bireylerin yetenek parametreleri ve maddelerin güçlük parametrelerini aynı ölçek düzeyinde kalibre edebildiğinden soruların güçlük düzeyiyle bireylerin yetenek düzeylerini karşılaştırma olanağı da sunarak bireyler ve test maddeleri hakkında doğru çıkarımlar yapmayı kolaylaştırmaktadır. MTK parametre sayısına göre 1, 2, 3 ve 4 parametreli lojistik modeller; boyut sayısına göre tek ve çok boyutlu modeller; puanlanışına göre ikili ve çoklu (çok kategorili) puanlama modelleri vb. farklı modeller sunmaktadır. Rasch Modeli, tek parametreli bir model olarak belirlenmiştir çünkü madde özellik eğrileri yalnızca madde güçlüğüne bağlıdır. Üç parametreli lojistik modelinde, maddenin karakteristik eğrileri aşağıdakilere bağlıdır: (i) Madde güçlük parametresi; (ii) madde ayırt edicilik parametresi; Ve (iii) "tahmin" parametresi olarak adlandırılabilir. Bu son parametre, çoktan seçmeli testte, tüm öğrencilerin, madde ne kadar zor olursa olsun, maddeye doğru cevap verme olasılığı ile ilgilidir.

Rasch Modeli, hem madde güçlüğü hem de öğrenci yeteneği bulunan simetrik bir süreklilik oluşturmak üzere tasarlanmıştır. Madde güçlüğü ve öğrenci yeteneği lojistik bir işlevle bağlantılıdır. Bu işlevle, bir öğrencinin bir maddeyi doğru yanıtlaması olasılığını hesaplamak mümkündür. Dahası, bu olasılık bağlantısından ötürü, tüm madde dizilerini her öğrenciye uygulamak şart değildir. Bazı bağlantı (anchor) maddeleri garanti edilirse, Rasch Modeli her maddenin ve her öğrencinin bulunduğu bir ölçek oluşturabilir. Rasch Modeli'nin bu son özelliği, bu modelin eğitim araştırmalarında ve özellikle PISA uygulamasında temel olmasının başlıca nedenlerinden birini oluşturmaktadır (Edition, 2009).

Rasch dikatomik verileri kullanarak öğrenci yeteneğini sürekli olarak tanımlayabilir. Üç temel ilke ile Rasch sürekliliğinin inşasının temeline inebiliriz.

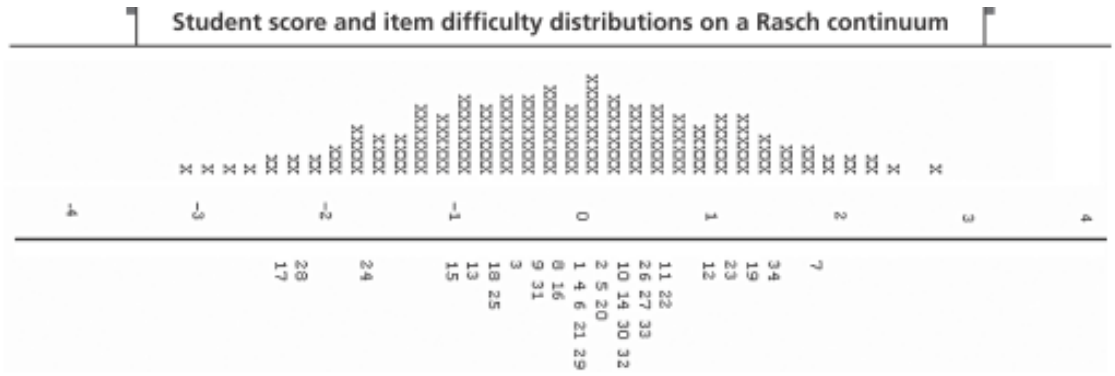
İlk ilke madde zorlukları ile ilgilidir. Örneğin iki sorudan oluşan iki madde ele alalım. 1 ve 2 numaralı maddelere verilen yanıtların deseni (0, 0) ve (1, 1) (1 başarı ve 0 başarısızlığı gösterir) olursa bu iki madde için zorluk karşılaştırması yapamayız. Öte yandan (1, 0) ve (0, 1) şeklinde elde edilen yanıt deseni karşılaştırma açısından bilgilendiricidir. Bu şekilde olan yanıt deseninde 50 öğrencinin (0, 1) şeklinde, 10 öğrencinin ise (1, 0) şeklinde yanıt verdiğini

varsayarsak ikinci maddenin ilk maddeden daha kolay olduğu sonucuna ulaşabiliriz. Aslında, 50 öğrenci ilk maddeye yanlış cevap verirken, ikinci maddeye doğru cevap vermiş oldu ve sadece 10 öğrenci ilk maddeyi doğru yanıtlarken ikinci maddeye yanlış cevap vermiş oldu. Bu durum, bir kişinin iki maddeden birini doğru yanıtladığında bu doğru yanıtlanan sorunun ikinci madde olması olasılığının birinci madde olması olasılığının 5 katı olduğunu gösterir. Bu yüzden ikinci soruyu doğru yanıtlamak, birinci soruyu doğru yanıtlamaktan daha kolaydır. Ancak iki maddenin göreceli olarak zorluğunun öğrenci yeteneklerinden bağımsız olduğunu göz ardı etmemeliyiz.

İkinci ilke referans noktası tanımlama ile ilgilidir. Rasch Modelinde ölçüm birimi, madde güçlüğü ve öğrenci yeteneğinin parametrelerini içeren olasılık fonksiyonu ile tanımlamıştır. Bu sebep ile de sadece bir referans noktası tanımlanması gerektiği kabul edilmiştir. En yaygın referans noktası madde zorluklarının sıfıra ortalanmasıdır. Bununla birlikte, öğrencinin yeteneğinde sıfırı merkez kabul etmek diğer bir izafi referans noktası olarak kullanılabilir.

Üçüncü ilke ile sürekliliğe vurgu yapılır. Kısmen farklı alt popülasyonlara sunulan maddelerin nispi zorluğunun hesaplanmasında süreklilik rol oynar. İlk maddenin tüm öğrencilere, ikinci maddenin de sadece düşük yetenekli öğrencilere verildiğini varsayalım. Maddelerin karşılaştırılması sadece her iki madde de çalışılan alt popülasyonda, yani düşük yetenekli öğrenci popülasyonu üzerinde gerçekleştirilecektir. Göreceli iki maddenin zorluğu bu öğrencilerin ortak alt grubuna dayanacaktır. (Edition, 2009)

Madde güçlükleri Rasch sürekliliğine yerleştirildiğinde, öğrenci puanları hesaplanabilir.



Şekil 4. Madde güçlük ve öğrenci puanlarının Rasch sürekliliğindeki dağılımı

Şekil 4'deki çizgi, Rasch sürekliliğini temsil etmektedir. Madde güçlükleri bu çizginin üstünde, madde numaraları çizginin altında bulunur.

Örneğin, madde 7, zor bir maddeyi ve madde 17, kolay bir maddeyi temsil eder. Bu test birkaç kolay madde, çok sayıda orta derecede madde ve birkaç zor madde içerir. Çizginin üstündeki x sembolleri, öğrenci puanlarının dağılımını temsil eder. (OECD, 2009a)

2.7.3 Rasch Modelde öğrenci puanının hesaplanması

Madde güçlükleri Rasch ölçeğinde belirlendikten sonra, öğrenci puanları hesaplanabilir. Bir öğrencinin B_i ile gösterilen yeteneği ile zorluk derecesi D_j olan j maddesine doğru bir cevap verme olasılığı şöyledir:

$$p(X_{i,j} = 1|\beta_i, \delta_i) = \frac{\exp(\beta_i - \delta_i)}{1 + \exp(\beta_i - \delta_i)}$$

Benzer şekilde yanlış cevap verme olasılığı ise ;

$$p(X_{i,j} = 0|\beta_i, \delta_i) = \frac{1}{1+\exp(\beta_i - \delta_i)}$$

şeklindedir.

Rasch Modeli maddelerin bağımsızlığını varsaymaktadır, yani doğru bir cevabın olasılığı diğer maddelere verilen cevaplara bağlı değildir. Sonuç olarak, iki maddeden başarılı olma ihtimali, iki başarı olasılığının çarpımına eşittir.

Rasch yetenek tahminleri genellikle maksimum olasılık tahmini (veya MLE) olarak belirtilir. Bu şekillerde gösterildiği gibi, her ham puan için, Rasch Modeli yalnızca bir maksimum olasılık tahmini belirlemektedir.

Warm, bu maksimum olasılık tahmininin yanlı olduğunu ve bu maddenin verebileceği bilgilere göre her maddenin katkısını ağırlıklandırmak için önerildiğini göstermiştir (Warm, 1989). Warm tahminleri ve MLE'ler, öğrencilerin bireysel yetenek tahminleri ile benzerdir.

Warm tahmini, MLE'deki küçük yanlılığı düzelttiğinde, genellikle bir kişinin yeteneğinin tahmini olarak tercih edilir. Bu nedenle, PISA'da, ağırlıklandırılmış olabilirlik tahminleri (WLE'ler), MLE'de önerdiği gibi yanlılığı hesaba katmak için MLE'ye ağırlık uygulanarak hesaplanır (OECD, 2009a).

2.7.4 Olası (Plausible) Değerler

Bir eğitim testinden plausible değerler üretmek, posterior dağılımlardan rasgele sayılar çizmekten oluşur. En temel anlamda “Plausible değerler bir öğrencinin makul derecede sahip olabileceği yeteneklerin bir göstergesi olduğunu söyler. Bir öğrencinin kabiliyetini doğrudan tahmin etmek yerine, bir öğrencinin Q için bir olasılık dağılımını tahmin edilir. Yani, WLE gibi bir Q için bir nokta hesabı almak yerine, bir öğrencinin Q için olası değerlerin bir aralığı ve bu değerlerin her biri için birleşik olasılıkla tahmin edilmektedir. Plausible değerler, bir öğrencinin Q’su için bu (tahmini) dağılımından rasgele çizgilerdir.” (Wu ve Adams, 2002).

Bütün bu metodoloji, süreksiz değişkenlerden oluşan bir kümeden (diğer bir deyişle test skoru) bir süreklilik oluşturmayı amaçlamaktadır. Göreceli olarak az sayıda madde kullanarak bir test yoluyla gözlemlenemeyen altta yatan yeteneği ölçmenin sonucu olarak yanlı çıkarımların önlenmesi amaçlanmıştır.

Son olarak, öğrenci yeteneğinin bireysel bir tahmini de posterior dağılımlardan türetilebilir. Bu türetilmiş bireysel tahmine, beklenen bir posteriori tahmincisi (EAP) denir. Posterior dağılımlardan bir dizi rasgele değer atamak yerine posterior dağılımların ortalaması verilir. Bu nedenle, EAP, belirli bir öğrenci için makul değerler grubunun ortalaması olarak düşünülebilir (Edition, 2009).

2.7.5 PISA Yeterlik Düzeyleri

PISA yeterlik düzeyleri, bilişsel süreçleri ve öğrenci bilgi düzeyini geniş ölçekli değerlendirme durumlarında karşılaştırmalı araştırmalar için kullanılmaya imkan tanıyan güçlü bir araçtır (Fischbach, Keller, Preckel, & Brunner, 2013).

Yeterlik düzeylerinin oluşturulmasında ilk aşama uzmanların her bir alanda raporlaştırma için kullanılabilecek muhtemel ölçekleri ve boyutları yazıya dökmesi ile başlamaktadır. Bu işlemin avantajı döngülerde yoğunlaşılan ağırlıklı alan için geliştirilen çoklu ölçeklerin geri bildirim ve raporlama amaçları için daha anlamlı, potansiyel olarak daha kullanışlı olmasıdır. İkinci aşama maddelerin ölçeklendirilmesi ile ilgilidir. Yani her bir soru maddesi, düşünülen bir ölçek ile ilişkilendirilir. Daha sonra uzmanlarca her maddenin özellikleri, değerlendirme çerçevesindeki sınıflandırmalara göre değerlendirilir. Ardından pilot uygulamadan elde edilen madde puanlarının istatistiksel analizi, maddelerin ölçekteki dağılımı ile ilgili objektif bir ölçüt elde etmek için kullanılır. Üçüncü aşamada beceriler denetlenir. Bu aşama her maddenin konu alanının, değerlendirme çerçevesindeki ilgili alt ölçeğin tanımı ile ilişkilendirilerek uzman tarafından ayrıntılı bir şekilde analiz edilmesini, kısmi puan verilen durumların ve puan basamaklarının değerlendirilmesini içerir. Her bir puan basamağına ulaşmak için gerekli bilgi ve beceriler tanımlanır ve açıklanır. Dördüncü aşamada pilot uygulama verilerinin analiz edilmesi söz konusudur. Öncelikle pilot uygulamadan elde edilen veriler madde tepki kuramına göre analiz edilir ve her bir başarı eşiği için madde güçlüğü hesaplanır. Genelde maddeler için tek bir başarı eşiği söz konusu iken kısmi puanlama gerektiren sorularda birden fazla başarı eşiği hesaplanabilir. Sonrasında ise her bir ölçek içinde başarı eşikleri, öğrenci becerileri ile ilişkilendirilerek sürekli bir güçlük süreci boyunca yerleştirilir. Beşinci aşama boyutların tanımlanmasını içerir. Alan uzmanı 3. ve 4. aşamada yapılan analizlerden elde edilen sonuçları birleştirir. Sonrasında her bir ölçek kümesi için madde puan basamakları ile ilişkilendirilen eşikler referans alınarak sıralanır ve daha sonra ilgili bilgi ve becerilerin açıklamaları ile bağlanır. Bu işlemler sonucu boyutu tanımlayan bilgi ve becerilerin bir hiyerarşisi oluşturulur. Altıncı aşama temeli kapsar. Bu aşamaya gelindiğinde artık esas uygulamadaki veriler kullanılabilir olduğundan, madde eşiklerinin bağıl güçlüğü ile ilgili istatistiksel analizlerden elde edilen bilgiler güncellenir. Sonrasında uzmanlarca gözden geçirme ve kontrol söz konusudur. Yedinci aşama geçerlik işlemini içerir. İlk olarak PISA maddeleri için tanımlanmış düzeylerin dayandığı göstergelerin değerlendirilebilmesini sağlayan gerekli materyallere sahip bilgili

uzmanlar istihdam edilir. Sonrasında ise tanımlanmış ölçeklerin tüm PISA ülkelerinin ulusal koordinatörlerinin dâhil olduğu danışma sürecine tabi tutulur. Bu aşama, tanımlanmış düzeyleri kullananların onları ne derece bilgilendirici bulduğu ile ilgili sonuçlara ulaşmayı sağlar(Fischbach et al., 2013; OECD, n.d.).

PISA Yeterlilik düzeylerinin açıklamalarını, her dönem gözden geçirir ve günceller; bu da, değerlendirme ve değerlendirme için geliştirilen yeni görevlerin çerçevesinde ve taleplerindeki değişiklikleri yansıtacak şekilde belirlenir. Yeterlik düzeylerinin en son açıklaması PISA 2012 değerlendirmesine dayanmaktadır (OECD, 2014). PISA sonuçları, en yüksek performans gösteren ve en hızlı gelişen eğitim sistemindeki öğrencilerin neler yapabileceğini göstererek, eğitimde mümkün olanı ortaya koymaktadır. (OECD, 2015a).

3 GEREÇ ve YÖNTEM

Araştırma kapsamında kullanılan gerek üst düzey düşünme becerileri gerekse kazanım temelli ölçme araçları proje sürecinde geliştirilmiştir. Bu süreçte amaca uygun soru havuzunu oluşturmak için soru yazarı eğitimleri, testlerin BTM özelliklerine uygun yapılandırılabilmesi için özelliklerin ve q matrislerin belirlenebilmesi için uzman çalışma grupları ve ön son testlerin PISA TIMSS sınavları ile eşdeğerliğinin kanıtlanması için eşitleme desenlerini içeren analiz yöntemleri kullanılmıştır.

3.1 Test Geliştirme Süreci

Bu süreçte, ölçülecek ve analiz edilecek özelliklerin belirlenmesi, soru yazma ekiplerinin eğitimi ve oluşturulması ve madde havuzunun hazırlanması işlemleri gerçekleştirilmiştir.

3.1.1 Hazırlık aşaması

Projenin ilk bölümünde 6. Sınıf matematik kazanımların ve bilişsel süreç becerilerinin belirlenmesi, Bilişsel Tanı Modellerine (BTM) uygun madde ve test türlerinin belirlenmesi ve soru yazım ekibinin kurulması için gerekli süreçler işletilmiştir.

Bilişsel süreç becerilerinin ve 6. sınıf matematik kazanımların belirlenmesi amacıyla proje ekibi ve aktif olarak görev yapan üç ortaokul Matematik öğretmeni ile bir komisyon kurulmuştur. Bu komisyon yaptıkları toplantılar ile matematik kazanımlarını ve bilişsel işlem süreçlerini belirlemişleridir. Bu toplantılarda bilişsel süreç becerileri, PISA (Programme of International Student Assesment), TIMSS (Trends in International Mathematics and Science Study) frameworklerinde (Mullis, I. V. S., Martin, 2013; OECD, 2015b) belirlenen üst düzey becerileri ve matematik alanında Smith & Stein (1998) tarafından geliştirilen taksonomi dikkate alınarak projenin amaçlarına uygun halde düzenlenmiştir.

BTM analizlerine uygun madde formatları ve test yapısı ile ilgili çalışmalar madde kazanım etkileşimi ve Q matris doygunluğu üzerinde gerçekleştirilmiştir (de la Torre & Chiu, 2016; DeCarlo, 2011). Ulaşılan bulgularda tek bir özellikle ilişkili maddelerin ve birden fazla özellikle ilişkili maddelerin model uyumuna etkilerinin farklılaştığı görülmüştür. Bu farklılaşmaların özellikle birden fazla kazanımla ilişkili maddelerin model veri uyumu üzerinde manidar bir katkısı olduğunu göstermektedir. Bu durum soru yazımı aşamasında birden fazla özellikle ilişkili maddelerin kazanımlarının doğru belirlenmesinin ve Q matris öğrenci örtük sınıfı arasındaki ilişkinin doğrulayıcı bir yöntemle analiz edilmesinin ihtiyacını ortaya koymuştur (Basokcu, Canpolat, & Gönülal, 2016).

Hazırlık aşamasında araştırılan bir diğer önemli konu Q matris geçerliğini belirlemeye yönelik yeni bir yaklaşım bulmaktır. Bu amaçla açımlayıcı örtük sınıf analizi (LCA) yöntemi kullanılarak Q matrisin geçerliği belirlenmeye çalışılmıştır. Açımlayıcı örtük sınıf modeli önsel bir bilgiye

dayanmadan, yapıdaki örtük sınıf sayısını belirlemeyi amaçlamaktadır. Bu yöntemle cevaplayıcıların kaç sınıf içinde yer aldıkları belirlenebilmektedir (L. M. Collins & Lanza, 2013; Dayton & Macready, 2006; McCutcheon, 1987). Çalışmada iki modelin bulgularını simülasyon çalışmalarıyla karşılaştırmayı amaçlamıştır ve iki modeli kullanarak bireylerin benzer yetenek sınıflarına atanıp atanmadıkları gözlenmiştir. Aynı zamanda LCA analizi sonucunda verilen madde cevaplama olasılıklarının başlangıç için belirlenen Q matris ile ne kadar uyumlu olduğu da belirlenmiştir (Basokcu & Güngör, 2016).

3.1.2 Matematiksel yeterlik matrislerinin belirlenmesi

Bu aşamasında üst düzey düşünme becerileri ve bilişsel süreç becerileri, 6 sınıf düzeyi matematik dersi kazanımları ve matematiksel hata örüntülerinin belirlenerek deneysel sürecin değerlendirilmesi için kullanılacak ön ve son testler ile deneysel süreçte kullanılacak izleme testlerinin q matrisleri için gerekli olan özellik listeleri hazırlanmıştır.

3.1.2.1 PISA'da Matematiksel Yeterlikler

PISA, matematik testinde içerik belirlenirken okul müfredatlarında yer alan ana temalardan faydalanılmıştır. PISA 2006 matematik maddelerinin yaklaşık %50'si işlemler, ölçme, temel cebir ve geometrik kavramlar ile ilgiliyken yaklaşık %20'si veri sunumuna odaklanmaktadır. Geri kalan maddeler ise sayılar, istatistik ve olasılık konularına dağılmakta, %5'lik az bir kısım ileri geometri ve kombinasyon gibi özel konu alanlarında yer almaktadır.

PISA, maddeleri geliştirdikten sonra öğrencilerin yanıtlarının analizleri sonucu, öğrencilerin sorulara doğru yanıt verebilmeleri için bazı yeterliklere sahip olması gerektiğine vurgu yapmışlardır. Bu yeterliliklere sahip olabilmenin de aslında matematik okuryazarlığı için ön koşul niteliğinde olduğu belirtilmiştir. Belirlenen yeterlikler her maddede dönüşümlü olarak sınanmıştır.

Genel olarak 7 temel başlıkta ifade edilen bu yeterlikler; İlişkilendirme, İletişim, Matematikleştirme, Çoklu Temsiller, Akıl Yürütme ve Kanıt Gösterme, Strateji Geliştirme, Sembolik, Formel ve Teknik Dil ve İşlem Kullanılmasıdır. Bu yeterlikler kısaca aşağıdaki gibi tanımlanmıştır. (Oecd, 2013)

İlişkilendirme: Matematik, sadece kurallar, semboller, şekiller ve işlemlerden ibaret değildir. İçinde bir anlam bütünlüğü olan düzenler ve ilişkiler ağıdır. Matematikle diğer disiplinler ve yaşam arasında da ilişkiler bulunmaktadır. Buna bağlı olarak ilişkilendirme becerisi, matematik kavramlarının kendi aralarında da, bir matematiksel kavramın diğer disiplinlerle ve günlük hayatla ilişkilendirilmesini kapsamaktadır.

İlişkilendirme becerilerinin geliştirilmesi için dikkate alınması gereken bazı göstergeler şunlardır: Kavramlar ve işlemler arasında ilişki kurma, Matematiksel kavram ve kuralları farklı

temsil biçimleriyle gösterme, ilişkilendirme ve birbirine dönüştürme, farklı Matematik kavramlarını birbiriyle ilişkilendirme, Matematiği diğer derslerde ve günlük yaşamda karşılaşılan konu ve durumlarla ilişkilendirme.

İletişim: Temel Matematik yeteneği olarak iletişim, problem çözme sürecinin aşamaları içerisindeki ya da arasındaki iletişimin sağlanabilmesi yeteneğini ifade etmektedir. Bir problemin çözümü; okumayı, yeniden kodlamayı, ifade, soru, görev ve verilenleri yorumlama gibi problemi anlama ve tanımaya yönelik birçok bireysel eylemin gerçekleştirilmesini, problemin netleştirilmesi ve formülleştirilmesi aşamalarını, çözüm için gerekli işlemlerin gerçekleştirilmesini ve çözümler üretilmesini içermektedir. Bir çözüme ulaşıldığında, bu çözümün sunulması, açıklanması ve doğruluğunun kanıtlanması, problemi çözen kişi ya da kişilerin eylemleri arasındadır. İletişim yeterliği ile ilgili dikkate alınması gereken bazı göstergeler şunlardır: Matematiğin sembol ve terimlerini etkili ve doğru kullanma, Matematiksel dili Matematiğin kendi içinde, farklı disiplinlerde ve yaşantısında uygun ve etkili bir biçimde kullanma, somut model, şekil, resim, grafik, tablo, sembol vb. farklı temsil biçimlerini kullanarak Matematiksel düşünceleri ifade etme, Matematiksel düşünceleri sözlü ve yazılı ifade etme, günlük dili, matematiksel dil ve sembollerle Matematiksel dili de günlük dil ve sembollerle ilişkilendirme, Matematiksel düşüncelerin doğruluğunu ve anlamını yorumlamadır.

Matematikleştirme: Gerçek dünyada karşılaşılabilecek bir problemi kesin bir Matematiksel forma dönüştürebilme becerilerinin bir bütününü ya da yeteneğini içermektedir. Gerçek yaşamda karşılaşılan problemler, doğası gereği açık ve kesin Matematiksel biçimde değildir. Bu tür problemlerin çözümünde, öncelikle Matematiksel biçime dönüştürerek tanımlama ve açıklama gerekebilmektedir.

Matematikleştirme, bir problemi kesin Matematiksel formda ifade etmeye yönelik modelleme, yapısal gösterim, varsayımlarla tanımlama, formülleştirme, kurulan bir yapı ya da modele yönelik matematiksel çıktıların elde edilmesi ve yorumlanması gibi spesifik eylemleri ifade etmektedir.

Çoklu temsiller: Çoklu temsil ya da gösterim sıklıkla matematiksel obje ya da durumların betimlenmesinde kullanılmaktadır. Matematik eğitiminde çoklu temsillerin önemi hem öğrenmeye araç hem de öğrenmenin bir göstergesi olmaları gibi iki önemli noktada kendini göstermektedir. Öğrencilerin Matematiksel kavramları anlamaları, bunları kullanarak etkili bir şekilde iletişim kurabilmeleri bunların değişik temsillerini kullanabilmeleri ve bu temsiller arasında geçiş yapabilmeleriyle ilgilidir. Uygulamada çoklu temsiller, grafik, tablo, diyagram, resim, denklem, formül ve diğer görsel araçları ifade etmektedir.

Akıl yürütme ve kanıt gösterme: Akıl yürütme (muhakeme), eldeki bilgilerden hareketle matematiğin kendine özgü araç (semboller, tanımlar, ilişkiler, vb.) ve düşünme tekniklerini (tümevarım, tümdengelim, karşılaştırma, genelleme, vb.) kullanarak yeni bilgiler elde etme süreci olarak tanımlanabilir. Akıl yürütme ve kanıt gösterme, matematik okuryazarlığının farklı aşamalarını ve bu aşamalardaki ilişkili eylemleri ifade eden bir beceriler bütünü ya da yeterliktir. Çıkarımların doğruluğunu ve geçerliliğini savunma, mantıklı genellemelerde ve çıkarımlarda bulunma, bir matematiksel durumu analiz ederken matematiksel örüntü ve ilişkileri açıklama ve kullanma bu becerinin temel göstergelerini oluşturmaktadır.

Strateji geliştirme: Strateji belirleme ya da tasarlama, etkili bir tanıma, formülleştirme ve problem çözmede bireylere kılavuzluk eden bir dizi kritik kontrol süreci içermektedir. Bu yeterlik, problem çözmede Matematiksel bilgi ve becerilerin kullanılmasına yönelik bir plan ya da stratejinin seçilmesi ya da tasarlanmasını ifade etmektedir. Aynı zamanda problem çözme sürecinin herhangi bir aşamasında kullanılabilecek bir yeterlik olarak görülmektedir.

Sembolik, formel ve teknik dil ve işlemlerin kullanılması: Matematik okuryazarlığı, sembolik, formel ve teknik dil ve işlemlerin kullanılmasını gerektirmektedir. Matematik okuryazarlığı açısından sembolik, formel ve teknik dilin kullanılması yeteneği, Matematiksel kurallarla ifade edilen bir Matematiksel içeriğe yönelik sembol gösterimlerini anlama, yorumlama, manipüle etme ve kullanma davranışlarının bir bütünü olarak ifade edilebilir. Aynı zamanda tanımlama ve kurallara dayalı formel yapıların ve formel sistemlerin anlaşılmasını ve kullanılmasını, ayrıca bu anlama ve kullanma süreçlerinde algoritmaların da kullanılmasını içermektedir.

3.1.2.2 Araştırmada Matematiksel Yeterliklerin belirlenmesi

P115K531 No'lu projede Matematiksel yeterlikler belirlenirken PISA standartlarından yola çıkılmıştır. Bununla birlikte proje araştırmacıları, danışmalar ve alan uzmanı öğretmenler ile birlikte 6. sınıf düzeyi göz önüne alınarak tekrar düzenlenmiştir. İlişkilendirme, İletişim, Matematikleştirme, Çoklu Temsiller, Akıl Yürütme ve Kanıt Gösterme Strateji Geliştirme, Sembolik, Formel ve Teknik Dil ve İşlem Kullanılması yeterlikleri 4 özellik altında toplanmıştır. Bu yeterlikler ve proje analiz kodları aşağıda Tablo 2'de verilmiştir.

Tablo 2. Proje Matematiksel Yeterlik Kodları ve Tanımları

KOD	YETERLİK
M1	İletişim ve İlişkilendirme
M2	Matematikleştirme
M3	Akıl Yürütme ve Strateji Geliştirme
M4	Sembolik ve Teknik Dil Kullanımı

İletişim ve İlişkilendirme (M1): İletişim becerisi; matematiksel dili, günlük dil ve sembollerle ilişkilendirme, matematiksel düşüncelerin doğruluğunu ve anlamını yorumlama gibi özellikleri kapsar. İlişkilendirme becerisi; matematik kavramlarının kendi aralarında, diğer disiplinlerle ve günlük hayatla ilişkilendirilmesini kapsamaktadır.

Matematikleştirme (M2): Bir problemi kesin matematiksel formda ifade etmeye yönelik modelleme, yapısal gösterim, varsayımlarla tanımlama, formülleştirme, kurulan bir yapı ya da modele yönelik matematiksel çıktıların elde edilmesi ve yorumlanması gibi spesifik eylemleri ifade etmektedir.

Akıl Yürütme ve Strateji Geliştirme (M3): Akıl yürütme, eldeki bilgilerden hareketle matematiğin kendine özgü araç (semboller, tanımlar, ilişkiler, vb.) ve düşünme tekniklerini (tümevarım, tümdengelim, karşılaştırma, genelleme, vb.) kullanarak yeni bilgiler elde etme sürecidir. Strateji geliştirme, problem çözmede matematiksel bilgi ve becerilerin kullanılmasına yönelik bir plan ya da stratejinin seçilmesi ya da tasarlanmasını ifade etmektedir.

Sembolik ve Teknik Dil Kullanımı (M4): Matematik okuryazarlığı açısından sembolik ve teknik dilin kullanılması yeteneği, Matematiksel kurallarla ifade edilen bir matematiksel içeriğe yönelik olarak sembol gösterimlerini anlama, yorumlama davranışlarının bir bütünüdür.

3.1.2.3 Matematiksel Hata Matrislerinin Belirlenmesi

Proje kapsamında öğrencileri izleme testleri sonucunda sadece kazanımlara ulaşma ve matematiksel yeterliklere sahip olma düzeyleri hakkında değil aynı zamanda hangi matematiksel hataları yaptıklarına yönelik bir geri bildirimde verilmiştir. Matematik alanında yaygın olarak gözlenen hatalar; problem durumunu anlayamama, matematiksel durumu anlama güçlüğü, kavramsal bilgi yetersizliği, işlemsel bilgi yetersizliği, dikkat eksikliği, matematiksel dile dönüştürememe, işlemsel bilgiyi kullanamamadır.

Pilot uygulama analizleri sonucunda proje kapsamında uygulanan izleme testlerinin geri bildirim sisteminde kullanılabilecek 3 temel matematiksel hata türüne göre q matrisler hazırlanmaya başlanmıştır. Bu hatalar ve proje kodları Tablo 3'de verilmiştir.

Tablo 3. Proje Hata Kodları ve Tanımları

KOD	HATALAR
E1	Matematiksel Durumu Anlama Güçlüğü
E2	Dönüştürme Güçlüğü
E3	İşlem Güçlüğü

Matematiksel Durumu Anlama Güçlüğü (E1): Öğrencinin bu hatayı gösterme oranı kendisine yöneltilen soruları yanıtlarken kavramsal bilgi yetersizliği, kavram yanılgısı, detayları görememe ve bilinenlerle bilinmeyenleri ayırt edememe düzeylerini göstermektedir.

Dönüştürme Güçlüğü (E2): Öğrencinin bu hatayı gösterme oranı kendisine yöneltilen soruları yanıtlarken Matematiksel dile dönüştürememe, temel ilişkileri görememe, İşlemsel bilgi yetersizliği düzeylerini göstermektedir.

İşlem Güçlüğü (E3): Öğrencinin bu hatayı gösterme oranı kendisine yöneltilen soruları yanıtlarken uygun çözüm yolunu seçememe ve dikkate dayalı işlemlerde yetersizlik düzeylerini göstermektedir.

3.1.3 Soru yazama ekibinin kurulması

Soru yazma ekibinin kurulması için 6. sınıf Matematik dersine giren öğretmenlere yönelik iki eğitim düzenlenmiştir. "Ortaokul Matematik Öğretmenleri İçin Üst Düzey Düşünme Becerilerine Yönelik Soru Yazma Eğitimi" başlıklı 9-10 Şubat 2016 tarihleri arasında gerçekleşen eğitime başvuran 692 öğretmen arasından seçilen 113 öğretmene 6. Sınıf düzeyinde PISA ve TIMSS testlerinde kullanılan özellikle üst düzey düşünme becerilerine yönelik sorular yazabilme becerisi kazandırmak amacıyla 2 gün boyunca 3 ayrı grup ve 5 ayrı oturumda çalıştaylar düzenlenmiş ve seminerler verilmiştir. Eğitim sonunda başarılı olan ve projede soru yazarı olarak çalışmak isteyen toplam 77 öğretmen soru yazarı olmayı kabul etmiş ve proje soru yazma sürecine ilişkin birinci etap başlamıştır.

3.1.4 Madde havuzu oluşturma süreci

Projenin bu aşamasında eğitim alan öğretmenler tarafından oluşturulan soru yazar ekibi ile projenin madde havuzunu oluşturma süreci gerçekleştirilmiştir. Bu süreçte 77 öğretmen proje amaçları doğrultusunda 2 aylık süreçte aşamalı olarak 324 soruyu proje web sitesi üzerinden sisteme yüklemişlerdir. Soru yazarlarının soruyu nasıl yazacakları ve sistem üzerinde soruyu nasıl tanımlayacakları yönergeler ile kendilerine bildirilmiştir. Bu süreçte yazarların sadece soruyu yazmaları değil aynı zamanda onlardan proje amaçlarına göre sorunun hangi özellikleri ölçtüğü, hangi kazanımlar ile tanımlanabileceği ve bilişsel süreç becerileri ya da üst düzey düşünme becerileri ile nasıl ilişkilendirilebileceğine dair bilgileri de tanımlamaları istenmiştir. Bu tanımlamalar proje kapsamında geliştirilen soruların daha sonra oluşturulacak Q matrisler ile ilişkilendirilmesinin ön çalışması olmuştur. Soru yazarları sorularını sisteme yükledikten sonra süreçten kopmamış ve soru değerlendirme sürecine de aktif olarak katılarak yazdıkları soruların daha üst düzey nitelikleri ölçebilmesi amacıyla geliştirilmesine katkı sağlamışlardır.

3.1.5 Soru değerlendirme süreci

Bu süreçte soru yazım ve değerlendirme aşamasına ilişkin çalışmalar başlamıştır. Projeye ait bir web sitesi kurulmuştur (www.cdmproject.org). Web sitesi projenin tanıtımının yanı sıra geliştirilen soruların değerlendirme sürecinin organizasyonuna da hizmet etmiştir. Projede soru yazarları dışında soruları değerlendirmek amacıyla bir hakem grubu belirlenmiştir. En az yüksek lisans yapmış matematik öğretmenleri ve deneyimli akademisyenlerden seçilen 28 kişiden oluşan hakem grubu yazarlar tarafından geliştirilen soruları değerlendirmişlerdir. Değerlendirme yapabilmek amacıyla geliştirilecek testlerde yer alan soruların proje amaçlarına uygunluğu ve beklenen standartları karşılayıp karşılamadığının belirlenebilmesi için soru yazarları ile soruları değerlendirecek hakemler arasında bir editörlük süreci işletilmiştir. Bu nedenle projede, soru yazarları ile hakemler arasında görüş alış verişinin etkin bir şekilde yürütülmesini sağlayan bir çevrimiçi editörlük sistemi geliştirilmiştir. Bu sistem, editörlük sürecinde yaşanan zorluklar göz önüne alarak değerlendirme sürecini iyileştirmek, hızlandırmak ve daha sonraki aşamalarda kullanılabilecek veriler toplamak amaçlarına hizmet etmiştir. Geliştirilen editörlük sisteminin diğer modellerden ayıran en önemli özellikleri sürecin çevrimiçi olması, hakem yazar arasında anlık iletişime izin vermesi ve bütün bu süreç boyunca hakem yazar anonimliğinin korumasıdır. Bu süreçte her soru üç hakem tarafından değerlendirilmiştir. Geliştirilen sistem sayesinde hakemler birbirinden bağımsız olarak soru yazarlarıyla değerlendirme sürecini yürütmüştür. Hakem – yazar arasındaki anlık bilgi ve doküman alışverişi süreci geleneksel soru değerlendirme yaklaşımından soruyu geliştirme anlayışına çevirmiştir. Yazar ve hakem arasındaki bilgi paylaşım süreci sadece iki kişi arasında olmakla birlikte kişisel bilgiler yönünden iki taraf için de tamamen anonim olmuştur. Her hakem soruyu ön ve son değerlendirme olarak iki kere puanlamıştır. Ön değerlendirme hakemin soru yazarı ile görüşmesinden önce gerçekleşmekteyken son değerlendirme yazar ile hakem arasındaki görüşmeler sona erdiğinde hakem tarafından yapılmıştır. Hakemler soru değerlendirmelerini, sistem üzerindeki değerlendirme formları aracılığı ile proje kuruluna göndermişlerdir. Bu sayede sorular hem amaca uygun olarak hızlı ve kör hakemlik kurallarıyla değerlendirilmiş hem de sorulara ilişkin değerlendirme verileri dijital olarak toplanabilmiştir. Sistemin kullanımına ilişkin açıklamalar ve kılavuzlar hakem ve yazarlar için ayrı hazırlanmıştır.

Projede kullanılacak soruları değerlendirebilmek amacıyla 15 maddeden oluşan 5'li puanlama anahtarı geliştirilmiştir(EK 1). Bu yöntemle 75 matematik öğretmeni tarafından 6. sınıf düzeyinde yazılan toplam 324 proje sorusu 28 hakem tarafından proje web sitesinde çevrimiçi değerlendirilmiştir.

3.2 Deneme Uygulaması ve Ön-Son Testin Hazırlanması

Araştırmada oluşturulan madde havuzunda yer alan sorular için ön-son test psikometrik özelliklerin belirlenmesi ve PISA-TIMSS testlerinin eşdeğerliğinin incelenmesi amacı ile deneme uygulaması yapılmıştır. Deneme uygulamalarının yapılacağı İzmir ilinde 7 ilçede yer alan toplam 12 okul tabakalama yöntemiyle belirlenmiştir. Deneme uygulaması soruların psikometrik özelliklerini belirleme amacı taşıdığından testler seçilen okullardaki hem 6. sınıflara (n=3303) hem de 7. sınıflara (n=2926) ait öğrencilerden toplam 6229 öğrenci arasından 3200 kişilik grup pilot uygulama örneklemi olarak belirlenmiştir. Deneme uygulamasında belirlen sorular projenin deneysel basamağında kullanılacak ön-test, son-test ve izleme testlerinin içeriğini oluşturmuştur. Bu nedenle soruların geniş bir grupta denenebilmesi ve aynı zamanda PISA ve TIMSS sınavlarıyla eşdeğerliğinin ölçülebilmesi amacıyla deneme uygulaması için zincir modelli eşitleme deseni kullanılmıştır (Holland, Dorans, & Petersen, 2006; Kolen & Brennan, 2004). Pilot uygulama deseni ankor (bağ) ve farklı sorulardan oluşan 6 kitapçık olarak tasarlanmıştır. Her kitapçıkta açık uçlu ve çoktan seçmeli formatında toplam 15 soru yer almıştır. Pilot uygulama deseninin karmaşık yapısı iki temel amaç için tasarlanmıştır. Öncelikle geliştirilen soruların PISA ve TIMSS soruları ile aynı düzey ve psikometrik özelliklere sahip olduğunun kanıtlanması ve ikinci olarak da deneme uygulamasındaki gruplarının denkliklerini sağlayarak proje sorularının psikometrik özelliklerinin farklı örneklemlerden gelmesine rağmen araştırılabilmesidir. Proje kapsamında geliştirilen soruların PISA ve TIMSS sorularıyla eşdeğerinin incelenebilmesi için tasarlanan 6 kitapçık içinde de daha önce çıkmış PISA ve TIMSS soruları kullanılmıştır. Uygulamada okul, sınıf ve şube farklarının maddelerin psikometrik özelliklerini etkilememesi için her okul ve sınıfta bütün kitapçıklar eşit dağıtılmıştır. Uygulama deseni Tablo 4'de verilmiştir.

Tablo 4. Pilot Uygulama 00 Deseni

NO	KAYNAK	KİTAPÇIK A	KİTAPÇIK B	KİTAPÇIK C	KİTAPÇIK D	KİTAPÇIK E	KİTAPÇIK F
1	PISA	p1	p1	p1	p1	p1	p1
2	TIMSS	t1	t1	t1	t1	t1	t1
3	PROJE	s2	s2	s2	s2	s2	s2
4	PROJE	s9	s9	s9	s9	s9	s9
5	PISA	p5	p2	p3	p3	p4	p7
6	PISA	p6	p5	p2	p4	p7	p6
7	TIMSS	t3	t2	t2	t5	t7	t7
8	TIMSS	t4	t4	t5	t6	t6	t3
9	PROJE	s22	s22	s21	s19	s18	s14
10	PROJE	s23	s23	s4	s20	s16	s15
11	PROJE	s12	s21	s19	s18	s14	s27
12	PROJE	s13	s4	s20	s16	s15	s29
13	PROJE	s32	s7	s24	s3	s26	s25
14	PROJE	s1	s28	s33	s31	s6	s34
15	PROJE	s5	s30	s8	s10	s11	s17

Tablo 4'de görüldüğü gibi desende PISA'ya ait 7, TIMSS'e ait 7 ve projede üretilen 34 soru olmak üzere toplam 48 soru uygulanmıştır. Her kitapçıkta 3 PISA, 3 TIMSS ve 9 proje sorusu bulunmaktadır.

3.2.1 Deneme uygulamasının puanlanması

Deneme uygulamasında hem çoktan seçmeli hem de açık uçlu sorular birlikte kullanılmıştır. Projenin temel amaçlarından biri açık uçlu soruların genel sınavlarda kullanılabilmesi ve puanlanmasına ilişkin dönütler alabilmektir. Deneme uygulamasında kullanılan 48 sorunun 25 tanesi çoktan seçmeli, 23 tanesi açık uçlu sorulardan oluşmuştur. Açık uçlu olarak tanımlanan sorular PISA ve TIMSS içeriğine benzer şekilde kısa cevaplı, uygulamalı ve doğru yanlış türü maddelerden oluşturulmuştur. Bu maddeler 5 kitapçığa eşit olarak (6) paylaştırılmış, sadece bir kitapçıkta 7 açık uçlu madde yer almıştır. Açık uçlu ve çoktan seçmeli maddelerin bir arada okutulabilmesi için özel bir optik form geliştirilmiştir. Öğrenciler sınav esnasında çoktan seçmeli maddeleri optik form üzerine işaretlemişler, açık uçlu soruları ise kitapçıkta çözmüşlerdir. Öğrencilerin açık uçlu sorulara verdiği cevapların tamamının dijital olarak kaydedilmesi ve optik okuyucu tarafından bir seferde okunması sağlanmıştır. Projede kullanılan optik form EK 2a ve 2b'de verilmiştir. Bu şekilde elde edilen veri farklı analizler ile en doğru puanlama anahtarını oluşturmaya uygun hale getirilmiş, aynı zamanda da öğrencilerin öğrenme eksikleri genel hata ve yanılgılarını analiz edebilmeyi mümkün kılmıştır. Bu süreçte açık uçlu soruyu değerlendiren hakemin katılık düzeyi puanlama sürecinden arındırılmaya çalışılmıştır. Muraki'nin puanlayıcı etkisi modeli ve Yao'nun puanlayıcı modelini kullanarak (Patz, Junker, Johnson, & Mariano, 2002), hakem katılığının puanlama üzerindeki etkisini belirlenmeye çalışılmıştır(Basokcu & Gökçe, 2016).

3.2.2 Proje soruları ile PISA, TIMSS testlerinin eşdeğerliği çalışması

Geliştirilen soruların PISA ve TIMSS soruları ile aynı düzey ve psikometrik özelliklere sahip olduğunun kanıtlanması projenin sürdürülebilirliği açısından oldukça önemlidir. Bu nedenle soruların testin geneline ilişkin istatistiklerinin hesaplanabilmesi için ortalamaları alınmıştır. Projede geliştirilen soruların PISA ve TIMSS soruları ile eşdeğer olup olmadığını belirlemek için tekrarlı ölçümler için ANOVA analizi yapılmıştır. Analiz sonuçlarına göre PISA, TIMSS ve proje sorularının ortalamaları arasında istatistiksel olarak anlamlı bir fark bulunmamıştır($F_{(896-2)}$ = 2,358, p>0,5). Dağılımların varyanslarının eşit olup olmadıkları Mauchly Küresellik testi ile incelenmiştir. Analiz sonuçlarına göre küresellik varsayımının bozulmadığı bir başka deyişle dağılımların varyanslarının eşit olduğu görülmüştür, ($\chi^2(2)=4.881, p>0,5$). Bu durum proje kapsamında üretilen soruların PISA ve TIMSS sorularıyla eşdeğer olduğunun bir göstergesidir. Bununla birlikte proje kapsamında üretilen soruların PISA ve TIMSS sorularıyla eşdeğerliği kapsamı ve ölçüt geçerliği de incelenmiştir. Ölçüt geçerliğine göre test puanları birbirleriyle

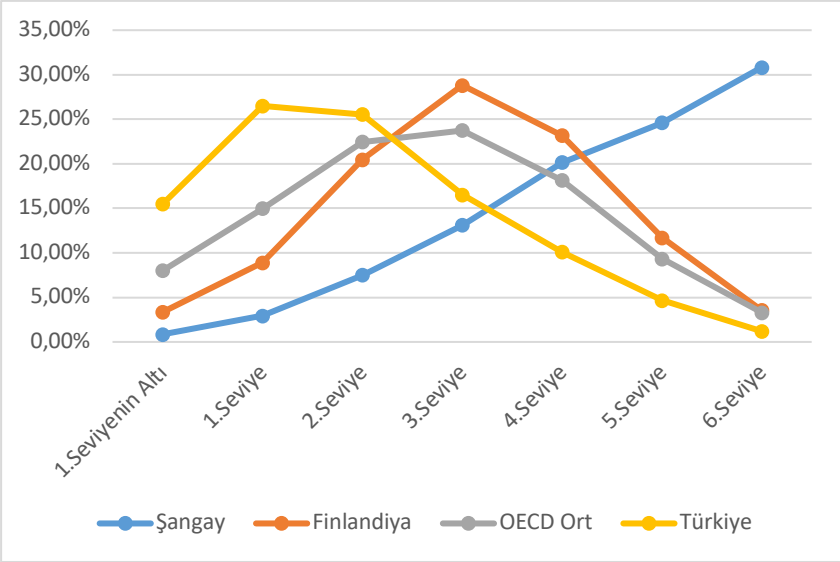
istatistiksel olarak anlamlı ve yüksek bir ilişki vermiştir (r=0.52, p<0,01). Kapsam geçerliği çalışması en az doçentlik seviyesinde olan, iki ölçme ve değerlendirme, iki matematik eğitimi ve bir program geliştirme alanında uzman proje araştırmacı ve danışmanlarından oluşan 5 kişilik ekip tarafından gerçekleştirilmiştir. Yapılan çalışma sonucunda projede üretilen soruların PISA ve TIMMS kapsamıyla uyumlu olduğu görülmüştür. Aynı zamanda testlerin eşdeğerliğine yönelik BTM içerikli eşitleme çalışması yapılmıştır. Lokal eşitleme ve BTM'ye göre eşitleme çalışmalarını içeren bu araştırma gerçek veri uygulaması ve simülasyon çalışması içermektedir(Basokcu & Canpolat, 2016; van der Linden & Wiberg, 2010; Xin & Zhang, 2015).

Araştırma PISA uygulamalarında öğrenciler için belirlenen matematiksel yeterlik düzeylerinin Türkiye örneklemindeki dağılımlarını incelemek ve benzer bir ölçümde PISA yaş aralığından daha erken bir dönemi hedef alarak belirlenen dağılımları karşılaştırmayı içermektedir. Bu amaçla öncelikle yeterlik düzeylerinin ülkelere ve genel ortalamada nasıl dağıldığı incelenmiştir. PISA 2012 sonuçları referans alınarak elde edilen bazı sonuçlar ve ülke bazlı kıyaslamalar aşağıda verilmiştir.

Tablo 5. PISA 2012 Matematiksel Yeterliklerin Dağılımı

Düzeyler	Şanghay	Finlandiya	OECD Ort	Türkiye
Level 1^-	0,85%	3,34%	8,02%	15,48%
Level 1	2,95%	8,92%	14,98%	26,50%
Level 2	7,51%	20,49%	22,46%	25,54%
Level 3	13,10%	28,82%	23,74%	16,52%
Level 4	20,17%	23,17%	18,15%	10,09%
Level 5	24,60%	11,71%	9,34%	4,67%
Level 6	30,83%	3,54%	3,31%	1,20%

Tablodaki yeterlik seviyeleri referans alınarak elde edilen yüzdelik dağılımlara bakarak 1.seviyenin altında Şanghay'da yüzde 0,85'lik bir kesim bulunurken, OECD ortalamasında ise bu yüzde 8,02'lik bir kısma karşılık gelmektedir. Türkiye'de ise bu oran %15,48 ile yüksek bir düzeydedir. 6.seviyeye baktığımızda Şanghay'da yüzde 30,83'lük bir kesim varken Türkiye'de ise bu seviyeye öğrencilerin sadece yüzde 1,20'lik kısım çıkabilmiştir. Aşağıdaki grafikte ülkelerin dağılımları daha açık görülmektedir.



Şekil 5. PISA 2012 Matematik Yeterlik Düzeylerinin Dağılım Grafiği

Grafik genel olarak incelendiğinde Şanghay'ın sola çarpık, Finlandiya ve OECD ortalamasının normal, Türkiye'nin ise sağa çarpık bir dağılım gösterdiği gözlenmektedir. Bu durum başarı ölçüsü olarak ülkelerin düzeylerini anlama konusunda yeterli bir fikir vermektedir. Aşağıdaki tablo, karşılaştırılan ülkelerin ortalama ve standart sapmalarını vermektedir.

Tablo 6.PISA 2012 Matematik Puanları Ortalama ve Standart Sapmaları

	Şanghay	Finlandiya	OECD Ortalama	Türkiye
Ortalama Puan	613	519	494	448
Standart Hata	3,3	1,9	0,5	4,8

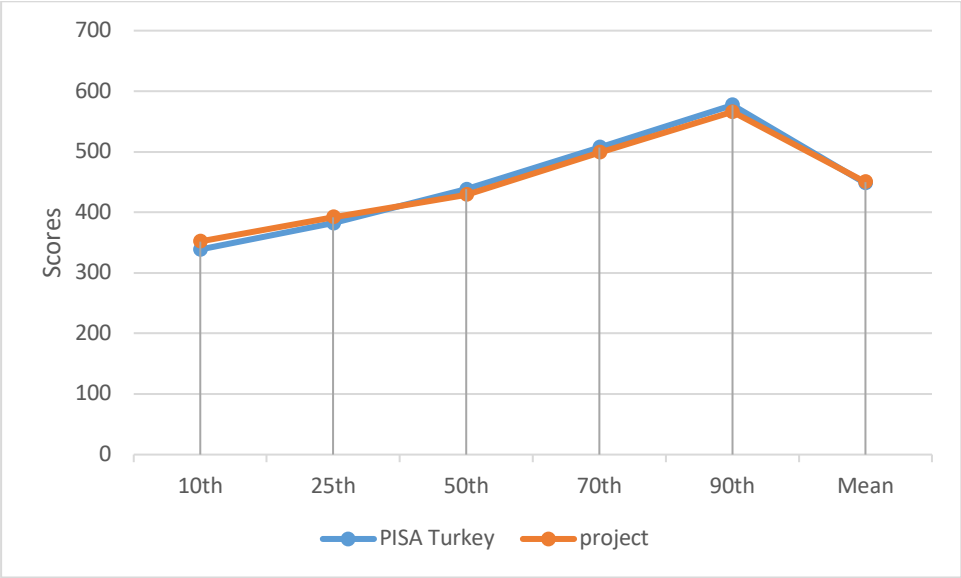
Tablo 6'daki verilere göre Şanghay'ın ortalama puanının OECD ortalamasına göre oldukça yüksek bir değerde olduğu görülmektedir. Türkiye'nin ortalama puanı incelendiğinde ise ölçüt alınan diğer ülkelere göre oldukça düşük olduğu görülmektedir.

Proje kapsamında PISA içeriği ve düzeyi ile eşdeğer olduğu kanıtlanan test puanları PISA prosedürüne uygun olarak öncelikle Rasch model ile analiz edilerek örneklemin yetenek puanları belirlenmiş ve daha sonrada olası puanlar hesaplanarak istatistiksel olarak daha güvenilir bir yetenek dağılımı elde edilmiştir. Aşağıda Tablo 7'de PISA 2015 Türkiye puanları ve Araştırmadan elde edilen puanların yüzdelik kesme (treshhold) puanları verilmiştir.

Tablo 7.PISA ve Proje ölçümleri için yüzdelik düzeylerin kesme puanlarının karşılaştırılması

	PISA Türkiye 15 yaş	Proje 12 yaş
10.	339	352
25.	382	392
50.	438	429
70.	507	499
90.	577	566
Ort.	448	450

Tablo 7 incelendiğinde PISA 2015 15 yaş Türkiye örneklemi ile 12 yaş Türkiye örnekleminin kesme puanlarının çok benzer olduğu görülmektedir. Aşağıdaki grafik iki test için kesme puanların örtüşme düzeyini göstermektedir.



Şekil 6. PISA 2015 ve Proje Kesme Puanların Çizgi Grafiği

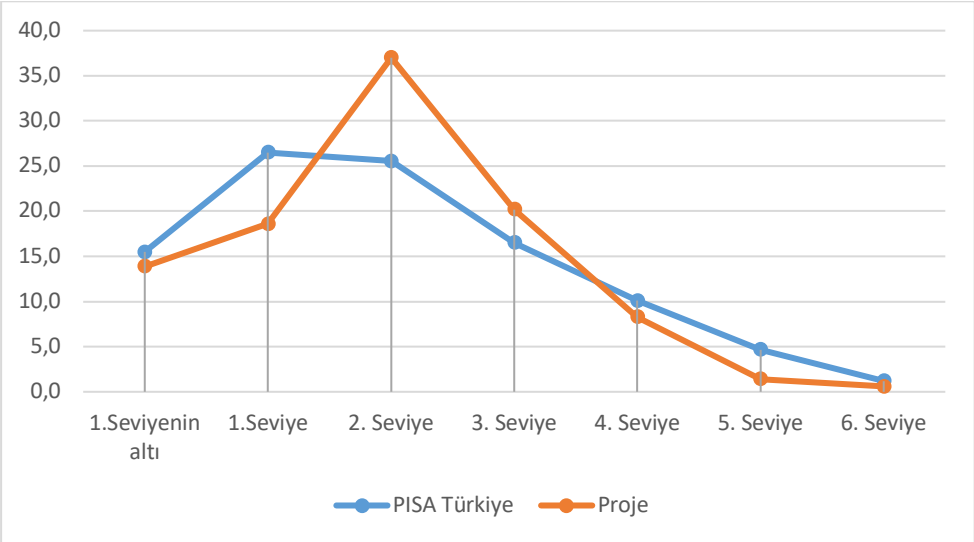
Şekil 6'da görüldüğü gibi iki grubun kesme puanları çok yakın ilerlemektedir. Bu durum iki örneklemin yüzdelik dağılımlarının birbirine çok benzediğini göstermektedir. Sonraki aşamada iki grubun matematiksel yeterlikler bakımından dağılımları yüzdelikler düzeyinde incelenmiştir. Aşağıdaki Tablo 8, 15 ve 12 yaş Türkiye dağılımının matematiksel yeterlik düzeylerini göstermektedir.

Tablo 8. PISA ve Proje ölçümleri için matematiksel yeterlik seviyeleri yüzdelik dağılımları

	PISA Türkiye 15 yaş	Proje 12 yaş
Level 1⁻	15,5	13,9
Level 1	26,5	18,6
Level 2	25,5	37,0
Level 3	16,5	20,2
Level 4	10,1	8,3
Level 5	4,7	1,4
Level 6	1,2	0,6

Tablo 8 incelendiğinde iki grupta da matematiksel yeterlik seviyelerindeki yüzdelik dağılımlarının benzer olduğu görülmektedir. 15 yaş grubunda 1. Seviyenin altı %15,5 iken 12 yaş da bu oran 13,9 dur. Benzer şekilde 15 grubun 1. Seviyeye düşme oranı daha yüksektir.

İkinci ve üçüncü seviyelerde ise 12 yaş grubunun daha yüksek yüzdede olduğu görülmektedir. 4, 5 ve 6. Seviyelerde ise 15 yaş grubunun oransal olarak daha yüksek olduğu görülmektedir. Aşağıdaki grafik, 12 ve 15 yaş grubunun matematiksel yeterlik seviyelerine göre dağılımını daha açık göstermektedir.



Şekil 7. PISA ve Proje ölçümleri için matematiksel yeterlik seviyeleri çizgi grafiği

Şekil 7 incelendiğinde 15 yaş grubunun 1. Seviye altı ve 1. Seviyede daha fazla olduğu görülmektedir. Bu durum 15 yaş örnekleminde 1 seviye ve altında, yani düşük düzeyde olan kişi sayısının 12 yaş grubundan daha fazla olduğunu göstermektedir. İkinci ve üçüncü seviye incelendiğinde ise 12 yaş grubunun yüzdelik dağılımının bu aralıkta daha yüksek olduğu görülmektedir. 12 yaş grubu için dağılımın orta düzeyde daha yoğun olduğu söylenebilir. 4, 5 ve 6. Seviyelerde ise 15 yaş grubunun 12 yaş grubuna göre daha yüksek oranda olduğu görülmektedir. Genel olarak grafik incelendiğinde iki grup için bazı farklılıklar görülse de bu farkların istatistiksel olarak anlamlı olmadığı belirlenmiştir. Bu durumda 12 ve 15 yaş grubu öğrencilerinin PISA matematiksel yeterlikleri açısından farklılık göstermedikleri anlaşılmaktadır.

3.2.3 Deneme uygulaması yapılan grupların denkliklerinin araştırılması

Deneme uygulamasında yer alan grupların denkliklerini sağlayarak proje sorularının psikometrik özelliklerinin farklı örneklemlerden gelmesine rağmen araştırılabilmesi amacıyla testlerde ankor (bağ) maddeler kullanılmıştır. Tablo 4'de verilen desende görüldüğü gibi 4 madde ankor olarak bütün kitapçıklarda yer almaktadır. Bu maddeler için kitapçıklara göre ortalamalar arasında fark olup olmadığı tek yönlü varyans analizi (ANOVA) ile incelenmiştir. Analizler SPSS 23 programı ile yapılmıştır. Tablo 9'da ANOVA sonuçları verilmektedir.

Tablo 9. Deneme Uygulaması Gruplarının Denkliklerine İlişkin ANOVA Sonuçları

Betimsel İstatistikler					ANOVA Sonuçları					
		N	Ortalama	S. Sapma	Varyans Kaynağı	Kareler toplamı	Kareler Ortalaması	S.D.	F	P
s201 PISA	Kitapçık A	520	,871	,335	Gruplar arası	1,056	0,211	5	1,987	0,077
	Kitapçık B	513	,903	,297						
	Kitapçık C	501	,904	,295	Gruplar içi	313,616	0,106	2952		
	Kitapçık D	491	,855	,352						
	Kitapçık E	484	,862	,346	Toplam	314,672		2957		
	Kitapçık F	449	,878	,328						
	Toplam	2958	,879	,326						
s501 TIMSS	Kitapçık A	520	,508	,500	Gruplar arası	2,205	0,441	5	1,781	0,113
	Kitapçık B	513	,446	,498						
	Kitapçık C	501	,463	,499	Gruplar içi	731,042	0,248	2952		
	Kitapçık D	491	,442	,497						
	Kitapçık E	484	,426	,495	Toplam	733,247		2957		
	Kitapçık F	449	,434	,496						
	Toplam	2958	,454	,498						
s1119 PROJE	Kitapçık A	520	,487	,500	Gruplar arası	0,229	0,046	5	0,183	0,969
	Kitapçık B	513	,481	,500						
	Kitapçık C	501	,487	,500	Gruplar içi	739,076	0,250	2952		
	Kitapçık D	491	,505	,500						
	Kitapçık E	484	,490	,500	Toplam	739,305		2957		
	Kitapçık F	449	,503	,501						
	Toplam	2958	,492	,500						
s6251 PROJE	Kitapçık A	520	,315	,465	Gruplar arası	2,358	0,472	5	2,132	0,059
	Kitapçık B	513	,380	,486						
	Kitapçık C	501	,339	,474	Gruplar içi	652,963	0,221	2952		
	Kitapçık D	491	,287	,453						
	Kitapçık E	484	,329	,470	Toplam	655,321		2957		
	Kitapçık F	449	,336	,473						
	Toplam	2958	,331	,471						

Tablo 9'da görüldüğü gibi kitapçık türlerine göre soruların ortalamaları arasında istatistiksel olarak anlamlı bir fark yoktur. Bu durum grup denkliklerinin bir göstergesidir. Grup denkliklerine ilişkin bir diğer analiz maddelerin ayırt edicilik parametrelerinin gruplar arası değişimidir. Ankor maddelere ilişkin ayırt edicilik parametreleri Tablo 10'da verilmiştir.

Tablo 10. Deneme Uygulaması Ankor Maddelerine İlişkin Madde Parametreleri

Kitapçık	Soru Kodu	rjx	Alt Grup	Üst Grup	Noktaçift serili r	Çift serili r
Kitapçık A	201	0.238	0.713	0.951	0.305	0.486
	501	0.665	0.140	0.805	0.540	0.677
	1119	0.535	0.203	0.738	0.442	0.555
	6251	0.378	0.147	0.524	0.349	0.456
Kitapçık B	201	0.219	0.775	0.972	0.379	0.652
	501	0.623	0.145	0.768	0.554	0.697
	1119	0.448	0.243	0.691	0.413	0.518
	6251	0.373	0.197	0.569	0.313	0.400
Kitapçık C	201	0.224	0.783	0.986	0.367	0.634

	501	0.637	0.152	0.789	0.541	0.679
	1119	0.491	0.250	0.741	0.436	0.546
	6251	0.381	0.163	0.544	0.328	0.425
Kitapçık D	201	0.296	0.652	0.949	0.391	0.586
	501	0.612	0.110	0.722	0.545	0.686
	1119	0.441	0.280	0.722	0.413	0.517
	6251	0.241	0.134	0.375	0.309	0.399
Kitapçık E	201	0.270	0.686	0.956	0.412	0.643
	501	0.694	0.137	0.831	0.555	0.700
	1119	0.466	0.291	0.757	0.411	0.516
	6251	0.330	0.177	0.507	0.317	0.386
Kitapçık F	201	0.348	0.645	0.992	0.453	0.731
	501	0.683	0.091	0.773	0.536	0.675
	1119	0.552	0.198	0.750	0.419	0.525
	6251	0.391	0.132	0.523	0.360	0.466

Tablo 10'da görüldüğü gibi soruların hem ayırt edicilik değerleri hem de madde test korelasyonları yeterlidir ve gruplar arası manidar farklılık göstermemektedir. Özellikle maddelerin nokta çift serili ve çift serili korelasyon katsayıları bütün maddelerin yeterli ayırt edicilik değerine sahip olduğuna işaret etmektedir. Ayırt edicilik değerlerinin kitapçıklar arasında farklılık gösterip göstermediği Friedman testi ile incelenmiştir. Her bir bağ sorusunun kitapçıklar için değerleri tekrarlı ölçüm olarak kabul edilmiş ve non-parametrik Friedman testi sonucuna göre soruların kitapçıklara göre ayırt edicilik değerleri arasında anlamlı bir fark olmadığı görülmüştür (χ^2= 1,331, p>0,05). Elde edilen bulgular ışığında deneme uygulaması için belirlenen grupların denk oldukları ve bu gruplardan gelen veriler ile proje için geliştirilen soruların psikometrik özelliklerinin incelenebileceği anlaşılmaktadır.

3.2.4 Proje Sorularının Psikometrik Özelliklerinin İncelenmesi

Proje kapsamında geliştirilen soruların psikometrik özellikleri farklı yöntemler ile incelenmiştir. Öncelikle Klasik Test Kuramına (KTK) göre madde ayırt edicilik indeksleri ve madde test korelasyonları belirlenmiş daha sonra Madde Tepki Kuramı (MTK) parametrelerine göre sorulara ilişkin bulgulara ulaşılmıştır. Bununla birlikte son olarak projenin temel istatistik modelini oluşturan BTM analizleri ile soruların yeterlik düzeyleri incelenmiştir.

KTK analizlerinde maddelerin güçlük değerleri (P_j), ayırt edicilik parametreleri (r_{jx}) ve nokta çift serili (r_{pbis}) ve çift serili (r_{bis}) korelasyon katsayıları hesaplanmıştır (Haladyna & Rodriguez, 2013; McDonald, 2013; Murphy & Davidshofer, 2005). Bu analizler sonucunda testte yer alan maddeler nitelik düzeyine göre sınıflanmıştır (S). Analizler Iteman 4.0. programı ile yapılmıştır. Maddelere ait KTK bulgularını Tablo 11'de verilmiştir.

Tablo 11. Deneme Uygulaması Maddelerine İlişkin KTK Parametreleri

No	P_j	r_{jx}	r_{pbis}	r_{bis}	S	No	P_j	r_{jx}	r_{pbis}	r_{bis}	S
1	87,6%	0,249	0,382	0,618	4	**25**	45,5%	0,690	0,592	0,744	1
2	10,1%	0,099	0,182	0,311	1	**26**	48,1%	0,601	0,474	0,594	4
3	76,0%	0,413	0,459	0,631	1	**27**	19,4%	0,301	0,357	0,514	2
4	67,1%	0,458	0,431	0,560	4	**28**	79,0%	0,469	0,540	0,764	1
5	45,3%	0,428	0,368	0,462	3	**29**	61,4%	0,613	0,549	0,698	4
6	41,8%	0,542	0,474	0,598	3	**30**	45,9%	0,713	0,610	0,767	4
7	63,1%	0,569	0,554	0,709	1	**31**	23,4%	0,393	0,417	0,576	4
8	45,3%	0,636	0,544	0,684	4	**32**	43,0%	0,555	0,484	0,610	4
9	73,3%	0,411	0,420	0,564	1	**33**	35,7%	0,544	0,496	0,637	2
10	70,0%	0,401	0,406	0,535	4	**34**	50,5%	0,511	0,448	0,563	3
11	77,3%	0,414	0,413	0,574	3	**35**	67,3%	0,443	0,445	0,579	4
12	50,1%	0,535	0,447	0,560	4	**36**	9,6%	0,211	0,374	0,647	3
13	35,8%	0,695	0,625	0,803	4	**37**	0,5%	0,012	0,107	0,540	1
14	72,3%	0,529	0,525	0,702	1	**38**	30,5%	0,402	0,385	0,506	1
15	68,1%	0,633	0,601	0,784	4	**39**	30,1%	0,457	0,459	0,604	1
16	54,8%	0,420	0,391	0,491	2	**40**	54,8%	0,698	0,574	0,722	4
17	49,2%	0,470	0,420	0,527	4	**41**	10,2%	0,276	0,409	0,695	1
18	20,2%	0,057	0,049	0,070	1	**42**	5,8%	0,155	0,381	0,766	1
19	52,2%	0,517	0,456	0,572	2	**43**	66,6%	0,539	0,489	0,634	4
20	43,8%	0,545	0,444	0,559	4	**44**	53,8%	0,446	0,409	0,513	1
21	65,3%	0,479	0,445	0,574	2	**45**	56,0%	0,622	0,550	0,692	4
22	63,5%	0,571	0,539	0,690	3	**46**	20,2%	0,353	0,391	0,557	3
23	32,5%	0,345	0,320	0,418	3	**47**	53,5%	0,583	0,519	0,651	4
24	23,2%	0,386	0,460	0,636	3	**48**	45,0%	0,549	0,497	0,625	4

Tablo 11'de görüldüğü gibi maddelere ait parametreler ve istatistikler göz önüne alınarak bir kalite sınıflaması yapılmıştır. Sınıflama sonucunda 20 madde "çok iyi" (4), 9 madde "iyi" (3), 5 madde "orta" (2) ve 14 madde "yetersiz" (1) olarak kategorilendirilmiştir.

Daha sonraki aşamada maddelerin MTK analizleri ile madde parametreleri ve Madde Karakteristik Eğrileri (MKE) belirlenmiştir (Embretson & Reise, 2013; Hambleton & Swaminathan, 2013; F. M. Lord, 1980; Reckase, 2006). Analizler IRTPro 2.1 ve MPlus 5 programı kullanılarak yapılmıştır. Tablo 8'de deneme uygulamasında kullanılan soruların MTK'da 3 Parametreli Lojistik Modele göre (3PLM) belirlenmiş; A (ayırt edicilik-eğim), B (güçlük-madde konumu) ve C (şans başarısı- düşük asimptot) parametreleri ve onların standart hataları verilmiştir. Her sorunun son hücresinde maddelerin 4 kategorili kalite sınıfları

tanımlanmıştır. Bu sınıflar belirlenirken aynı zamanda MKE'de dikkate alınmıştır. Madde Karakteristik eğrileri raporda çok geniş yer tutacağından EK 3'de verilmiştir.

Tablo 12. Deneme Uygulaması Maddelerine İlişkin MTK Parametreleri

No	A	s.e.	C	s.e.	B	s.e.	S	No	A	s.e.	C	s.e.	B	s.e.	S
1	0.48	0.12	-0.18	0.09	0.37	0.21	1	**25**	0.47	0.13	-1.14	0.11	2.42	0.64	2
2	0.94	0.16	-0.09	0.10	0.10	0.11	3	**26**	1.18	0.18	-1.50	0.14	1.27	0.17	4
3	0.70	0.16	-1.56	0.13	2.24	0.47	1	**27**	1.71	0.23	-0.49	0.13	0.28	0.08	4
4	1.45	0.30	-2.99	0.31	2.06	0.28	3	**28**	1.31	0.18	-0.68	0.12	0.52	0.10	4
5	2.14	1.00	-7.77	1.97	3.62	0.91	4	**29**	0.76	0.13	0.07	0.10	-0.09	0.13	2
6	0.86	0.17	-1.57	0.14	1.82	0.31	3	**30**	1.47	0.20	0.35	0.12	-0.24	0.08	4
7	0.43	0.12	-0.22	0.09	0.50	0.25	1	**31**	1.22	0.21	2.29	0.20	-1.87	0.24	4
8	0.81	0.15	0.00	0.10	-0.01	0.12	3	**32**	0.83	0.14	0.98	0.12	-1.17	0.20	2
9	0.82	0.15	0.72	0.11	-0.89	0.18	3	**33**	1.47	0.20	1.46	0.17	-1.00	0.12	4
10	0.33	0.12	-0.50	0.09	1.54	0.60	1	**34**	2.03	0.27	1.25	0.19	-0.61	0.08	4
11	0.95	0.17	0.95	0.12	-1.00	0.18	2	**35**	0.73	0.12	-0.04	0.10	0.06	0.14	3
12	1.22	0.64	-5.87	0.85	4.80	2.06	3	**36**	1.68	0.21	-0.25	0.14	0.15	0.08	4
13	2.53	0.65	-4.88	0.83	1.92	0.21	2	**37**	1.80	0.26	1.92	0.22	-1.07	0.11	4
14	0.70	0.14	0.17	0.10	-0.24	0.14	2	**38**	1.93	0.24	0.02	0.15	-0.01	0.08	2
15	1.24	0.19	1.53	0.15	-1.23	0.16	3	**39**	1.45	0.18	0.28	0.13	-0.19	0.09	4
16	0.70	0.14	1.18	0.12	-1.68	0.32	3	**40**	1.53	0.27	2.69	0.27	-1.76	0.21	4
17	2.26	0.34	-0.96	0.18	0.42	0.07	4	**41**	1.45	0.20	0.75	0.14	-0.52	0.09	4
18	0.74	0.13	-0.06	0.10	0.08	0.13	3	**42**	0.52	0.13	-0.72	0.10	1.40	0.37	3
19	0.83	0.14	0.17	0.10	-0.21	0.12	3	**43**	1.77	0.28	2.10	0.23	-1.19	0.12	4
20	0.97	0.15	0.78	0.11	-0.80	0.14	3	**44**	1.42	0.20	0.66	0.13	-0.47	0.09	4
21	1.22	0.17	0.18	0.11	-0.15	0.09	3	**45**	0.98	0.17	-1.41	0.14	1.44	0.23	2
22	1.17	0.20	2.03	0.17	-1.74	0.24	4	**46**	1.00	0.16	-1.01	0.12	1.02	0.17	3
23	1.94	0.27	-0.99	0.16	0.51	0.08	4	**47**	1.87	0.36	-3.27	0.37	1.75	0.20	3
24	-0.26	0.13	-1.40	0.11	-5.36	2.68	1	**48**	1.11	0.17	0.86	0.13	-0.78	0.13	3

Tablo 12'de görüldüğü gibi maddelere MTK parametreleri ve MKE göz önüne alınarak bir kalite sınıflaması yapılmıştır. Sınıflama sonucunda 18 madde "çok iyi" (4), 17 madde "iyi" (3), 8 madde "orta" (2) ve 5 madde "yetersiz" (1) olarak kategorilendirilmiştir.

Son olarak deneme uygulaması verileri BTM'ye göre DINA (Deterministic Input Noisy And gate) model kullanılarak analiz edilmiştir (de la Torre, 2008; de la Torre & Lee, 2010; Henson & Douglas, 2005; Henson, Roussos, Douglas, & Xuming He, 2008; J. Leighton & Gierl, 2007). Bu analizler için gerekli olan Q matrisler Ege Üniversitesi Eğitimde Ölçme Değerlendirme yüksek lisans öğrencileri ve ODTU İlköğretim Bölümü Matematik Eğitimi doktora öğrencileri tarafından hazırlanmıştır. Oluşturulan Q matrisler son olarak proje araştırmacı ve

danışmanlarından oluşan uzman ekibi tarafından düzenlenmiştir. BTM analizlerine göre elde edilen bulgular Tablo 13'de verilmiştir.

Tablo 13. Deneme Uygulaması Maddelerine İlişkin BTM Parametreleri

No	G	s.e.	S	s.e	S	No	G	s.e.	S	s.e	S
1	0,39	0,04	0,34	0,03	2	**25**	0,54	0,03	0,27	0,02	1
2	0,16	0,04	0,24	0,03	4	**26**	0,32	0,04	0,28	0,04	3
3	0,35	0,03	0,43	0,04	2	**27**	0,22	0,03	0,82	0,03	1
4	0,02	0,01	0,73	0,04	2	**28**	0,08	0,02	0,61	0,03	3
5	0,29	0,03	0,28	0,04	3	**29**	0,24	0,03	0,29	0,04	3
6	0,11	0,02	0,69	0,03	2	**30**	0,35	0,03	0,29	0,03	3
7	0,09	0,02	0,67	0,03	2	**31**	0,32	0,03	0,10	0,03	4
8	0,36	0,03	0,17	0,05	3	**32**	0,50	0,04	0,16	0,02	3
9	0,18	0,03	0,24	0,04	4	**33**	0,55	0,04	0,17	0,03	2
10	0,35	0,03	0,30	0,04	3	**34**	0,39	0,04	0,07	0,02	4
11	0,38	0,03	0,34	0,04	2	**35**	0,16	0,03	0,27	0,03	4
12	0,50	0,03	0,08	0,03	3	**36**	0,07	0,04	0,01	0,04	4
13	0,01	0,02	0,99	0,01	1	**37**	0,31	0,03	0,23	0,03	3
14	0,00	0,00	0,85	0,03	1	**38**	0,39	0,05	0,35	0,03	2
15	0,43	0,03	0,18	0,05	3	**39**	0,53	0,05	0,43	0,03	1
16	0,09	0,02	0,87	0,02	1	**40**	0,36	0,06	0,36	0,03	3
17	0,61	0,03	0,05	0,02	3	**41**	0,02	0,09	0,53	0,03	3
18	0,63	0,03	0,14	0,03	2	**42**	0,23	0,04	0,03	0,06	4
19	0,42	0,03	0,08	0,03	3	**43**	0,18	0,06	0,40	0,03	3
20	0,19	0,03	0,46	0,04	3	**44**	0,16	0,04	0,24	0,03	4
21	0,48	0,03	0,17	0,03	3	**45**	0,34	0,06	0,41	0,03	2
22	0,19	0,03	0,57	0,04	2	**46**	0,36	0,05	0,25	0,03	3
23	0,33	0,03	0,20	0,03	3	**47**	0,46	0,06	0,39	0,03	1
24	0,10	0,02	0,36	0,04	4	**48**	0,17	0,04	0,29	0,03	4

Tablo 13'de görüldüğü gibi BTM analizlerine göre maddelere ait G (tahmin) ve S (kaydırma) parametreleri hesaplanmıştır. Daha sonra parametreler göz önüne alınarak maddeler için kalite sınıflaması yapılmıştır. Sınıflama sonucunda 10 madde "çok iyi" (4), 20 madde "iyi" (3), 11 madde "orta" (2) ve 7 madde "yetersiz" (1) olarak kategorilendirilmiştir.

3.2.5 Madde havuzunun değerlendirilmesi

Bütün analizler birlikte değerlendirilerek en iyi psikometrik özelliklere sahip maddeler proje kapsamında uygulanacak ön test ve izleme testleri için seçilmiştir. Seçim aşamasında, maddelere ait sınıflama puanlarının ortalamaları dikkate alınmıştır. Her üç kuramla yapılan analizlerde en iyi sonuçları veren ortak maddeler öncelikle seçilmiştir. Bununla birlikte özellikle üst düzey düşünme becerisi, 6. Sınıf kapsamı ve bilişsel süreç becerilerini daha iyi temsil eden maddeler proje araştırmacıları ve danışmanlarının oluşturduğu uzman grup tarafından değerlendirilmiştir.

3.2.6 Ön-test sınavının belirlenmesi

Deneme uygulaması sonuçlarına göre belirlenen madde havuzundan 6. Sınıf programının tamamını temsil edecek ön-son test formları hazırlanmıştır. Bu süreçte uzman grubuyla sorular kalitelerine göre ayrıldıkları sınıflar içinde kazanımları bakımından değerlendirilmiş ve 30 soruluk proje ön-testi belirlenmiştir. Ön-test de örneklemden elde edilecek sonuçların daha geniş bir dağılım göstermesi ve ön-son test arasındaki farkların istatistiksel olarak daha yüksek güvenirlikte olabilmesi için 30 soruluk bir test tasarlanmıştır. Testte yer alan maddelerden 5 tanesi PISA, 6 tanesi TIMSS, 19 tanesi proje sorularından oluşmuştur. Ön-test de PISA ve TIMSS soruları kullanmaktaki amaç deney ve kontrol gruplarının belirlenmesi aşamasında grupların denkliğini sağlama kriterlerini arttırmaktır. Bununla birlikte ön-son test arasındaki farka ilişkin analizler için daha sağlam kanıtlar elde etmektir.

Ön-test hazırlanırken öğrenci performansı ve test uzunluğu göz önüne alınarak PISA, TIMSS uygulamalarındaki gibi 15 soruluk iki oturum planlanmıştır. Her oturum için öğrenci etkileşimini azaltmak amacıyla A ve B kitapçıkları oluşturulmuştur. Aynı zamanda iki oturumda da eşit sayıda açık uçlu (6 adet) soru bulunmaktadır. Test hazırlanırken soruların deneme uygulamalarından elde edilen güçlük indekslerine göre test içindeki yerleri belirlenmiş iki oturumda yer alan iki kitapçık için zor sorular testin ortasına gelecek şekilde dengeli bir dağılımı elde edilmiştir. Her kitapçık için öğrenciler görece kolay sorulardan başlamış ve soru zorluğu giderek arttırılmıştır. Testtin her kitapçık için 11. sorusundan sonra ise madde güçlükleri giderek arttan maddeler seçilmiş ve görece daha kolay sorularla test tamamlanmıştır. Bu dağılım öğrencinin gerçek bilgi düzeyini en az hata kaynağıyla ölçmeye imkan vermektedir.

3.3 Araştırma Örneklemi ve Deney-Kontrol Grupları

Araştırmanın örnekleminin İzmir ili geneli için temsil düzeyini yükseltmek ve deneysel süreç bulgularının genellenebilmesi ve iç-dış geçerlik problemlerinin önüne geçilebilmesi için mümkün olduğu kadar geniş bir popülasyon üzerinde çalışılmıştır. Aynı zamanda araştırma sürecince kullanılacak farklı istatistiksel yöntemlerin varsayımlarını karşılamak ve sınırlılıklarına takılmamak için de örneklem büyüklüğünün geniş tutulmuştur.

3.3.1 Araştırma Örnekleminin Belirlenmesi

Araştırma evrenini İzmir İli Milli Eğitim Müdürlüğüne bağlı 30 ilçe içerisindeki 448 Devlet Ortaokulu oluşturmaktadır. Bu okullarda 6. Sınıf düzeyinde toplam 1822 şube ve 45069 öğrenci yer almaktadır. Örneklem büyüklüğü belirlenirken güven düzeyi (confidence level) ve güven aralığı (confidence interval) istatistiklerinden yararlanılmıştır (Oulte, 2011; Thompson, 2012). Güven düzeyi %99 ve güven aralığı t = 2 olarak belirlendiğinde 45069 kişilik evren için örneklem büyüklüğü n= 3809 belirlenmiştir.

Sonraki aşamada bu örneklem büyüklüğü göz önüne alınarak tabakalı yöntemle tesadüfi küme örneklem yöntemi kullanılarak okul ve sınıflar belirlenmiştir (Lodico, Spaulding, & Voegtle, 2006). Tabakalamada göz önüne alınan kriterler aşağıda sıralanmıştır:

I. Okulların 2015-2016 Kasım - Nisan TEOG genel başarı ortalamaları
II. Okulların 2015-2016 Kasım - Nisan TEOG matematik ortalamaları
III. Okulların 6. Sınıflarda şubeye düşen öğrenci sayıları
IV. İlçelere düşen okul, şube ve öğrenci sayıları

Bu kriterler göz önüne alınarak seçgisiz olarak 10 ilçeden 20 okul araştırma örneklemini oluşturmuştur. Araştırma örnekleminde yer alan okullar şube ve öğrenci sayıları Tablo 14'de verilmiştir.

Tablo 14. Araştırma Örnekleminde Yer Alan Okullar

İLÇE	OKUL	6. SINIF ŞUBE SAYISI	ÖĞRENCİ SAYISI
ALİAĞA	Aliağa Ortaokulu	4	107
BAYRAKLI	Zihni Üstün Ortaokulu	10	282
BORNOVA	Yavuz Selim Ortaokulu	5	182
	Doktor Cavit Özyeğin Ortaokulu	6	169
	Bornova Karacaoğlan Mahallesi Ortaokulu	9	233
	Evrenosoğlu Ortaokulu	6	190
BUCA	Hasan Ali Yücel Ortaokulu	4	182
	30 Ağustos Ortaokulu	8	292
	Saadet Emir Ortaokulu	14	465
ÇİĞLİ	Çiftehavuzlar Ortaokulu	4	128
GAZİEMİR	Nevvar Salih İşgören Ortaokulu	10	240
	Remzi Doğan Ortaokulu	6	160
KARABAĞLAR	Şerif Remzi Ortaokulu	7	214
	Ali Akatlar Ortaokulu	5	192
KARŞIYAKA	Eren Şahin Eronat Ortaokulu	7	272
	Selçuk Yaşar Alaybey Ortaokulu	7	200
KONAK	Güzelyalı Ortaokulu	18	596
	Gazi Ortaokulu	7	186
	Gültepe Ortaokulu	7	173
BALÇOVA	Asil Nadir Ortaokulu	4	129
BORNOVA	Suphi Koyuncuoğlu Ortaokulu*	7	249
TOPLAM		148	4592

Tablo 14'de görüldüğü gibi İzmir ilinde yer alan okullar ilçelere göre belirlenmiş ve araştırma örneklemi için 20 okulda toplam 148 şube ve 4592 öğrenci araştırmaya dahil edilmiştir. Örneklem büyüklüğünün artmasının temel sebebi okullarda yer alan 6. Sınıf şubeleri arasında seçim yapılmamasıdır. Bu sayede hem örneklemin evreni temsil gücü arttırılmış hem de bir okulda yer alan şubelerin bazıların uygulamadan mahrum kalmaması ve okul için eşitliğin bozulmaması sağlanmıştır. Örneklem büyüklüğünün son hali ile örneklemin güven aralığı t= 1,77 düzeyine düşürülmüş yani evreni temsil etme gücü arttırılmıştır.

Bununla birlikte deneme uygulamasındaki soru sayısının kısıtlı olması, projedeki soruların değerlendirme sürecinin uzaması ve ikinci eğitim sonunda proje soru yazarı ekibine dahil olan yeni öğretmenlerin yazdığı soruların sisteme sürekli eklenmesinden dolayı örneklemin dışında evren parametreleri ortalamalarına en yakın düzeyde bulunan ve örneklem büyüklüğü açısından yeterli görünen Bornova İlçesi Suphi Koyuncuoğlu Ortaokulu pilot uygulama okulu olarak belirlenmiştir. Belirlenen pilot okul hazırlanan her testti önceden alarak genel örneklemdeki tasarım ve uygulama hata ve eksikliklerinin ortaya çıkartılmasında yardımcı olmuştur. Bununla birlikte pilot uygulamada yapı geçerliğine ilişkin çalışmaların yapılması da amaçlanmıştır.

3.3.2 Deney ve Kontrol Gruplarının Belirlenmesi

Ön-test uygulaması sonuçlarına göre araştırma örneklemini oluşturan 20 okul öğrenci ve şube sayılarına göre deney ve kontrol grupları olarak belirlenmiştir. Bu aşamada öğrencilerin ön test puan ortalamalarına göre okul başarı sıralamaları oluşturulmuş ve Deney-1, Deney-2 ve Kontrol grupları belirlenmiştir. Grup oluşturma aşamasında okullar küme örneklem olarak alınmıştır. Bunun öncelikli sebebi aynı okulda yer alan farklı şubelere farklı uygulama yapmamaktır. Okul içindeki bazı 6. sınıfların araştırmada farklı manipülasyona tabi tutulması okul iklimi ve idareci-veli-öğretmen ilişkisinde sorunlar çıkartabileceği düşünülmüştür. Pilot ve ön-test uygulaması sırasında uygulamaların özellikle 6. sınıflara yönelik olduğu belirtilmesine rağmen okul yöneticilerinden 5. ve 7. sınıf velilerinin kendi çocuklarının da bu uygulamaya katılması konusunda ısrar ettikleri bilgisi proje ekibine defalarca ulaştırılmıştır. Bu noktada okulların uygulamaları bütün olarak alması hem sonuçların bilimsel tutarlılığı hem de uygulamanın okul içi problemlere neden olmaması bakımından tercih edilmiştir.

Araştırma örnekleminde yer alan 20 okula yukarıda belirtilen kriterler göz önüne alınarak Tablo 15'de gösterildiği biçimde gruplara atanmıştır.

Tablo 15. Araştırma Örneklemi Deney ve Kontrol Grupları

GRUP	KOD	OKUL	İLÇE	ŞUBE	N
DENEY-1	28	ASİL NADİR ORTAOKULU	BALÇOVA	4	129
	03	YAVUZ SELİM ORTAOKULU	BORNOVA	5	182
	18	REMZİ DOĞAN ORTAOKULU	GAZİEMİR	6	160
	05	HASAN ALİ YÜCEL ORTAOKULU	BUCA	4	182
	06	NEVVAR SALİH İŞGÖREN ORTAOKULU	GAZİEMİR	10	240
	16	DOKTOR CAVİT ÖZYEĞİN ORTAOKULU	BORNOVA	6	169
	20	GAZİ ORTAOKULU	KONAK	7	186
	22	BORNOVA KARACAOĞLAN MAHALLESİ O.O.	BORNOVA	9	233
Toplam				51	1481

DENEY-2	07	EREN ŞAHİN ERONAT ORTAOKULU	KARŞIYAKA	7	272
	19	ŞERİF REMZİ ORTAOKULU	KARABAĞLAR	7	214
	27	SAADET EMİR ORTAOKULU	BUCA	14	465
	23	EVRENOSOĞLU ORTAOKULU	BORNOVA	6	190
	17	30 AĞUSTOS ORTAOKULU	BUCA	8	292
Toplam				42	1433
KONTROL	14	ÇİFTEHAVUZLAR ORTAOKULU	ÇİĞLİ	4	128
	10	GÜZELYALI ORTAOKULU	KONAK	18	596
	15	ALİAĞA ORTAOKULU	ALİAĞA	4	107
	21	ZİHNİ ÜSTÜN ORTAOKULU	BAYRAKLI	10	282
	13	SELÇUK YAŞAR ALABEY ORTAOKULU	KARŞIYAKA	7	200
	24	ALİ AKATLAR ORTAOKULU	KARABAĞLAR	5	192
	25	GÜLTEPE ORTAOKULU	KONAK	7	173
Toplam				55	1678

Tablo 15'de görüldüğü gibi her okul için özgün bir Kod belirlenmiştir. Araştırma süresince rapor ve geri bildirimlerde okullar için bu kodlar kullanılmış ve gerek okulların gerekse idari kurumların hangi okulların araştırmaya katıldığı ve hangi grup içinde olduğuna dair bilgiler gizli tutulmuştur.

3.4 Deneysel İşlem Süreci

Araştırmanın deneysel işlem basamaklarında ön test uygulaması, izleme testleri ve daha sonrasında gerçekleştirilen geri bildirim süreci ve son test uygulaması yer almaktadır. Tablo 16 araştırmanın deneysel sürecini göstermektedir.

Tablo 16. Deneysel Desen Basamakları

			İzleme 1			İzleme 2			İzleme 3			İzleme 4		
Grup	Öntest	ED*	Test	GB**	ED	Test	GB	ED	Test	GB	ED	Test	GB	Sontest
Deney-1	X	→	X	X	→	X	X	→	X	X	→	X	X	X
Deney-2	X	→	X		→	X		→	X		→	X		X
Kontrol	X	→			→			→			→			X

* ED: Eğitim Durumları (Matematik Dersi Öğrenme ve Öğretme Süreçleri)
** GB: Geribildirim (BTM'ye göre verilen geribildirim).

Tablo 16'da görüldüğü gibi araştırmanın deneysel süreci üç grup üzerinden yürütülmüştür. Deney grupları araştırma içinde iki farklı grup olarak ayrılmıştır. Deney-1 grubu, izleme testlerinin uygulandığı ve aynı zamanda da uygulama sonucunda BTM analizlerine göre kazanım, bilişsel süreç ve hata konularında detaylı geri bildirim almaktadırlar. Deney -2 grubu test etkisinin belirlenmesi amacıyla oluşturulmuştur. Bu gruba da izleme testleri uygulanmış sadece testler sonucunda toplam puan üzerinden bir geri bildirim verilmiştir. Deney-1 ve

Deney-2 grubu arasında fark, araştırmada BTM'ye göre geri bildirim vermenin etkisini ortaya koyacaktır. Kontrol grubu ise sadece ön ve son-testleri almışlardır. Bu grup ise hem izleme testlerinin hem de geri bildirimin etkisini ortaya koymak amacı ile analizlere dahil edilmiştir.

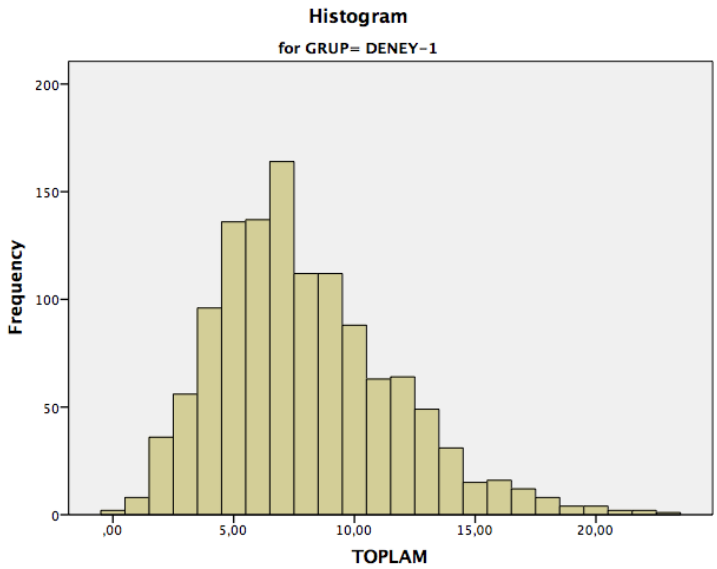
3.4.1 Ön-test uygulaması

Ön test uygulaması 20 ayrı okulda 6. Sınıf düzeyinde 55 sınıfta toplam 4419 öğrenciye uygulanmıştır. Ön test betimsel istatistikleri Tablo 17'de verilmiştir.

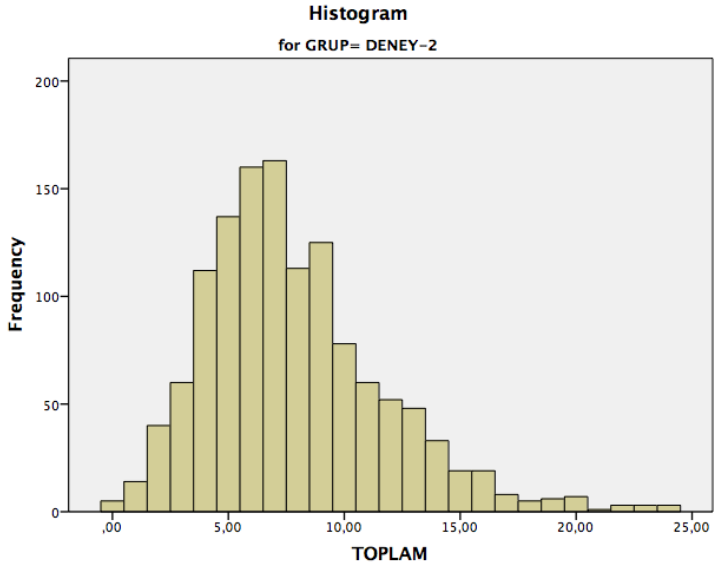
Tablo 17. Deney ve Kontrol Grupları Betimsel İstatistikleri

	DENEY-1	DENEY-2	KONTROL
N	1481	1433	1678
ORTALAMA	7,96	7,84	7,91
MEDYAN	7,00	7,00	7,00
VARYANS	13,66	15,06	14,55
STANDART S	3,70	3,88	3,81
MİNİMUM	0	0	0
MAKSİMUM	23	24	22
RANJ	23	24	22
ÇARPIKLIK	0,774	0,975	0,72

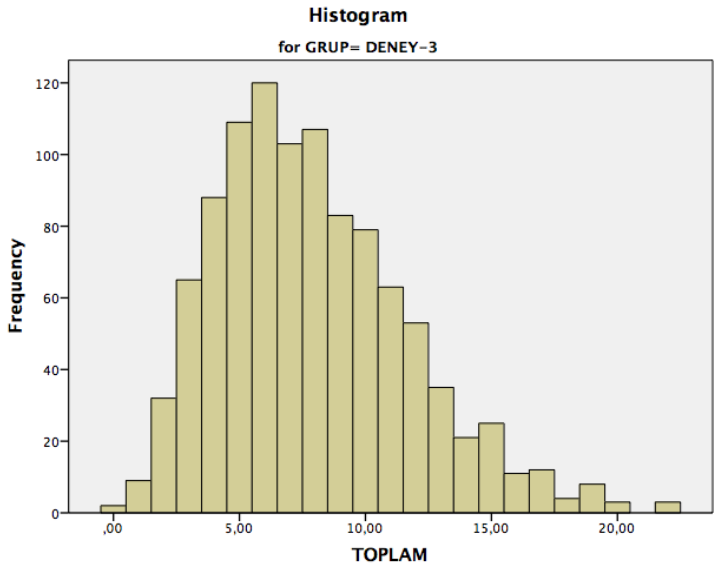
Tablo 17'de görüldüğü gibi Deney-1 grubunda 8 okul (N= 1481), Deney-2 grubunda 5 okul (N=1433) ve Kontrol grubunda 7 okul (N=1678) yer almaktadır. Aşağıda Deney-1, Deney-2 ve Kontrol gruplarının histogram grafikleri Şekil 8, Şekil 9 ve Şekil 10'da verilmiştir.



Şekil 8. Deney-1 Ön-Test puanlarına Ait Histogram Grafiği



Şekil 9. Deney-2 Ön-Test puanlarına Ait Histogram Grafiği



Şekil 10. Kontrol Ön-Test puanlarına Ait Histogram Grafiği

Tablo 17 ve şekiller incelendiğinde gruplara ait puan dağılımlarının normal dağılımdan aşırı sapma göstermediği görülmektedir. Grupların düzeylerinin denk olup olmadığı varyans analizi ile incelenmiştir. Tek yönlü ANOVA sonuçları Tablo 18'de verilmiştir.

Tablo 18. Deney ve Kontrol Gruplarına ilişkin Varyans Analizi Sonuçları

Betimsel İstatistikler				ANOVA sonuçları					
	N	**Ortalama**	**S.S.**	**Varyans Kaynağı**	**Kareler Toplamı**	**Kareler Ortalaması**	**Sd.**	**F**	**p**
DENEY-1	1452	7,957	3,696	**Gruplar içi**	9,120	4,560	2	0,316	0,729
DENEY-2	1414	7,837	3,880	**Gruplar arası**	50836,542	14,426	4376		
KONTROL	1513	7,905	3,814						
Total	4379	7,898	3,797	**Toplam**	50845,662		4378		

Tablo 18 incelendiğinde Deney-1, Deney-2 ve Kontrol gruplarının ön-test puan ortalamaları arasında istatistiksel olarak anlamlı bir fark olmadığı görülmektedir ($F_{2\text{-}4376}$ = 0,316, p>0,05). Bu durum araştırma gruplarının denkliğinin bir göstergesidir.

3.4.2 Ön-test Psikometrik Özellikleri

Ön test uygulamasında yer alan soruların psikometrik özellikleri KTK ve MTK kullanılarak incelenmiştir. Öncelikle KTK için madde güçlüğü ve madde test korelasyonları hesaplanmıştır. Tablo 19'da KTK analizi sonuçları verilmiştir.

Tablo 19. Ön-Test KTK Madde Analizleri

No	P_j	r_{jx}	r_{pbis}	r_{bis}	No	P_j	r_{jx}	r_{pbis}	r_{bis}
1	0,593	0,327	0,372	0,345	**16**	0,594	0,456	0,365	0,462
2	0,193	0,334	0,321	0,430	**17**	0,465	0,524	0,411	0,516
3	0,370	0,401	0,344	0,439	**18**	0,062	0,337	0,487	0,566
4	0,415	0,631	0,499	0,631	**19**	0,153	0,311	0,480	0,580
5	0,264	0,368	0,349	0,471	**20**	0,359	0,323	0,404	0,462
6	0,296	0,336	0,399	0,495	**21**	0,220	0,296	0,328	0,392
7	0,090	0,309	0,479	0,669	**22**	0,206	0,313	0,379	0,396
8	0,038	0,277	0,418	0,506	**23**	0,375	0,442	0,486	0,593
9	0,114	0,139	0,407	0,442	**24**	0,116	0,284	0,332	0,381
10	0,157	0,283	0,342	0,366	**25**	0,167	0,265	0,327	0,487
11	0,258	0,338	0,325	0,339	**26**	0,310	0,273	0,356	0,336
12	0,136	0,274	0,469	0,580	**27**	0,179	0,400	0,583	0,708
13	0,291	0,289	0,335	0,411	**28**	0,400	0,498	0,430	0,545
14	0,401	0,442	0,369	0,467	**29**	0,310	0,309	0,403	0,497
15	0,404	0,364	0,406	0,488	**30**	0,022	0,264	0,555	0,710
Ortalama						0,265	0,347	0,405	0,490

Tablo 19'da maddelerin güçlük değerleri (P_j), ayırt edicilik parametreleri (r_{jx}) ve nokta çift serili (r_{pbis}) ve çift serili (r_{bis}) korelasyon katsayıları verilmiştir. Görüldüğü gibi ön-test maddelerinin gerek ayırt edicilik değerleri gerekse madde test korelasyonları test maddelerinin yeterli ayırt ediciliğe sahip olduğunu göstermektedir. Ön-testte ilişkin güvenirlik analizlerinde KR-20 katsayısı incelenmiştir. Tüm test için KR-20 = 0,72 bulunmuştur. Okul düzeyinde yapılan analizlerde testtin iç tutarlılık katsayısı 0,64 ile 0,76 arasında değişmektedir. Bu durum testin yeterli güvenirliğe sahip olduğunu göstermektedir.

Bunula birlikte test güvenirliği MTK ile de incelenmiştir. MTK'da marjinal güvenirlik Rasch Model'de (1 Parametreli Model) 0,96 ve 2 Parametreli Model'de 0,72 tahminlenmiştir.

Analizlerde veri setinin 1-0 formatına dönüştürüldüğü göz önüne alındığında Rasch Modelin veri ile daha uyumlu olduğu anlaşılmaktadır. Ön test uygulamasına ilişkin MTK madde parametreleri (EK 4), Madde Karakteristik eğrileri (EK 5) Test karakteristik Eğrisi (EK 6) ve Test Bilgi Fonksiyonu (EK 7) eklerde verilmiştir.

3.4.3 Geri bildirim sistemi

Araştırma kapsamında Deney-1 ve Deney-2 grupları arasındaki temel manipülasyon geri bildirim üzerinden kurgulanmıştır. Bu amaç doğrultusunda öğrenciye uygulan izleme testlerinin hemen ardından Deney-1 gruplarına detaylı geri bildirim formları verilmiştir (EK 8). Deney-2 gruplarına ise sadece doğru ve yanlış sayılarını gösteren bir geri bildirim formu verilmiştir (EK 9). Detaylı geri bildirim formu testten elde edilen bir öğrenci profilidir. Bu profil içinde öğrencinin hangi özelliklere ne olasılıkla sahip olduğu, hangi üst düzey becerileri gösterebildiği ve en önemlisi hangi tür hataları tekrar ettiği görülmektedir. Deney-1 ve Deney-2 grupları arasındaki bu geri bildirim farklılığı aynı zamanda test etkisini görmeye de imkan vermektedir. BTM modellerine uygun olarak tasarlanan yazılım ile öğrencilerin kazanım, yeterlik ve hata profilleri dijital olarak üretilebilmiş aynı zamanda da bu sistem özellikle açık uçlu sorular için oldukça yeni ve esnek bir geribildirim modeli oluşturmuştur.

3.4.4 İzleme Testleri

Araştırmanın deneysel işlem sürecinde Deney-1 ve Deney-2 gruplarına 4 adet izleme testi uygulanmıştır. Her izleme testi uygulamasından önce testin pilot uygulaması gerçekleştirilmiştir. Test takvimi Tablo 20'de verilmiştir.

Tablo 20. Deney Süreci Sınav Takvimi

Test	Tarih	Kişi Sayısı	kod
ÖNTEST	02-04.11.2016	4730	11
İZLEME_1 PİLOT	17.11.2016	231	20
İZLEME_1	05-07.12.2016	2918	21
İZLEME_2 PİLOT	23.12.2016	230	30
İZLEME_2	09-11.01.2017	2832	31
İZLEME_3 PİLOT	22.3.2017	238	40
İZLEME_3	04-06.04.2017	2875	41
İZLEME_4 PİLOT	4.05.2017	219	50
İZLEME_4	16-18.05.2017	2837	51
SONTEST	22.05-02.06.2017	4299	61
KALICILIK			71

3.4.4.1 İzleme-1 Uygulaması

İlk izleme uygulamasının pilot uygulaması 17/11/2016 tarihinde 231 öğrenciye uygulanmıştır. Bu uygulama madde parametreleri EK 10'da verilmiştir. İzleme-1'e ait betimsel istatistikler Tablo 21'de verilmiştir.

Tablo 21. İzleme-1(21) Testi Betimsel İstatistikleri

Soru Sayısı	11
Ortalama	3,65
Medyan/ortanca	3
Standart Sapma	2,07
Varyans	4,28
Çarpıklık	0,63
Basıklık	0,02

Tablo 21'de verildiği gibi testin puan ortalaması 3,65 ve Medyan/ortanca puanının 3 olduğu görülmektedir. Teste dair standart sapma 2.07'dir ve varyans 4,48 olarak hesaplanmıştır. Çarpıklık ve basıklık değerleri ise sırasıyla 0,626 ve 0,016 olarak bulunmuştur. Çarpıklık değerinin 1'den küçük oluşu dağılımın normal dağılımdan aşırı sapma göstermediği anlamına gelmektedir. İzleme-1 testine ilişkin okul ve şube düzeyindeki istatistikler EK 11'de verilmiştir. İzleme-1 testine ait madde güçlük ve ayırt edicilik parametreleri Tablo 22'de listelenmiştir.

Tablo 22. İzleme-1(21) Testi Madde Parametreleri

No	Soru Kodu	Madde Güçlüğü (Pj)	Madde Ayırt Ediciliği (rbis)
1	9796	0,41	0,5
2	3690	0,21	0,22
3	1027	0,18	0,45
4	1025	0,27	0,55
5	1033	0,51	0,48
6	1021	0,68	0,45
7	1032	0,4	0,41
8	7339	0,01	0,19
9	6728	0,29	0,38
10	5158_B	0,38	0,52
11	5158_C	0,31	0,49
	Ortalama	0,33	0,43

Tablo 22'de görüldüğü gibi 1. İzlemede bulunan her bir maddeye dair madde güçlükleri 0,18 ile 0,68 arasında, madde ayırt edicilik değerleri ise 0,22 ile 0,55 arasında değerler almıştır. Ortalama madde güçlüğü 0,33 ve ortalama ayırt edicilik değeri 0,42 olarak hesaplanmıştır. Buradan maddelerin öğrencilerin büyük çoğunluğuna zor geldiği söylenebilir. Ayrıca teste ait ortalama çift serili korelasyon 0,42, ortalama nokta çift serili korelasyon 0,24 olarak

hesaplanmıştır. Üst grup (n=816) için en düşük puan 5 ve alt grup(n=867) için en yüksek puan ise 2'dir.

İzleme-1 uygulaması için, öğrencilerin matematiksel yeterliklerini, kazanımlara sahip olma düzeyleri ve hatalara göre hazırlanan 3 farklı Q matris kullanılarak CDM analizleri yapılmıştır. İzleme-1 için elde edilen Q Matrisler EK 12'de verilmiştir.

Yapılan BTM analizleri ile elde edilen madde parametreleri Tablo 23'de listelenmiştir.

Tablo 23. İzleme-1(21) BTM Analizlerine Göre Madde Parametreleri

	Matematiksel Yeterlikler				**Kazanımlar**				**Hatalar**			
No	**g**	**gse**	**s**	**sse**	**g**	**gse**	**s**	**sse**	**g**	**gse**	**s**	**sse**
9796	0,261	0,012	0,299	0,02	0,038	0,023	0,045	0,033	0,295	0,02	0,267	0,012
3690	0,189	0,009	0,694	0,035	0,192	0,009	0,731	0,025	0,77	0,02	0,198	0,009
1027	0,106	0,008	0,174	0,054	0,101	0,008	0,527	0,025	0,563	0,023	0,096	0,008
1025	0,122	0,009	0,329	0,023	0,122	0,009	0,334	0,023	0,216	0,029	0,115	0,009
1033	0,368	0,013	0,213	0,019	0,393	0,012	0,182	0,02	0,199	0,02	0,382	0,013
1021	0,561	0,013	0,103	0,014	0,584	0,012	0,086	0,014	0,083	0,015	0,597	0,011
1032	0,269	0,016	0,382	0,025	0,089	0,03	0,146	0,043	0,41	0,022	0,314	0,012
7339	0,001	0,001	0,969	0,007	0,00	0,005	0,969	0,008	0,971	0,007	0,001	0,001
6728	0,227	0,011	0,539	0,022	0,238	0,01	0,371	0,04	0,563	0,02	0,219	0,011
5158_B	0,256	0,011	0,18	0,023	0,209	0,012	0,202	0,022	0,22	0,023	0,258	0,011
5158_C	0,211	0,01	0,311	0,025	0,169	0,011	0,351	0,023	0,356	0,024	0,205	0,01
Ort.	0,234	0,01	0,381	0,024	0,194	0,013	0,359	0,025	0,422	0,02	0,241	0,01

Tablo 23'de görüldüğü gibi BTM analizi sonucunda maddelerin g parametreleri ortalamaları matematiksel yeterlikler için 0,234, kazanımlar için 0,194 ve hatalar için 0,422; s parametrelerinin ortalamaları ise matematiksel yeterlikler için 0,381, kazanımlar için 0,359 ve hatalar için 0,241dir. Her iki parametrenin matematiksel yeterlikler, kazanımlar ve hatalar için g ve s parametre değerleri testtin model uyumunun yeterli olduğunu göstermektedir.

3.4.4.2 İzleme-2 Uygulaması

İzleme-2 için 23/12/2016 tarihinde 230 öğrenciye yapılan Pilot uygulama madde parametreleri EK 13'de verilmiştir. Yapılan analizler sonucunda 2 madde testten çıkarılmış ve gerçek uygulamaya geçilmiştir.

İzleme-2, Pilot uygulamada kullanılan maddeler arasında yeterli ayırt ediciliğe sahip ve düzeltilen 10 maddeden oluşmuş ve Deney-1 ve Deney-2 gruplarına uygulanmıştır. Uygulamaya toplam 2832 öğrenci katılmıştır. İzleme-2 ait betimsel istatistikler Tablo 24'de verilmiştir.

Tablo 24. İzleme-2(31) Testi Betimsel İstatistikleri

Soru sayısı	10
Ortalama	2,68
Medyan/ortanca	3
Standart Sapma	1,66
Varyans	2,74
Çarpıklık	0,66
Basıklık	0,56

Tablo 24'de görüldüğü gibi testin puan ortalaması 2,68'dir. Medyan/ortanca puan 3 tür. Teste dair standart sapma 1,66 ve varyans 2,74 olarak hesaplanmıştır. Çarpıklık ve basıklık değerleri ise sırasıyla 0,65 ve 0,56 olarak hesaplanmıştır. Çarpıklık değerinin 1'den küçük oluşu dağılımın normal dağılımdan aşırı sapma göstermediği anlamına gelmektedir. İzleme-2 testine ilişkin okul ve şube düzeyindeki istatistikler EK 14'de verilmiştir. İzleme-2 testine ait madde güçlük ve ayırt edicilik parametreleri Tablo 25'te listelenmiştir.

Tablo 25. İzleme-3(31) Testi Madde Parametreleri

No	Soru Kodu	Madde Güçlüğü (Pj)	Madde Ayırt Ediciliği (rbis)
1	1011_A	0,29	0,45
2	1011_B	0,27	0,46
3	1012_A	0,29	0,40
4	1011_C	0,35	0,40
5	1012_B	0,30	0,38
6	1014_A	0,43	0,45
7	1015_A	0,08	0,32
8	1014_B	0,05	0,32
9	1015_B	0,38	0,43
10	1014_C	0,24	0,27
	Ortalama	0,27	0,39

Tablo 25'te görüldüğü gibi İzleme-2 testinde bulunan her bir maddeye dair madde güçlükleri 0,05 ile 0,43 arasında, madde ayırt edicilik değerleri ise 0,27 ile 0,46 arasında değerler almıştır. Ortalama madde güçlüğü 0,27 ve ortalama ayırt edicilik değeri 0,39 olarak hesaplanmıştır. Buradan maddelerin, öğrencilerin büyük çoğunluğuna zor geldiği söylenebilir. Ayrıca teste ait ortalama çift serili korelasyon 0,39, ortalama nokta çift serili korelasyon 0,35 olarak hesaplanmıştır. Üst grup (n=775) için en düşük puan 4 ve alt grup(n=1387) için en yüksek puan ise 2'dir.

2832 öğrencinin katıldığı İzleme-2 uygulaması için, öğrencilerin matematiksel yeterliklerini, kazanımlara sahip olma düzeyleri ve hatalara göre hazırlanan 3 farklı Q matris EK 15'de verilmiştir. Yapılan CDM analizleri ile elde edilen madde parametreleri Tablo 26'da listelenmiştir.

Tablo 26. İzleme-2(31) BTM Analizlerine Göre Madde Parametreleri

	Matematiksel Yeterlikler				Kazanımlar				Hatalar			
No	**g**	**gse**	**s**	**sse**	**g**	**gse**	**s**	**sse**	**g**	**gse**	**s**	**sse**
1011_A	0,113	0,019	0,451	0,028	0,224	0,011	0,387	0,035	0,419	0,033	0,224	0,011
1011_B	0,170	0,011	0,425	0,028	0,209	0,009	0,290	0,040	0,400	0,031	0,174	0,011
1012_A	0,250	0,010	0,048	0,077	0,128	0,018	0,481	0,026	0,560	0,028	0,250	0,011
1011_C	0,294	0,012	0,485	0,027	0,313	0,010	0,392	0,038	0,482	0,028	0,300	0,011
1012_B	0,260	0,009	0,176	0,067	0,186	0,016	0,557	0,022	0,582	0,030	0,263	0,011
1014_A	0,325	0,014	0,330	0,025	0,362	0,012	0,261	0,033	0,306	0,029	0,352	0,012
1015_A	0,043	0,005	0,699	0,032	0,046	0,005	0,617	0,041	0,768	0,023	0,041	0,005
1014_B	0,001	0,004	0,830	0,016	0,021	0,004	0,682	0,035	0,799	0,021	0,014	0,004
1015_B	0,312	0,011	0,211	0,043	0,064	0,031	0,164	0,043	0,205	0,050	0,318	0,011
1014_C	0,228	0,011	0,743	0,021	0,107	0,040	0,664	0,032	0,724	0,040	0,230	0,010
Ort.	0,200	0,011	0,440	0,036	0,166	0,016	0,450	0,035	0,525	0,031	0,217	0,010

Maddelerin *g* parametreleri ortalamaları matematiksel yeterlikler için 0,200, kazanımlar için 0,16 ve hatalar için 0,525; *s* parametrelerinin ortalamaları ise matematiksel yeterlikler için 0,440, kazanımlar için 0,450 ve hatalar için 0,217dir. Her iki parametrenin matematiksel yeterlikler, kazanımlar ve hatalar için hata değerlerinin de genel olarak düşük olduğu görülmektedir.

3.4.4.3 İzleme-3 Uygulaması

İzleme-3 uygulamasının pilot uygulaması 22/3/2017 tarihinde 238 öğrenciye uygulanmıştır. Bu uygulama madde parametreleri EK 16'da verilmiştir. İzleme-3'e ait betimsel istatistikler aşağıda Tablo 27'de verilmiştir.

Tablo 27. İzleme-3(41) Testi Betimsel İstatistikleri

Soru sayısı	13
Ortalama	3,71
Medyan/ortanca	4
Standart Sapma	1,94
Varyans	3,75
Çarpıklık	0,55
Basıklık	0,29

Tablo 27'de görüldüğü gibi testin puan ortalaması 3,71dir. Medyan/ortanca puan 4 tür. Teste dair standart sapma 1,94'tür ve varyans 3,75 olarak hesaplanmıştır. Çarpıklık ve basıklık değerleri ise sırasıyla 0,55 ve 0,29 olarak hesaplanmıştır. Çarpıklık değerinin 1'den küçük oluşu dağılımın normal dağılımdan aşırı sapma göstermediği anlamına gelmektedir. İzleme-3 testine ilişkin okul ve şube düzeyindeki istatistikler EK 17'de verilmiştir. İzleme-3 testine ait madde güçlük ve ayırt edicilik parametreleri Tablo 28'de listelenmiştir.

Tablo 28. İzleme-3(41) Testi Madde Parametreleri

No	Soru Kodu	Madde Güçlüğü (Pj)	Madde Ayırt Ediciliği (rbis)
1	1023_A	0,42	0,39
2	1023_B	0,52	0,46
3	1024	0,55	0,49
4	1028	0,39	0,45
5	1022_A	0,07	0,06
6	1022_B	0,33	0,37
7	1026	0,18	0,25
8	1020_A	0,29	0,41
9	1020_B	0,05	0,29
10	1020_C	0,19	0,22
11	1021_A	0,13	0,4
12	1021_B	0,43	0,44
13	1021_C	0,16	0,22
	Ortalama	0,29	0,34

Tablo 28'de görüldüğü gibi İzleme-3'de bulunan her bir maddeye dair madde güçlükleri 0,05 ile 0,55 arasında, madde ayırt edicilik değerleri ise 0,22 ile 0,49 arasında değerler almıştır. Ortalama madde güçlüğü 0,29 ve ortalama ayırt edicilik değeri 0,34 olarak hesaplanmıştır. Buradan maddelerin öğrencilerin büyük çoğunluğuna zor geldiği söylenebilir. Ayrıca teste ait ortalama çift serili korelasyon 0,34, ortalama nokta çift serili korelasyon 0,31 olarak hesaplanmıştır. Üst grup (n=816) için en düşük puan 5 ve alt grup(n=867) için en yüksek puan ise 2dir.

2875 öğrencinin katıldığı İzleme-3 uygulaması için, öğrencilerin matematiksel yeterliklerini, kazanımlara sahip olma düzeyleri ve hatalara göre hazırlanan 3 farklı Q matris EK 18'de verilmiştir. Yapılan CDM analizleri ile elde edilen madde parametreleri Tablo 29'da listelenmiştir.

Tablo 29. İzleme-3(41) BTM Analizlerine Göre Madde Parametreleri

	Matematiksel Yeterlikler				Kazanımlar				Hatalar			
No	g	gse	s	sse	g	gse	s	sse	g	gse	s	sse
1	0,346	0,011	0,336	0,026	0,348	0,011	0,335	0,026	0,363	0,023	0,334	0,012
2	0,356	0,014	0,186	0,021	0,384	0,013	0,170	0,021	0,131	0,022	0,420	0,012
3	0,396	0,013	0,024	0,022	0,399	0,013	0,102	0,019	0,064	0,019	0,423	0,012
4	0,254	0,013	0,345	0,022	0,306	0,011	0,283	0,026	0,290	0,025	0,293	0,011
5	0,100	0,007	0,989	0,007	0,102	0,008	0,980	0,008	0,980	0,008	0,098	0,008
6	0,172	0,021	0,517	0,023	0,278	0,010	0,493	0,026	0,418	0,030	0,244	0,012
7	0,169	0,009	0,776	0,021	0,168	0,009	0,787	0,018	0,777	0,021	0,169	0,009
8	0,222	0,010	0,460	0,026	0,226	0,010	0,464	0,026	0,375	0,032	0,185	0,012
9	0,023	0,004	0,843	0,017	0,025	0,004	0,819	0,021	0,863	0,014	0,018	0,004
10	0,190	0,009	0,801	0,021	0,189	0,010	0,802	0,019	0,802	0,019	0,189	0,010
11	0,067	0,006	0,601	0,025	0,013	0,009	0,545	0,029	0,609	0,024	0,065	0,006

12	0,264	0,015	0,242	0,026	0,207	0,018	0,219	0,026	0,235	0,026	0,336	0,011
13	0,151	0,009	0,833	0,018	0,151	0,009	0,833	0,018	0,829	0,016	0,148	0,010
Ort.	0,208	0,011	0,535	0,021	0,215	0,010	0,526	0,022	0,518	0,021	0,225	0,010

Tablo 29'da görüldüğü gibi maddelerin *g* parametreleri ortalamaları matematiksel yeterlikler için 0,208; kazanımlar için 0,215 ve hatalar için 0,518; *s* parametrelerinin ortalamaları ise matematiksel yeterlikler için 0,535; kazanımlar için 0,526 ve hatalar için 0,225'dir. Her iki parametrenin matematiksel yeterlikler, kazanımlar ve hatalar için hata değerlerinin de genel olarak düşük olduğu görülmektedir.

3.4.4.4 İzleme-4 Uygulaması

İzleme-4 testinin pilot uygulaması 04/05/2017 tarihinde 219 öğrenciye uygulanmıştır. Bu uygulama madde parametreleri EK 19'da verilmiştir. İzleme-3 testine ait betimsel istatistikler aşağıda Tablo 30'da verilmiştir.

Tablo 30. İzleme-4(51) Testi Betimsel İstatistikleri

Soru sayısı	9
Ortalama	3,34
Medyan/ortanca	3
Standart Sapma	1.57
Varyans	2,48
Çarpıklık	0,29
Basıklık	-0,01

Tablo 30'da verildiği gibi testin puan ortalaması 3,34 ve Medyan/ortanca puan 3 olduğu görülmektedir. Teste dair standart sapma 1,57'dir ve varyans 2,48 olarak hesaplanmıştır. Çarpıklık ve basıklık değerleri ise sırasıyla 0,299 ve 0,018 olarak bulunmuştur. Çarpıklık değerinin 1'den küçük oluşu dağılımın normal dağılımdan aşırı sapma göstermediği anlamına gelmektedir. İzleme-4 testine ilişkin okul ve şube düzeyindeki istatistikler EK 20'de verilmiştir. İzleme-4 testine ait madde güçlük ve ayırt edicilik parametreleri Tablo 31'de listelenmiştir.

Tablo 31. İzleme-4(51) Testi Madde Parametreleri

No	**Soru Kodu**	**Madde Güçlüğü (P_j)**	**Madde Ayırt Ediciliği (r_{bis})**
1	5017_A	0,3	0,42
2	5017_B	0,02	0,23
3	5036_A	0,42	0,54
4	5036_B	0,25	0,37
5	5036_C	0,14	0,26

6	5058_A	0,74	0,5
7	5058_B	0,53	0,52
8	5059	0,22	0,42
9	5057	0,74	0,44
	Ortalama	0,37	0,41

Tablo 31'de görüldüğü gibi 4. İzlemede bulunan her bir maddeye dair madde güçlükleri 0,02 ile 0,74 arasında, madde ayırt edicilik değerleri ise 0,23 ile 0,54 arasında değerler almıştır. Ortalama madde güçlüğü 0,37 ve ortalama ayırt edicilik değeri 0,41 olarak hesaplanmıştır. Buradan maddelerin öğrencilerin büyük çoğunluğuna zor geldiği söylenebilir. Ayrıca teste ait ortalama çift serili korelasyon 0,41, ortalama nokta çift serili korelasyon 0,24, KR20(Alpha) 0,57, KR-21 değeri 0,41, SEM (KR-21'den) değeri 1,38 olarak hesaplanmıştır. Üst grup (n=1229) için en düşük puan 4 ve alt grup (n=852) için en yüksek puan ise 2'dir.

2834 öğrencinin katıldığı İzleme-4 uygulaması için, öğrencilerin matematiksel yeterliklerini, kazanımlara sahip olma düzeyleri ve hatalara göre hazırlanan 3 farklı Q matris EK 21'de verilmiştir. Yapılan CDM analizleri ile elde edilen madde parametreleri Tablo 32'de listelenmiştir.

Tablo 32. İzleme-4(51) BTM Analizlerine Göre Madde Parametreleri

	Matematiksel Yeterlikler				**Kazanımlar**				**Hatalar**			
No	**g**	**gse**	**s**	**sse**	**g**	**gse**	**s**	**sse**	**g**	**gse**	**s**	**sse**
5017_A	0,171	0,023	0,329	0,068	0,193	0,018	0,389	0,052	0,155	0,022	0,561	0,024
5017_B	0,030	0,002	0,641	0,010	0,020	0,008	0,770	0,008	0,000	0,003	0,642	0,011
5036_A	0,274	0,013	0,167	0,029	0,204	0,031	0,367	0,031	0,284	0,013	0,196	0,029
5036_B	0,196	0,012	0,611	0,029	0,153	0,018	0,660	0,020	0,208	0,010	0,645	0,023
5036_C	0,121	0,009	0,811	0,022	0,118	0,014	0,841	0,015	0,115	0,012	0,838	0,013
5058_A	0,544	0,018	0,067	0,014	0,481	0,023	0,008	0,019	0,521	0,020	0,047	0,016
5058_B	0,298	0,019	0,235	0,019	0,371	0,017	0,001	0,045	0,416	0,013	0,138	0,024
5059	0,033	0,052	0,595	0,053	0,000	0,515	0,562	0,515	0,001	0,074	0,563	0,075
5057	0,672	0,012	0,061	0,017	0,619	0,017	0,138	0,014	0,590	0,018	0,109	0,015
Ort.	0,260	0,018	0,391	0,029	0,240	0,073	0,415	0,080	0,254	0,021	0,415	0,026

Tablo 32'de görüldüğü gibi BTM analizi sonucunda maddelerin *g* parametreleri ortalamaları matematiksel yeterlikler için 0,26; kazanımlar için 0,24 ve hatalar için 0,254; *s* parametrelerinin ortalamaları ise matematiksel yeterlikler için 0,391; kazanımlar için 0,415 ve hatalar için 0,415'dir. Her iki parametrenin matematiksel yeterlikler, kazanımlar ve hatalar için g ve s parametre değerleri testtin model uyumunun yeterli olduğunu göstermektedir.

3.4.5 Üstbilişsel İzleme Ölçümleri Uygulaması

Projenin 3 izleme sürecinin ardından öğrencilere ve öğretmenlere yönelik üstbilişsel ölçümler alma aşaması gerçekleştirilmiştir. Bu süreçte izleme testini grubunda yer alan Deney-1 ve Deney-2 öğrencilerinin tamamına yönelik ölçüm stratejileri geliştirilmiştir. Bu ölçümler üç temel bölümde gerçekleştirilmiştir. İlk aşama temel üstbilişsel izleme becerisine yönelik ölçümlerin alındığı her soru için öğrencilerin soruyu çözüp çözemeyeceklerine ve sorudan sonrada cevabının eminliğine dair ölçümler alınmıştır. Aşağıda bu ölçümlerin nasıl alındığını gösteren şekiller bulunmaktadır.

Bu soruyu "çözebileceğimi **düşünüyorum**" ◯
"çözebileceğimi **düşünmüyorum**" ◯ → *Bu seçeneği işaretledi yseniz bile, **yine de** soruyu size en doğru gelen yanıtla cevaplayınız!*

A) $4a \times 2a$ **B)** $\frac{2a \times 2a}{2}$ **C)** $2a \times a$ **D)** $\frac{4a \times 2a}{2}$

Şekil 11.Üst Bilişsel İzleme Çoktan Seçmeli Soru

"Yanıtınızın doğru olduğundan ne kadar eminsiniz?" *Lütfen bir sayıyı yuvarlak içine alarak belirtiniz.*

Hiç emin değilim	Emin değilim	Biraz eminim	Eminim	Kesinlikle eminim
1	2	3	4	5

Şekil 12. Üst Bilişsel İzleme Eminlik Düzeyi Ölçümü

SORU 2. BU SORUNUN CEVABINI KİTAPÇIK ÜZERİNE YAZINIZ VE OPTİK FORMDA "Y" KUTUSUNU DOLDURUNUZ P5017_B

Tangram parçalarından oluşan şekilde AEFG karesinin çevresinin uzunluğu 10 cm'dir. Buna göre aşağıdaki parçaların alanlarını işlemlerinizi göstererek hesaplayınız.

SORU	Bu soruyu çözebileceğimi		CEVAP	Yanıtınızın doğru olduğundan ne kadar eminsiniz? (aşağıdaki sayılardan uygun olanı işaretleyiniz)				
	"Düşünüyorum"	"Düşünmüyorum"						
Siyah üçgen	○	○		1	2	3	4	5
Mavi paralelkenar	○	○		1	2	3	4	5
Yeşil kare	○	○		1	2	3	4	5

Şekil 13. Üst Bilişsel İzleme Doğru Yanlış Soru

Üstbilişsel izlemeye yönelik ikinci tür ölçümler ise öğrencilerin test geneline ilişkin kendi ve sınıf arkadaşlarının performanslarına yönelik tahminlerini içermektedir. Bu ölçümler aşağıdaki sorular yardımı ile alınmıştır.

Test sorularına başlamadan önce, bir tahminle başlayalım.
Sizce, testte bulunan 9 adet matematik sorusundan kaç tanesini doğru olarak yanıtlarsınız? *Lütfen belirtiniz ……..*

Lütfen aşağıdaki soruları cevaplayınız

- Sizce, sorulardan kaç tanesine doğru yanıt vermişsinizdir? ..... adet
- Sizce sınıftaki arkadaşlarınız ortalama olarak kaç soruya doğru yanıt vermiştir? ..... adet

Şekil 14. Öğrenci Performans Tahmin Formu

Sorularda da görüldüğü gibi öğrencilerin kendi performanslarına ilişkin ön tahminde bulunması istenmiştir. Test sonunda ise hem kendi hem de sınıf arkadaşlarına ilişkin performans kestirimleri alınmıştır.

Üstbilişsel izlemenin öğretmen ayağına ilişkin ölçümler ise testten sonra öğretmenlere gönderilen formlar ile elde edilmiştir. Bu formlarda her bir soru için ve genel sınıf performansı için öğretmen görüşleri istemiştir.

00 değişik desen

	Bu soruyu sınıfınızdan kaç öğrenci doğru cevaplamıştır				
Şubeler	A	C			
Öğrenci sayısı	15	8			

Şekil 1. Öğrenci sayısı

1600 değişik desen

	Bu soruyu sınıfınızdan kaç öğrenci doğru cevaplamıştır				
Şubeler	A	C			
Öğrenci sayısı	%50	%25			

Şekil 2. Doğru Cevaplanma Yüzdesi

Şekil 15. Üst Bilişsel İzleme Öğretmen Soru Formu

Yukarıdaki şekillerde görüldüğü öğretmenler dersine girdiği şubeler için her soruda öğrenci performansını tahmin etmişlerdir. Bununla birlikte testteki ortalama performansa ilişkinde tahminde bulunmuşlardır.

Lütfen Aşağıdaki bilgileri test sorularını inceledikten sonra doldurunuz.

Dersine girdiğiniz şubeler					
Bu testteki toplam 9 soruda sınıf ortalamasının kaç olacağını düşünüyorsunuz?					
Sizce şubenizde kaç öğrencinin 7 veya daha fazla doğru cevabı olacaktır?					
Sizce şubenizde kaç öğrencinin 3 veya daha az doğru cevabı olacaktır?					

Şekil 16. Üst Bilişsel İzleme Öğretmen Test Formu

Öğretmenler aynı zamanda yukarıdaki şekilde görüldüğü gibi sınıflarındaki öğrencilerin hem ortalama hem de yüksek ve düşük performansa ilişkin tahminlerine yer vermişlerdir.

Bu ölçümlerin tamamı önce pilot grupta yer alan 230 öğrenciye ve daha sonra da İzleme 4'e katılan 2834 öğrenciye uygulanmıştır.

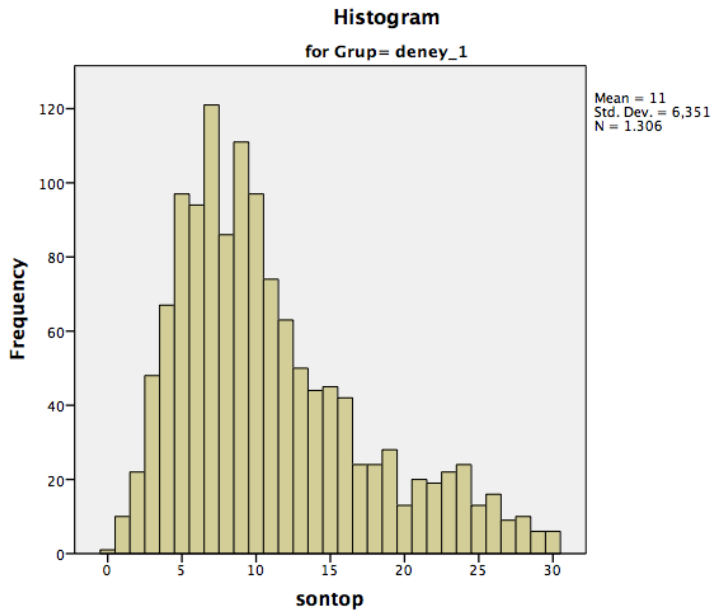
3.4.6 Son Test Uygulaması

Son test uygulamasına 4069 öğrenci katılmıştır. Ancak bu öğrencilerden ön test uygulamasına girmeyenler araştırma bulgularında yer almayacağı son test analizler 3705 öğrenci üzerinden gerçekleştirilmiştir. Son test betimsel istatistikleri Tablo 33'de verilmiştir.

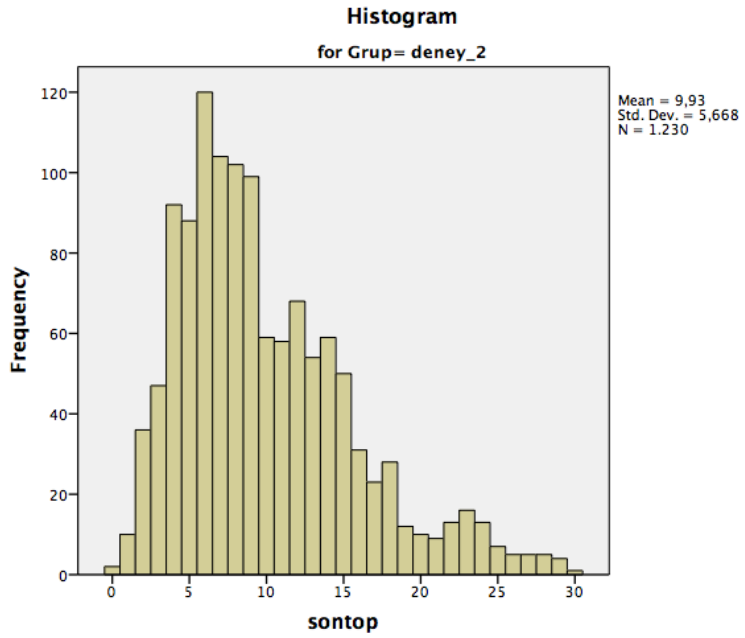
Tablo 33. Deney ve Kontrol Grupları Betimsel İstatistikleri

	DENEY-1	DENEY-2	KONTROL
N	1306	1230	1169
ORTALAMA	11,01	9,93	9,24
MEDYAN	9,00	9	8,00
VARYANS	40,34	32,12	30,35
STANDART S	6,35	5,69	5,51
MİNİMUM	0	0	0
MAKSİMUM	30	30	30
RANJ	30	30	30
ÇARPIKLIK	0,945	0,989	0,581

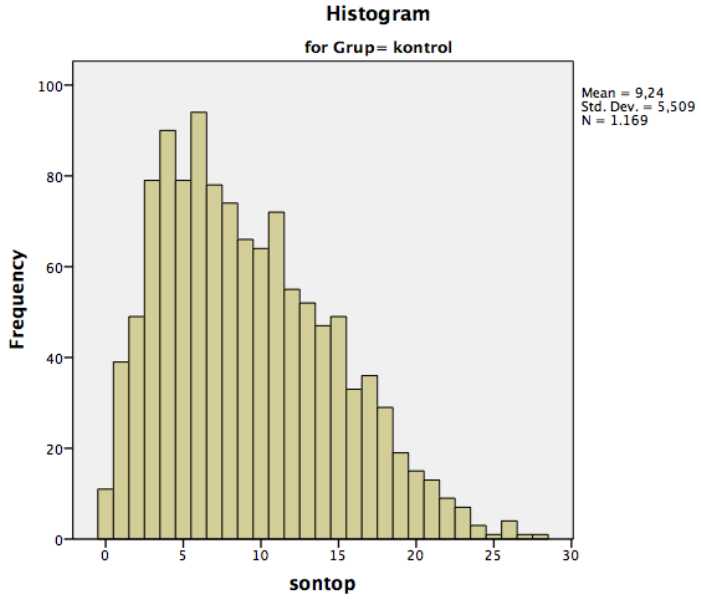
Aşağıda Deney-1, Deney-2 ve Kontrol gruplarının son test puanlarına ait histogram grafikleri aşağıda verilmiştir.



Şekil 17. Deney-1 Son-Test puanlarına Ait Histogram Grafiği



Şekil 18. Deney-2 Son-Test puanlarına Ait Histogram Grafiği



Şekil 19. Kontrol Son-Test puanlarına Ait Histogram Grafiği

3.4.7 Son-test Psikometrik Özellikleri

Son test uygulamasında yer alan soruların psikometrik özellikleri Klasik Test Teorisi(KTK) ve Madde Tepki Kuramı(MTK) kullanılarak incelenmiştir. Öncelikle KTK için madde güçlüğü ve madde test korelasyonları hesaplanmıştır. Tablo 34'de KTK analizi sonuçları verilmiştir.

Tablo 34. Son-Test KTK Madde Analizleri

No	P_j	r_{jx}	r_{pbis}	r_{bis}	No	P_j	r_{jx}	r_{pbis}	r_{bis}
1	0,63	0,47	0,39	0,32	**16**	0,63	0,58	0,46	0,39
2	0,56	0,59	0,47	0,4	**17**	0,49	0,72	0,57	0,50
3	0,41	0,51	0,43	0,36	**18**	0,38	0,66	0,55	0,49
4	0,5	0,76	0,59	0,53	**19**	0,21	0,53	0,6	0,55
5	0,36	0,51	0,47	0,4	**20**	0,41	0,39	0,36	0,29
6	0,36	0,52	0,45	0,38	**21**	0,28	0,34	0,36	0,29
7	0,14	0,39	0,56	0,52	**22**	0,24	0,35	0,39	0,33
8	0,11	0,27	0,46	0,42	**23**	0,38	0,53	0,46	0,39
9	0,16	0,21	0,31	0,25	**24**	0,21	0,23	0,3	0,24
10	0,20	0,31	0,38	0,32	**25**	0,28	0,48	0,47	0,41
11	0,29	0,30	0,30	0,22	**26**	0,31	0,31	0,33	0,25
12	0,22	0,50	0,54	0,49	**27**	0,23	0,57	0,62	0,57
13	0,28	0,28	0,30	0,23	**28**	0,41	0,62	0,53	0,47
14	0,5	0,51	0,41	0,34	**29**	0,36	0,41	0,39	0,32
15	0,48	0,43	0,36	0,28	**30**	0,06	0,17	0,48	0,44
Ortalama						0,336	0,448	0,443	0,380

Tablo 34'de maddelerin güçlük değerleri (P_j), ayırt edicilik parametreleri (r_{jx}) ve nokta çift serili (r_{pbis}) ve çift serili (r_{bis}) korelasyon katsayıları verilmiştir. Görüldüğü gibi son-test maddelerinin gerek ayırt edicilik değerleri gerekse madde test korelasyonları test maddelerinin yeterli ayırt ediciliğe sahip olduğunu göstermektedir. Son-testte ilişkin güvenirlik analizlerinde KR-20 katsayısı incelenmiştir. Tüm test için KR-20 = 0,86 bulunmuştur. Okul düzeyinde yapılan analizlerde testtin iç tutarlılık katsayısı 0,78 ile 0,89 arasında değişmektedir. Bu durum testtin yeterli güvenirliğe sahip olduğunu göstermektedir.

Bunula birlikte test güvenirliği MTK ile de incelenmiştir. MTK'da marjinal güvenirlik Rasch Model'de (1 Parametreli Model) 0,96 ve 2 Parametreli Model'de 0,72 tahminlenmiştir. Analizlerde veri setinin 1-0 formatına dönüştürüldüğü göz önüne alındığında Rasch Modelin veri ile daha uyumlu olduğu anlaşılmaktadır. Son test uygulamasına ilişkin MTK madde parametreleri (EK 22), Madde Karakteristik eğrileri (EK 23) Test karakteristik Eğrisi (EK 24) ve Test Bilgi Fonksiyonu (EK 25) eklerde verilmiştir.

4 BULGULAR

Bu bölümde araştırma sürecinde elde edilen bulgular verilmiştir. Bulgular üç ana başlıkta incelenmiştir. Öncelikle araştırmanın temel problem cümlesi doğrultusunda deneysel işlem sürecinin etkililiğine ilişkin bulgular daha sonra ise BTM analizleri ve öğrenci profilleri ile ilgili bulgular ve son olarak da üst bilişsel izleme, gibi konularda ulaşılan bulgular raporlanmıştır.

4.1 Deneysel Sürece İlişkin Bulgular

Araştırmanın deneysel süreci ile ilgili bulgulara ulaşmak için ön-test grup denklikleri, son-test farkları, tekrarlı ölçümler için ANOVA ve çok değişkenli faktöriyel analizler gerçekleştirilmiştir.

4.1.1 Ön-Son Testler için Varyans Analizleri

Araştırmada grupların ön-test puan ortalamaları arasındaki farkların anlamlı olup olmadığı tek yönlü varyans analizi (ANOVA) ile incelenmiştir.

Tablo 35. Ön-Test Toplam Puanlarına ait Betimsel İstatistikleri

Grup	N	$\bar{X}$	s	Çarpıklık
Deney_1	1306	8,04	3,729	,711
Deney_2	1230	8,07	3,802	,763
Kontrol	1169	8,00	3,678	,610
Toplam	3705	8,04	3,737	,699

Tablo 35'de görüldüğü gibi ön-test toplam puanlarının bütün gruplar için çarpıklık katsayılarının -1,1 değeri arasındadır. Bu durum bütün grupların normal dağılımdan aşırı sapma göstermediğini anlamına gelmektedir (Leech, Barrett, & Morgan, 2007). Bununla birlikte Levene istatistiği (p>0,05) grupların varyanslarının homojen olduğunu göstermektedir. Bu durum veri setinin ANOVA analizinin varsayımlarını karşıladığına işaret etmektedir.

Tablo 36. Grupların Ön-Test Puanlarına ait ANOVA Analizi

	Kareler Toplamı	sd	Kareler ortalaması	F	P
Gruplar arası	2,739	2	1,37	0,098	0,907
Gruplar içi	51712,11	3702	13,969		
Toplam	51714,85	3704			

ANOVA sonuçları incelendiğinde gruplara ait ön-test ortalamaları arasında istatistiksel olarak anlamı bir fark yoktur (F (2,3702) = 0,098, p = 0,907). Bu durum deney ve kontrol gruplarının ön-test puanlarının eşit olduğunu göstermektedir.

Araştırmada grupların son-test puan ortalamaları arasındaki farkların anlamlı olup olmadığı tek yönlü varyans analizi (ANOVA) ile incelenmiştir.

Tablo 37. Son-Test Toplam Puanlarına ait Betimsel İstatistikleri

Grup	N	$\bar{X}$	s	Çarpıklık
Deney_1	1306	11	6,351	,945
Deney_2	1230	9,93	5,668	,989
Kontrol	1169	9,24	5,509	,581
Toplam	3705	10,09	5,914	,894

Tablo 37'de görüldüğü gibi son-test toplam puanlarının bütün gruplar için çarpıklık katsayılarının -1,1 değeri arasındadır. Bu durum bütün grupların normal dağılımdan aşırı sapma göstermediği anlamına gelmektedir (Leech et al., 2007). Bununla birlikte Levene istatistiği (p > 0,05) grupların varyanslarının homojen olduğunu göstermektedir. Bu durum veri setinin ANOVA analizinin varsayımlarını karşıladığına işaret etmektedir.

Tablo 38. Grupların Son-Test Puanlarına ait ANOVA Analizi

	Kareler Toplamı	sd	Kareler ortalaması	F	P
Gruplar arası	1978,212	2	989,106	28,703	0,000
Gruplar içi	127571,9	3702	34,46		
Toplam	129550,1	3704			

ANOVA sonuçları incelendiğinde gruplara ait son-test ortalamaları arasında istatistiksel olarak anlamı bir fark bulunmaktadır (F (2,3702) = 28,703, p = 0,000). Gruplar arasındaki farklar Post Hoc LSD testi ile incelenmiştir. LSD testi sonuçlarına göre Deney-1 grubu son test ortalaması (M=11, s=6,35) istatistiksel olarak Deney-2 (M=9,93, s=5,67) ve Kontrol grubu son test ortalamalarından (M=9,24, s=5,51) daha yüksektir. Aynı şekilde Deney-2 grubu ile kontrol grubu son test ortalamaları arasında da Deney-2 grubu lehine istatistiksel olarak anlamlı bir fark görülmektedir.

4.1.2 Ön-Son Test Uygulamaları Arasında Tekrarlı Ölçümler Analizi

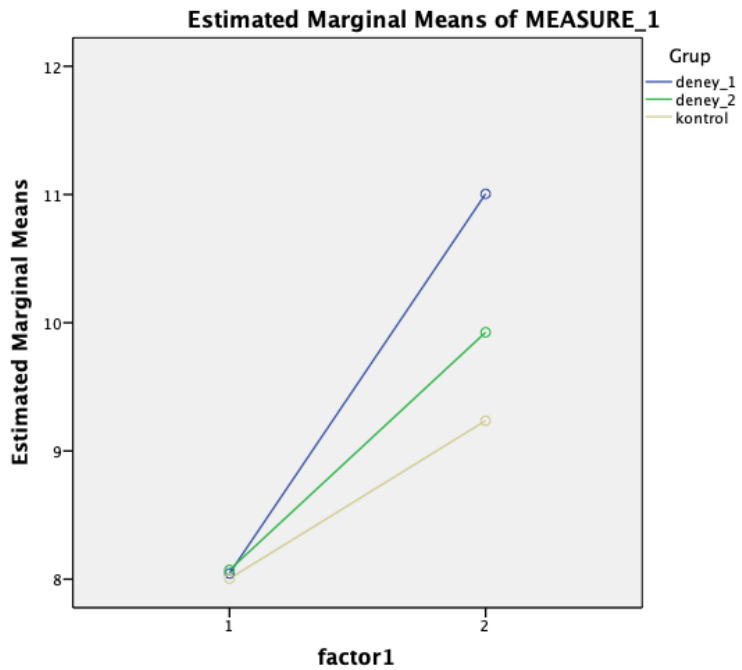
Deneysel sürecin değerlendirilmesi tekrarlı ölçümler için faktöriyel Varyans Analizi (ANOVA) ile incelenmiştir. Bu analiz ön-son test kullanılan deneysel desenlerde grupları ve testleri ayrı bir faktör olarak inceleyerek deneyin başarısı konusunda en gerçekçi istatistiksel sonuçları vermektedir (Büyüköztürk, 2010; Leech et al., 2007). Araştırmada 2 x 3 iki faktörlü desen kullanılmıştır. İlk faktör zaman (ön-test, son-test) ikinci faktör ise gruptur (Deney-1, Deney-2, Kontrol).

Analiz sonuçlarına göre dağılımların normallik ve küresellik varsayımı karşılanmıştır. Ön test - son test ANOVA sonuçları Tablo 39'da verilmiştir.

Tablo 39. Ön-Son Test Puanlarının ANOVA Sonuçları

Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Denekler arası	743015,874	3704			
Grup	1025,013	2	512,506	13,885	0,007
Hata	136646,2	3702	36,911		
Denekler içi	51101,624	3705			
Ölçüm	7507,806	1	7507,806	651,859	0,000
Grup*Ölçüm	955,938	2	477,969	41,499	0,000
Hata	42637,88	3702	11,518		
Toplam	794117,498	7409			

Ön-son test puan ortalamalarının gruplara göre farklılaşma düzeyleri incelendiğinde öğrencilerin bulundukları grup ile ön-son test puanları arasındaki farklılaşmanın etkili olduğu görülmektedir ($F(2, 3702) = 41,499$, $p = 0,00$). Hangi gruplar arası farkların anlamlı olduğunu belirlemek için Post Hoc LSD testi uygulanmıştır. Analiz sonuçlarına göre Deney-1, Deney-2 (p= 0,00), Deney-1, Kontrol (p= 0,00) ve Deney-2, Kontrol (p=0,03) arasındaki farklar ilk yazılan grup lehine anlamlıdır. Bir diğer değişle deneysel işlem süreci sonucunda Deney-1 grubunun Deney-2'den ve Deney-2 grubunun Kontrol grubundan ön-son test ortalamaları arasındaki farklar daha yüksektir. Marjinal ortalamalardaki değişim Şekil 20'de verilmiştir.



Şekil 20. Gruplara Ait Ön-Son Test Profil Grafiği

Şekil 20 incelendiğinde ön-son test arasında grupların ortalamalarındaki değişim görülmektedir. Her grup da ön-son test arasındaki fark istatistiksel olarak anlamlı bulunmuştur. Bununla birlikte ortak etki göz önüne alındığında Deney-1 grubunun her iki gruptan, Deney-2

grubunun ise Kontrol grubundan istatistiksel olarak daha yüksek düzeyde son test puan ortalamalarının arttığı görülmektedir.

4.1.3 Deneysel Sürece ilişkin Kovaryans Analizleri

Ön-son test kullanılan deneysel desenlerde sıkça kullanılan bir diğer analiz kovaryans (ANCOVA) analizidir. Ön-test puanlarının kovaryet olarak alındığı bu analizde özellikle gruplar arasındaki farklar ön-test puanları kontrol ederek hesaplanmaktadır. Araştırma deseninde ön-test puanlarının gruplar arasında farklılaşmadığı görülmekle birlikte örneklem büyüklüğü göz önüne alınarak deneysel işlemin tekrar kovaryans analizi ile de incelenmesine gerek görülmüştür. Tablo 40'da Son-test puanlarına ilişkin ANCOVA sonuçları verilmiştir.

Tablo 40. Son Test Düzeltilmiş Puanlarının Gruba Göre ANCOVA Sonuçları

	Kareler Toplamı	sd	Kareler ortalaması	F	P
Ön-test	42724,79	1	42724,79	1863,639	0,00
Grup	1917,668	2	958,834	41,824	0,00
Hata	84847,14	3701	22,925		
Toplam	506619	3705			

Analiz sonuçlarına göre gruplar arasındaki farkların anlamlı olduğu görülmüştür ($F(2,3701) = 41,824$, $p=0,00$). Hangi gruplar arasındaki farkın anlamlı olduğu Post Hoc LSD testi ile incelenmiştir. Analiz sonuçlarına göre Deney-1 ($M=11,001$) ile Deney-2 ($M=9,897$) ve Kontrol ($M=9,269$) grupları arasında Deney-1 lehine anlamlı bir fark olduğu görülmüştür. Aynı zamanda Deney-2 grubu ortalaması da Kontrol grubundan istatistiksel olarak anlamlı düzeyde büyüktür.

4.1.4 Deneysel Sürecin Kalıcılığına İlişkin Tekrarlı Ölçümler Varyans Analizi

Deneysel süreç tamamlandıktan sonra elde edilen bulgularının kalıcılığının incelenebilmesi amacı ile örnekleme kalıcılık testi uygulanmıştır. Bu test son testten yaklaşık 8 ay sonra gerçekleştirilmiştir. Kalıcılık testi puanlarına ilişkin betimsel istatistikler Tablo 41'de verilmiştir.

Tablo 41. Kalıcılık Testi Örneklemine ait Ön-Son ve Kalıcılık Test Puanları Betimsel İstatistikleri

		Ön-Test		Son-Test		Kalıcılık	
Grup	N	ort	ss	ort	ss	ort	ss
Deney-1	606	8,18	3,76	11,15	6,39	11,42	6,24
Deney-2	560	7,99	3,79	9,65	5,37	9,65	5,48
Kontrol	507	8,11	3,61	9,53	5,66	9,20	5,15
Toplam	1673	8,10	3,72	10,16	5,89	10,15	5,75

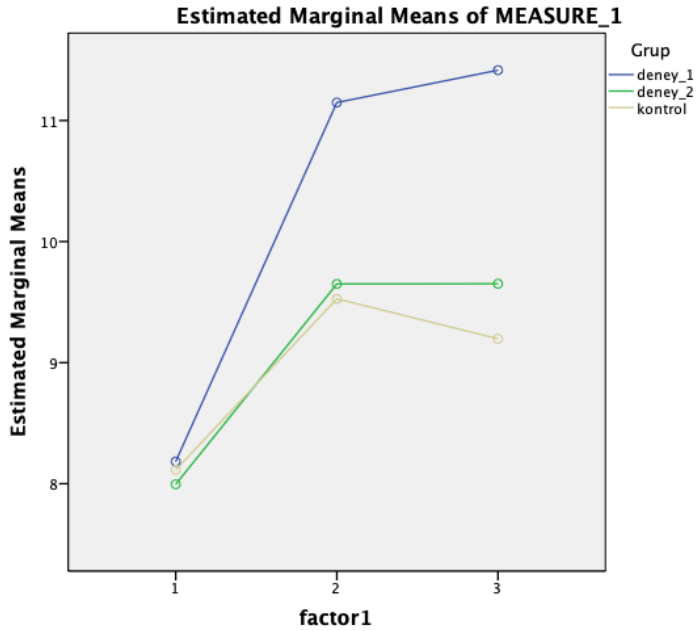
Tablo 41'de görüldüğü gibi Kalıcılık testi örneklemi araştırma örneklemine göre daha küçüktür. Bunun temel sebebi öğrencilerin 6. Sınıftan 7. Sınıf düzeyine geçerken okul ve şube düzeyinde değişiklikler olmasıdır. Araştırmada uygulanan 3 testin tamamını almış olan örneklem üzerinden kalıcılık analizleri gerçekleştirilmiştir. Araştırmanın deneysel sürecinde ulaşılan düzeylerin kalıcılığı tekrarlı ölçümler için faktöriyel Varyans Analizi(ANOVA) ile incelenmiştir. Araştırmada 3 x 3 iki faktörlü desen kullanılmıştır. İlk faktör zaman (ön-test, son-test, kalıcılık testi) ikinci faktör ise gruptur (Deney-1, Deney-2, Kontrol).

Analiz sonuçlarına göre dağılımların normallik ve küresellik varsayımları karşılanmıştır. Ön-son-kalıcılık testleri ANOVA sonuçları aşağıda verilmiştir.

Tablo 42. Ön-Son-Kalıcılık Test Puanlarının ANOVA Sonuçları

Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Denekler arası	66131,327	1672			
Grup	1752,693	2	876,347	22,733	0,000
Hata	64378,634	1670	39,550		
Denekler içi	38250,056	1673			
Ölçüm	3297,972	1	3297,972	156,380	0,000
Grup*Ölçüm	704,137	2	352,068	16,694	0,000
Hata	34247,947	1670	21,089		
Toplam	104381,383	3345			

Ön-son-kalıcılık test puan ortalamalarının gruplara göre farklılaşma düzeyleri incelendiğinde öğrencilerin bulundukları grup ile ön-son test puanları arasındaki farklılaşmanın etkili olduğu görülmektedir ($F(2, 1670) = 16,694$, $p = 0,00$). Hangi gruplar arası farkların anlamlı olduğunu belirlemek için Post Hoc LSD testi uygulanmıştır. Analiz sonuçlarına göre Deney-1, Deney-2 ($p= 0,00$), Deney-1, Kontrol ($p= 0,00$) ve Deney-2, Kontrol ($p=0,03$) arasındaki farklar ilk yazılan grup lehine anlamlıdır. Kalıcılık testi ortalamaları ise bütün gruplar için son testle arasında istatistiksel olarak anlamlı bir fark bulunmamıştır. Bir diğer değişle deneysel işlem süreci sonucunda yapılan Kalıcılık testi ortalamaları bütün gruplarda son testle aynı düzeyde farklılaşmakla birlikte, her grup için son ve kalıcılık testi arasında anlamlı bir fark gözlenmemiştir. Aşağıdaki marjinal ortalamalardaki değişim Şekil 21'de verilmiştir.



Şekil 21. Gruplara Ait Ön-Son-Kalıcılık Testi Profil Grafiği

Şekil 21 incelendiğinde grupların son ve kalıcılık testi ortalamaları arasında anlamlı bir fark olmadığı görülmektedir. Bununla birlikte Deney-1 grubunun son teste göre kalıcılık testinde başarı düzeylerinde bir artış olduğu gözlenmektedir. Aynı zamanda kontrol grubunda da belirgin düzeyde son test ile kalıcılık testi arasında bir düşüş belirlenmiştir. Bu farkların anlamlılığına ilişkin yapılan tekrarlı ölçümler t testi sonuçlarına göre bütün gruplarda son-kalıcılık testi ortalamaları arasında istatistiksel olarak anlamlı bir fark olmadığı görülmüştür (Deney-1 $t(605)= -0,805$, $p=0,421$; Deney-2 $t(559)=-0,06$, $p=0,996$; Kontrol $t(506)=1,007$, $p=0,314$).

4.2 Öğrenci Yetenek Düzeyleri ve BTM Analizlerine İlişkin Bulgular

Araştırmanın temel hipotezi PISA ve TIMSS başarısını arttırmaya yönelik bir modelin etkililiği üzerine kurulduğu için analizlerin PISA ve TIMSS standartlarında gerçekleştirilmesi hipotez ile ilgili bulguları daha genellenebilir yapacağı düşünülmüştür. Bu nedenle deney ve kontrol gruplarındaki öğrencilerin puanları sadece KTK ile değil aynı zamanda PISA uygulamasındaki gibi (OECD, 2009b) MTK modelleri ile de hesaplanmıştır. Aynı zamanda araştırma test puanları için PISA matematik yeterlik düzeyleri belirlenerek araştırma hipotezi bu düzeyler üzerinden de test edilmiştir.

Aynı zamanda araştırma geri bildirim sisteminin temelini oluşturan BTM analizleri kullanılarak öğrencilerin yetenek sınıflarında ve olasılıklarındaki değişimler incelenmiştir. Bununla birlikte deneysel süreçteki değişimin öğrenci yeterlik düzeylerine göre değişimleri de analiz edilmiştir.

4.2.1 MTK Analizleri ve Öğrenci Yetenek Düzeylerindeki Değişim

Araştırmada ön-son test için öğrenci yetenek düzeyleri PISA analizlerinde kullanılan Rash model ile hesaplanmıştır. Daha sonra PISA yeterlik düzeylerinin belirlendiği standart puanlar elde edilmiştir. Aşağıda Tablo 43'de 2015 PISA uygulaması matematik yeterlikleri için Türkiye örneklemi dağılımları ve araştırma sonucunda elde edilen ön test dağılımları yüzdelik olarak verilmiştir.

Tablo 43. 2015 PISA ve Proje Matematiksel Yeterlik Seviyeleri

Seviyeler	PISA Türkiye	Proje
1.Seviyenin altı	15,5%	13,8%
1.Seviye	26,5%	25,6%
2. Seviye	25,5%	27,0%
3. Seviye	16,5%	19,3%
4. Seviye	10,1%	10,1%
5. Seviye	4,7%	3,4%
6. Seviye	1,2%	0,8%

Tablo 43'de görüldüğü gibi PISA 2015 uygulaması ile araştırmada elde edilen yetenek düzeyleri birbirlerine çok yakın değerler almaktadır. Bu durum araştırmada kullanılan testlerin PISA ile yüksek düzeyde benzer olduğunun bir göstergesidir. Ön ve son testlerde gruplara göre farklılaşmalar için yapılan tek yönlü varyans analizi sonuçları aşağıda Tablo 44'de verilmiştir.

Tablo 44. Ön-Son Test Yetenek Puanları için Gruplara Göre ANOVA Sonuçları

		Kareler Toplamı	sd	Kareler ortalaması	F	P
Ön-test	Gruplar arası	17764,59	2	8882,293	1,244	0,288
	Gruplar içi	26431981	3702	7139,919		
	Toplam	26449746	3704			
Son-test	Gruplar arası	510884,3	2	255442,2	30,348	0,00
	Gruplar içi	31160271	3702	8417,145		
	Toplam	31671155	3704			

ANOVA sonuçları incelendiğinde gruplara ait ön-test ortalamaları arasında istatistiksel olarak anlamı bir fark bulunmaktadır ($F(2,3702) = 1,244$, $p = 0,288$). Gruplara ait son-test ortalamaları incelendiğinde ise aralarında istatistiksel olarak anlamı bir fark bulunduğu görülmektedir ($F(2,3702) = 30,348$, $p = 0,00$). Gruplar arasındaki farklar Post Hoc LSD testi ile incelenmiştir. LSD testi sonuçlarına göre Deney-1 grubu son test ortalaması (M=464,27, s=97,38) istatistiksel olarak Deney-2 (M=448,52, s=88,89) ve Kontrol grubu son test ortalamalarından (M=435,59, s=88,16) daha yüksektir. Aynı şekilde Deney-2 grubu ile kontrol grubu son test

ortalamaları arasında da Deney-2 grubu lehine istatistiksel olarak anlamlı bir fark görülmektedir.

Araştırma deseninde ön-test puanlarının gruplar arasında farklılaşmadığı görülmekle birlikte örneklem büyüklüğü göz önüne alınarak deneysel işlemin tekrar kovaryans analizi ile de incelenmesine gerek görülmüştür. Tablo 45'de Son-test puanlarına ilişkin ANCOVA sonuçları verilmiştir.

Tablo 45. Son Test Düzeltilmiş Puanlarının Gruba Göre ANCOVA Sonuçları

	Kareler Toplamı	sd	Kareler ortalaması	F	P
Ön-test	11852551,7	1	11852551,7	2271,95	0,00
Grup	392391,4	2	196195,593	37,60	0,00
Hata	19307719,4	3701	5216,893		
Toplam	781893246	3705			

Analiz sonuçlarına göre gruplar arasındaki farkların anlamlı olduğu görülmüştür ($F(2,3701) = 37,60$, $p=0,00$). Hangi gruplar arasındaki farkın anlamlı olduğu Post Hoc LSD testi ile incelenmiştir. Analiz sonuçlarına göre düzeltilmiş Deney-1 ($M=462,365$) ile Deney-2 ($M=449,007$) ve Kontrol ($M=437,193$) grup ortalamaları arasında Deney-1 lehine anlamlı bir fark olduğu görülmüştür. Aynı zamanda Deney-2 grubu ortalaması da Kontrol grubundan istatistiksel olarak anlamlı düzeyde büyüktür.

4.2.1.1 Matematiksel Yeterlik Düzeylerindeki Değişimlere İlişkin Bulgular

Araştırmada ön ve son test puanlarına göre MTK analizleri ile belirlenen Öğrenci yetenek puanları daha sonra PISA standart puanlarına dönüştürülmüş ve bu şekilde PISA matematiksel yeterliklerin dağılımlarına ulaşılmıştır. Tablo 46'da ön test için gruplara göre matematiksel yeterliklere düşen öğrenci sayıları verilmiştir.

Tablo 46. Ön Test Puanlarına Göre Öğrencilerin Matematiksel Yeterlik Dağılımları

Seviyeler	Deney-1	Deney-2	Kontrol	Toplam
1. Seviye altı	173	167	171	511
1. Seviye	327	313	309	949
2. Seviye	343	354	303	1000
3. Seviye	269	230	216	715
4. Seviye	137	109	127	373
5. Seviye	47	41	38	126
6. Seviye	10	16	5	31
Toplam	1306	1230	1169	3705

Grup dağılımlarında istatistiksel olarak bir fark olup olmadığı ki kare testi ile incelenmiş ve test sonuçlarına göre üç grubun dağılımın istatistiksel olarak farklı olmadığı görülmüştür (χ^2(12,3705)=13,966, p=0,303). Tablo 47'de son test için gruplara göre matematiksel yeterliklere düşen öğrenci sayıları verilmiştir.

Tablo 47. Son Test Puanlarına Göre Öğrencilerin Matematiksel Yeterlik Dağılımları

Seviyeler	Deney-1	Deney-2	Kontrol	Toplam
1. Seviye altı	153	186	253	592
1. Seviye	321	321	260	902
2. Seviye	366	321	287	974
3. Seviye	224	245	233	702
4. Seviye	115	86	110	311
5. Seviye	73	48	23	144
6. Seviye	54	23	3	80
Toplam	1306	1230	1169	3705

Son test yeterlik grup dağılımlarında istatistiksel olarak bir fark olup olmadığı ki kare testi ile incelenmiş ve test sonuçlarına göre üç grubun dağılımın istatistiksel olarak anlamlı farklı olduğu görülmüştür (χ^2(12,3705)=117,972, p=0,000).

Tablo 48 Deney-1 grubundaki öğrencilerin yetenek düzeylerinde ön ve son test arasındaki değişimlerini göstermektedir.

Tablo 48. Deney-1 Grubu Ön-Son Test Yeterlik Düzeyleri Çapraz Tablosu

		ÖN-TEST						
		1. Seviye altı	1. Seviye	2. Seviye	3. Seviye	4. Seviye	5. Seviye	6. Seviye
SON-TEST	1. Seviye altı	40	66	27	15	2	3	0
	1. Seviye	76	113	84	37	8	3	0
	2. Seviye	46	102	119	72	26	1	0
	3. Seviye	9	38	68	68	32	9	0
	4. Seviye	1	8	26	45	22	11	2
	5. Seviye	1	0	10	25	26	8	3
	6. Seviye	0	0	9	7	21	12	5

Tablo 48 incelendiğinde Deney-1 grubu öğrencilerin %21'inin (n=401) son-testte ön-teste göre daha düşük bir seviyeye indiği, %41'inin (n=530) daha yüksek seviyeye çıktığı ve %28'inin (n=375) aynı seviyede kaldığı görülmektedir. Tablo 49 Deney-2 grubundaki öğrencilerin yetenek düzeylerinde ön ve son test arasındaki değişimlerini göstermektedir.

Tablo 49. Deney-2 Grubu Ön-Son Test Yeterlik Düzeyleri Çapraz Tablosu

		ÖN-TEST						
		1. Seviye altı	1. Seviye	2. Seviye	3. Seviye	4. Seviye	5. Seviye	6. Seviye
SON-TEST	1. Seviye altı	63	80	33	8	2	0	0
	1. Seviye	76	125	80	34	4	2	0
	2. Seviye	24	77	122	68	25	3	2
	3. Seviye	4	26	93	73	37	11	1
	4. Seviye	0	4	18	32	25	5	2
	5. Seviye	0	0	4	8	12	16	8
	6. Seviye	0	1	4	7	4	4	3

Tablo 49 incelendiğinde Deney-2 grubu öğrencilerin %33'ünün (*n*=405) son-testte ön-teste göre daha düşük bir seviyeye indiği, %45'inin (*n*=427) daha yüksek seviyeye çıktığı ve %32'sinin (*n*=398) aynı seviyede kaldığı görülmektedir. Tablo 50 Kontrol grubundaki öğrencilerin yetenek düzeylerinde ön ve son test arasındaki değişimlerini göstermektedir.

Tablo 50. Kontrol Grubu Ön-Son Test Yeterlik Düzeyleri Çapraz Tablosu

		ÖN-TEST						
		1. Seviye altı	1. Seviye	2. Seviye	3. Seviye	4. Seviye	5. Seviye	6. Seviye
SON-TEST	1. Seviye altı	82	103	52	13	3	0	0
	1. Seviye	56	97	70	27	10	0	0
	2. Seviye	30	79	94	63	20	1	0
	3. Seviye	2	26	69	80	47	9	0
	4. Seviye	1	4	16	31	38	19	1
	5. Seviye	0	0	2	2	9	6	4
	6. Seviye	0	0	0	0	0	3	0

Tablo 50 incelendiğinde Kontrol grubu öğrencilerin %38'inin (*n*=447) son-testte ön-teste göre daha düşük bir seviyeye indiği, %34'ünün (*n*=397) daha yüksek seviyeye çıktığı ve %28'inin (*n*=330) aynı seviyede kaldığı görülmektedir.

4.2.2 BTM Analizleri ve Öğrenci Örtük Sınıflarındaki Değişim

Öğrenci örtük sınıfları BTM analizleri ile belirlenmiştir. Analizlerde DINA, G-DINA ve Fusion(RUM) model kullanılmıştır. Araştırmada data ile en yüksek uyumu gösteren modeller incelenmiş ve ön-son test için DINA model sınıflamalarının daha uygun olduğu belirlenmiştir. Araştırma sürecinde uygulanan izleme testleri ve ön-son test için matematiksel yeterlikleri

gösteren 4 özellik belirlenmiş ve Q matrisler bu özelliklere göre belirlenmiştir (Bkz. Bölüm 3.1.2.). Ön-son test için belirlenen Q matris EK 26'da verilmiştir. Araştırma kapsamında belirlenen 4 özellik için 16 örtük sınıf oluşmuştur. Aşağıda bu 16 sınıf için ön ve son testlerden elde edilen dağılımların frekansları verilmiştir.

Tablo 51. Gruplara Göre Ön ve Son Test Örtük Sınıf Frekansları

	Deney-1		Deney-2		Kontrol	
	Ön-test	Son-test	Ön-test	Son-test	Ön-test	Son-test
0000	246	109	251	205	237	231
0001	57	89	48	109	44	110
0010	137	54	130	60	101	83
0011	47	106	59	63	48	56
0100	49	59	49	66	64	53
0101	55	31	44	32	52	24
0110	58	62	51	75	55	76
0111	81	70	79	61	76	71
1000	69	79	99	50	53	58
1001	71	49	60	49	62	46
1010	52	35	39	47	47	38
1011	66	46	59	57	47	47
1100	38	36	31	23	25	30
1101	57	26	51	28	57	21
1110	45	84	25	82	36	42
1111	178	371	155	223	165	183
Toplam	1306	1306	1230	1230	1169	1169

Tablo 51'de görüldüğü gibi Deney gruplarında ön teste düşük sayıda özelliğe sahip olan sınıflarda yer alan öğrenci sayılarının son testte giderek azaldığı görülmektedir. Sınıflar arasındaki değişimin oranları Tablo 52'de verilmiştir.

Tablo 52. Gruplara Göre Ön-Son Test Arasında Örtük Sınıf Değişim Yüzdeleri

	Artış	Sabit	Düşüş
Deney-1	51%	33%	16%
Deney-2	40%	33%	26%
Kontrol	29%	44%	28%

Tablo 52, ön ve son test arasındaki öğrencilerin atandıkları sınıflardaki değişimi göstermektedir. Değişim miktarının hesaplandığı tablolar EK 27, 28 ve 29'da verilmiştir. Örtük sınıflar öğrencilerin sahip olduğu matematiksel yeterlik sayısının bir göstergesidir. Örneğin "0000" sınıfındaki öğrenci hiçbir özelliğe sahip değilken "1001" sınıfındaki öğrenci iki özelliğe sahiptir. Tablo 52'de artış olarak ifade edilen bölüm ön testte göre son testteki örtük sınıfının

daha fazla özellik içerdiğini göstermektedir. Sabit sütununda ise öğrencilerin ön ve son testte sahip olduğu özellik sayısı eşittir. Düşüş sütünü ise ön son test arasında sahip oldukları özellik sayısında düşüş olan öğrenci oranını göstermektedir. Tablo 52 incelendiğinde Deney-1 grubundaki öğrencilerini %51'i son testte ön teste göre daha yüksek bir sınıfa atanmıştır. Bu oran Deney-2 grubu için %40, Kontrol grubu için ise %29 düzeyindedir. Düşüş sütunu incelendiğinde Deney-1 grubundaki öğrencilerini %16'sı son testte ön teste göre daha düşük bir sınıfa atanmıştır. Bu oran Deney-2 grubu için %26, Kontrol grubu için ise %28 düzeyindedir. Ön ve son testte örtük sınıf düzeyleri değişmeyen öğrencilerin oranları Deney-1'de %33, Deney-2'de %33 ve Kontrol grubunda %44 düzeyindedir. Bu durum deneysel sürecin öğrencilerin daha yüksek örtük sınıflara geçmesinde etkili olduğuna işaret etmektedir.

4.3 Araştırma Sürecinde Ulaşılan Diğer Bulgular

Bu bölümde üst bilişsel izleme ölçümleri, madde formatı ile ilgili bulgular ve üst düzey düşünme becerileri ile ilgili bulgular raporlandırılmıştır.

4.3.1 Üst Bilişsel İzleme Ölçümlerine İlişkin Bulgular

Araştırmada üst bilişsel izleme performansına ait bulgular iki yönden incelenmiştir. Öncelikle üstbilişsel izleme performansı ile BTM analizlerine göre elde edilen matematiksel yeterlikler ile ilişki bulgular ve üst bilişsel izleme performansının öğrenci başarı düzeyleri ile ilişkisine yönelik bulgular olarak toparlanmıştır.

4.3.1.1 BTM Örtük Sınıflarına Göre Üstbilişsel İzleme Performansındaki Değişim

Araştırmada izleme-4 testinde uygulanan üstbilişsel izleme ölçümleri ile öğrencilerin matematiksel yeterliklerinin ne düzeyde ilişkili olduğuna yönelik analizler gerçekleştirilmiştir. Bu amaç doğrultusunda BTM modelleri ile elde edilen örtük sınıf dağılımları ve Tip-2 sinyal tespit kuramı ile elde edilen üstbilişsel izleme becerisi karşılaştırılmıştır. Katılımcıların örtük sınıflara ait sonsal olasılıkları, bu sınıflara ait testten alınan ortama doğru sayısı ve örtük sınıfların üstbilişsel izleme performansları (AUC) Tablo 53'de verilmiştir.

Tablo 53. Örtük Sınıflara ait Sonsal Olasılıklara göre Ortalamalar ve AUC Değerleri

Örtük sınıf	Sonsal olasılık	Ortalama doğru sayısı	Üstbilişsel izleme becerisi (AUC)
"0000"	0,0658	1,12	0,16
"1000"	0,0631	2,50	0,14
"1100"	0,0631	1,75	0,13
"1010"	0,0538	2,50	0,18
"1001"	0,0644	2,21	0,13
"0110"	0,0572	2,21	0,22
"1110"	0,0476	4,67	0,34
"1101"	0,0527	3,89	0,34

"1111"	0,0936	6,36	0,41

Tablo 53'te görüldüğü gibi öğrencilerin sahip oldukları özellik sayısı arttıkça üstbilişsel izleme becerisinde de bir artış olduğu görülmektedir. Hangi belli özelliklerin, üstbilişsel izleme performansında anlamlı olarak farklılaştığını belirlemek için, özelliklere sahip olup olmama durumu ile üstbilişsel izleme becerisi arasındaki farklar, tüm beceriler ayrı ayrı olmak üzere, ilişkisiz örneklemler t-test ile analiz edilmiştir; bkz. Tablo 54.

Tablo 54. Matematiksel Yeterlikler ve Üstbilişsel İzleme Performanslarının ait İlişkisiz Örneklemler t-testi Sonuçları

	Beceriye sahip olup olmama	*n*	ort.	Ss	*t*	*df*	*p*
İletişim ve ilişkilendirme	1	60	0,225	0,293	0,953	118	0,343
	0	60	0,175	0,280			
Matematikleştirme	1	40	0,267	0,265	1,823	118	0,071
	0	80	0,167	0,292			
Akıl Yürütme ve Strateji Geliştirme	1	42	0,301	0,302	2,914	118	0,004
	0	78	0,146	0,264			
Sembolik ve Teknik Dil Kullanımı	1	44	0,233	0,309	0,945	118	0,347
	0	76	0,181	0,273			

Tablo 54'de gösterilen t-testi sonuçlarına göre, sadece "akıl yürütme ve strateji geliştirme" özelliğine sahip olmanın üstbilişsel izleme üzerinde etkili olduğu bulunmuştur. Diğer bir ifadeyle, akıl yürütme ve strateji geliştirme özelliğine sahip olan öğrencilerin üstbilişsel izleme puanları ortalaması (0,30), bu özelliğe sahip olmayan öğrencilerden (0,15) istatistiksel olarak anlamlı düzeyde büyüktür. İletişim ve ilişkilendirme, matematikleştirme ve sembolik ve teknik dil kullanımı özelliklerine sahip olup olmama durumunda, öğrencilerin üstbilişsel izleme puanları arasında istatistiksel olarak anlamlı bir fark bulunmamıştır (Basokcu & Güzel, 2017).

4.3.1.2 Öğrenci Başarı Düzeylerine Göre Üstbilişsel İzleme Performansındaki Değişim

Araştırmada İzleme-4 testinin bütün örneklemi için üst bilişsel izleme performansları ve aynı zamanda öğrencilerin önsel ve sonsal başarı tahminleri ölçümlenmiştir. İzleme performanslarını ve önsel ve sonsal tahminlerin test başarısı açısından karşılaştırılabilmesi amacıyla, öğrenci puanları üç performans grubuna ayrılmıştır. Bunlar ortalamadan manidar uzaklaşma düzeyine göre düşük, orta ve yüksek performanslı gruplar. Bu gruplara ait betimsel istatistikler Tablo 55'te verilmiştir.

Tablo 55. Başarı Gruplarına İlişkin Betimsel İstatistikler

Grup	N	Ortalama	S.s.	AUC Ort

Düşük	851	1,53	0,64	0,15
Orta	1354	3,44	0,5	0,27
Yüksek	627	5,58	0,82	0,42

Grupların üstbilişsel izleme yetenekleri açısından farklılık gösterip göstermediğini saptamak için tek yönlü varyans analizi (ANOVA) yapılmıştır.

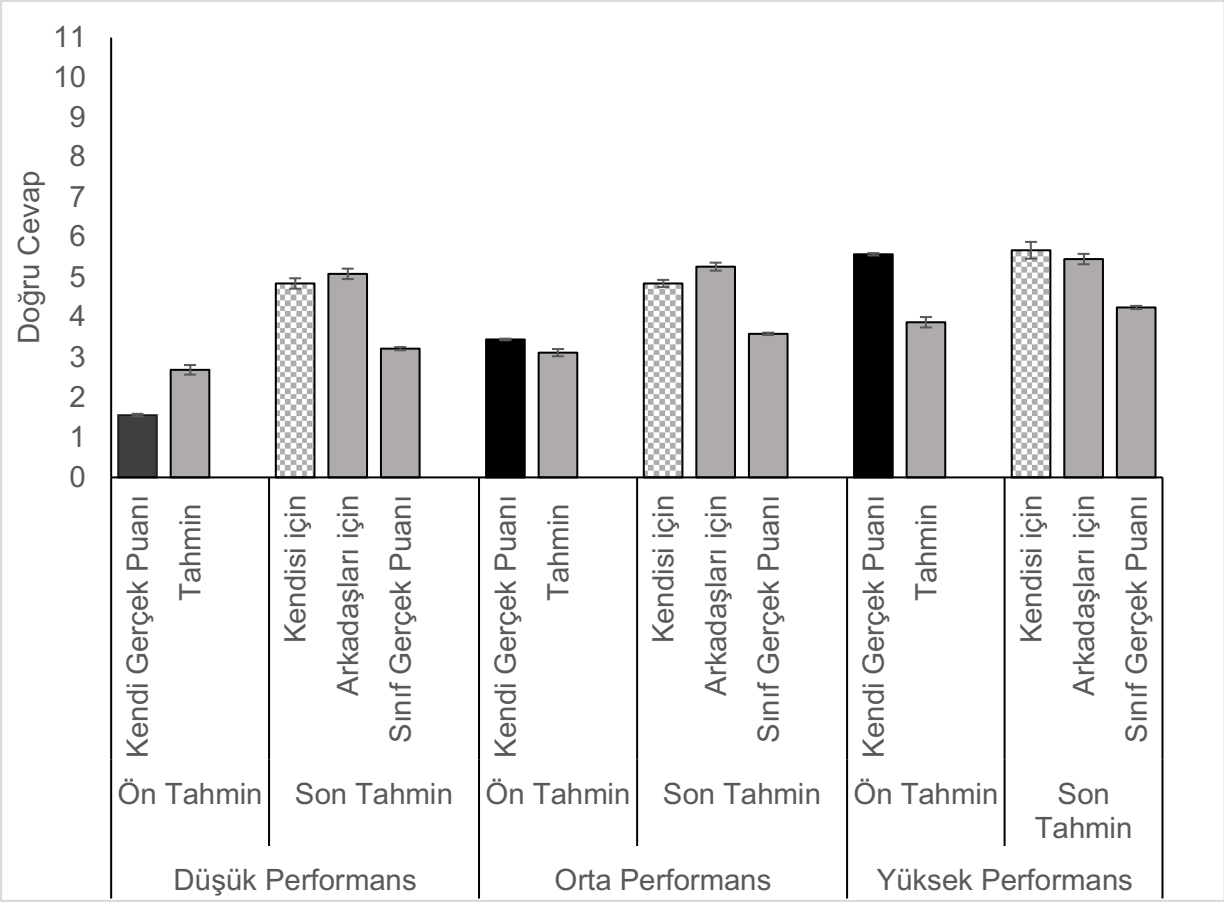
Tablo 56. Grupların Son-Test Puanlarına ait ANOVA Analizi

	Kareler Toplamı	sd	Kareler ortalaması	F	P
Gruplar arası	14,601	2	7,3	81,687	0,00
Gruplar içi	252,831	2829	0,089		
Toplam	267,432	2831			

Sonuçlar, üst bilişsel izleme performansları açısından grupların birbirlerinden anlamlı olarak farklı olduğunu göstermektedir (F (2,2829) = 81,687; p <0,000). Post-hoc karşılaştırmalarına göre, düşük-performans grubunun orta ve yüksek performanslı gruplardan (M = 0,27, s = 0,32; M = 0,42, s =0,82) daha düşük üst bilişsel izleme performansı (M = 0,15, s = 0,64) göstermiştir. Öğrencilerin testteki performansı daha iyi olduğunda, doğru ve yanlış yanıtlarını saptama yeteneklerinin de daha yüksek olduğu bulunmuştur.

4.3.2 Öğrencilerin Test Performans Farkındalıklarına İlişkin Bulgular

Proje kapsamında üstbilişsel ölçümlerin bir parçası da öğrencilerin kendi ve arkadaşlarının performansları hakkındaki farkındalıklarının ölçümlerini içermektedir. Bu amaçla öğrencilerin İzleme-4 uygulamasında test öncesinde ve sonrasında kendi performansları ve sınıf arkadaşlarının performansları hakkında bazı ölçümler alınmıştır. Analizlerde öğrencilerin kendi ve arkadaşlarının performansları hakkındaki farkındalık düzeyleri gerçek performanslarına göre belirlenen yüksek, orta ve düşük performans grupları için karşılaştırılmıştır. Aşağıda Şekil 22, bu gruplar için öğrencilerin tahminlerini göstermektedir.



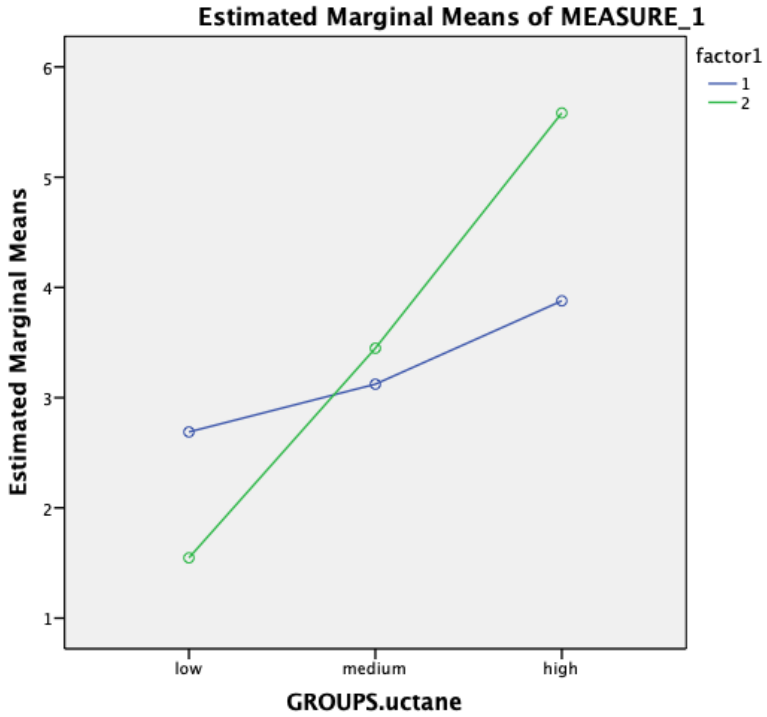
Şekil 22. Öğrencilerin Gruplara Göre Performans Tahminleri

Şekil 22'de görüldüğü gibi öğrenciler test öncesinde kendileri için bir ortalama tahmininde bulunmuşlardır. Test bittikten sonra ise kendi performansları ve arkadaşları hakkında tekrar bir tahmin yapmışlardır. Şekil 22'de öğrencilerin tahminleri ile birlikte aynı zamanda kendilerinin ve sınıflarının gerçek performansları da verilmiştir. Grupların ön ve son tahminleri arasındaki farkların anlamlılığı karışık faktöriyel varyans analizi ile incelenmiştir. Ön tahmin ve gerçek performans ortalamaların gruplar arasındaki farklara ilişkin analizler Tablo 57'de verilmiştir.

Tablo 57. Gruplara Göre Ön Tahmin ve Gerçek Puanlarının ANOVA Sonuçları

Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Denekler arası	21021,704	2769			
Grup	4813,016	2	2406,508	410,817	0,000
Hata	16208,69	2767	5,858		
Denekler içi	16986,52	2770			
Ölçüm	109,945	1	109,945	19,719	0,000
Grup*Ölçüm	1449,042	2	724,521	129,946	0,000
Hata	15427,54	2767	5,576		
Toplam	38008,231	5539			

Tablo 57 incelendiğinde öğrencilerin bulundukları grup ile ön tahminleri ve test puanları arasındaki farklılaşmanın etkili olduğu görülmektedir ($F(2, 2767) = 19{,}719$; $p = 0{,}00$). Hangi gruplar arası farkların anlamlı olduğunu belirlemek için Post Hoc LSD testi uygulanmıştır. Analiz sonuçlarına göre Yüksek, Orta (p= 0,00), Orta, Düşük (p= 0,00) ve Yüksek, Düşük (p=0,00) performans grupları arasındaki farklar ilk yazılan grup lehine anlamlıdır. Aşağıda marjinal ortalamalardaki değişim Şekil 23'de verilmiştir.



Şekil 23.Gruplara Ait Ön Tahmin Test Performans Puanları Profil Grafiği

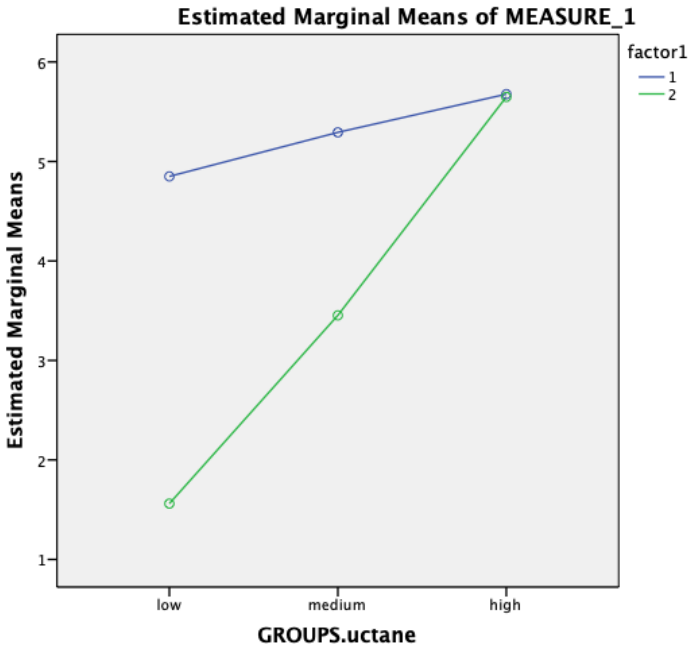
Şekil 23 incelendiğinde performans grupları ile tahmin ve gerçek puan arasındaki kesişim açıkça görülmektedir. Düşük performans grubunun kendi test performansları ile ilgili olduğundan daha yüksek, Yüksek performans grubunun ise olduğundan daha düşük bir beklenti ile teste başladıkları görülmektedir. Öğrencilere testi tamamladıktan sonra kendi performansları ile ilgili tekrar tahminde bulunmaları istenmiştir. Bu tahminlerin gerçek performansları ile arasındaki farkların gruplara göre analizleri tekrarlı ölçümler için varyans analizi ile gerçekleştirilmiştir. Tablo 58 bu analizlerin sonuçlarını vermektedir.

Tablo 58. Gruplara Göre Son Tahmin ve Gerçek Puanlarının ANOVA Sonuçları

Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Denekler arası	9960,145	1933			
Grup	2985,924	2	1492,962	413,366	0,000
Hata	6974,221	1931	3,612		
Denekler içi	10757,64	1934			
Ölçüm	2602,111	1	2602,111	734,554	0,000

Grup*Ölçüm	1315,090	2	657,545	185,62	0,000
Hata	6840,441	1931	3,542		
Toplam	20717,787	3868			

Tablo 58 incelendiğinde öğrencilerin bulundukları grup ile ön tahminleri ve test puanları arasında anlamlı bir fark bulunmuştur; $F(2, 1931) = 185,62$; $p = 0,00$. Hangi gruplar arası farkların anlamlı olduğunu belirlemek için Post Hoc LSD testi uygulanmıştır. Analiz sonuçlarına göre Yüksek, Orta (p= 0,00), Orta, Düşük (p= 0,00) ve Yüksek, Düşük (p=0,00) performans grupları arasındaki farklar ilk yazılan grup lehine anlamlıdır. Aşağıda marjinal ortalamalardaki değişim Şekil 24'de verilmiştir.



Şekil 24.Gruplara Ait Son Tahmin Test Performans Puanları Profil Grafiği

Şekil 24 incelendiğinde Düşük performans grubunun test performanslarını olduğundan çok yüksek gördüğü anlaşılmaktadır (p=0,00). Bununla birlikte Yüksek grubun test performanslarını gerçek performanslarına çok yakın tahminledikleri ve bu iki değer arasında istatistiksel olarak anlamlı bir fark olmadığı görülmektedir (p=0,844). Bulgular incelendiğinde üst düzey düşünme becerileri içinde önemli bir yer tutan üstbiliş becerisi araştırma örnekleminde de literatür ile uyumlu bir dağılım sergilemiştir. Bununla birlikte araştırma kapsamında geliştirilen testlerin üst düzey düşünme becerilerini ölçebildiğine dair de bir bilgi sunmaktadır (Basokcu & Güzel, 2018).

4.3.3 Açık Uçlu Sorulara İlişkin Bulgular

Araştırmada geliştirilen izleme testlerde hem çoktan seçmeli hem de yapılandırılmış yanıtlı maddeler birlikte kullanılmıştır. Aynı test içinde farklı madde formatlarının birlikte kullanıldığı

geniş ölçekli sınav sayısı oldukça azdır (PISA, TIMSS, PEARLS). Bununla birlikte değişik madde formatlarının kullanıldığı ve BTM modellerine göre geliştirilmiş ve analiz edilmiş bir test uygulanması literatürde bulunmamaktadır. Bu nedenle proje kapsamında kullanılan farklı madde formatlarından oluşan testlerin psikometrik yapılarına ilişkin çalışmalara da yer verilmiştir.

Araştırma, üst düzey düşünme becerilerini ölçmeye yönelik hazırlanan çoktan seçmeli ve yapılandırılmış yanıtlı maddelerden oluşan 4 izleme testinde yer alan maddelerin BTM parametreleri acısından karşılaştırılmasına dayanmaktadır. Her bir izleme testine katılan öğrenci sayıları ve madde sayıları aşağıdaki Tablo 59'da verilmiştir.

Tablo 59. İzleme Testlerine Ait Madde Dağılımı

	Kişi Sayısı(N)	Çoktan Seçmeli	Yapılandırılmış Yanıtlı	Toplam
İzleme 1	2918	7	4	11
İzleme 2	2832	8	5	13
İzleme 3	2875	10	3	13
İzleme 4	2834	7	2	9

Araştırma kapsamında ele alınan yapılandırılmış ve çoktan yanıtlı maddelerin DINA analizleri için OX programından yararlanılmıştır. Çoktan seçmeli ve yapılandırılmış yanıtlı maddelere ait g ve s parametreleri aşağıdaki Tablo 60'ta verilmiştir.

Tablo 60. Çoktan Seçmeli Test Madde Parametreleri

Soru kodu	G	S	Soru kodu	G	S	Soru kodu	G	S
1021	0,50	0,15	1018	0,26	0,60	1020_C	0,20	0,55
3690	0,18	0,41	1014_C	0,23	0,76	1021_B	0,17	0,48
6728	0,22	0,52	1012_A	0,20	0,41	1021_C	0,15	0,53
1032	0,29	0,49	1012_B	0,21	0,46	5017_A	0,17	0,33
1033	0,32	0,31	1023_B	0,19	0,50	5036_A	0,27	0,17
5158_B	0,21	0,11	1024	0,25	0,46	5036_B	0,20	0,61
5158_C	0,23	0,19	1028	0,26	0,23	5058_A	0,54	0,07
1011_A	0,15	0,56	1022_B	0,13	0,28	5058_B	0,30	0,24
1011_C	0,28	0,46	1026	0,39	0,43	5059	0,03	0,60
1015_B	0,28	0,33	1020_A	0,34	0,36	5057	0,67	0,06
1017	0,14	0,51	1020_B	0,21	0,45			

Tablo 60'ta verilen çoktan seçmeli maddelere ait *g* ve *s* parametrelerinin ortalamaları *g* için 0,25, ve s için 0,39 olarak hesaplanmıştır. Testlerin ortalama güçlük düzeylerinin düşük olmasından dolayı *s* parametrelerinin daha yüksek olduğu görülmektedir. Bununla birlikte maddelerin kabul edilebilecek parametre aralıklarında olduğu söylenebilir.

Yapılandırılmış yanıtlı maddelere ait *g* ve *s* parametreleri ölçme değişmezliği açısından incelenmiştir. Yapılandırılmış yanıtlı maddelerin çoktan seçmeli maddeler ile birlikte hesaplanan parametreleri ile tek başına analizleri sonucunda elde edilen parametreler arasındaki farklar Fisher Z istatistiği ile hesaplanmıştır. Tablo 61'de hem iki koşul (YY Birleşik = Yapılandırılmış Yanıtlı Birleşik/ YY Ayrı=Yapılandırılmış Yanıtlı Ayrı) için madde parametreleri hem de Z testi sonuçları verilmiştir.

Tablo 61. Yapılandırılmış Yanıtlı Maddeleri Ait Parametre Değişmezliğinin İncelenmesi

	YY Birleşik		YY Ayrı		
	G	S	G	S	p
1027	0,11	0,28	0,11	0,46	.001
1025	0,11	0,24	0,06	0,14	.064
7339	0,00	0,57	0,00	0,58	.896
9796	0,20	0,36	0,22	0,38	.891
1011_B	0,16	0,40	0,11	0,42	.658
1016	0,21	0,32	0,14	0,32	.989
1015_A	0,03	0,40	0,03	0,43	.429
1014_A	0,25	0,39	0,25	0,45	.152
1014_B	0,00	0,60	0,00	0,61	.925
1023_A	0,07	0,47	0,07	0,58	.071
1022_A	0,11	0,66	0,08	0,65	.921
1021_A	0,02	0,50	0,02	0,57	.150
5017_B	0,01	0,54	0,00	0,57	.528
5036_C	0,12	0,51	0,11	0,55	.605

Tablo 61'de verilen parametre değerlerinin ortalamaları incelendiğinde birinci koşul için *g*=0,10 *s*= 0,44 olarak ikinci koşul için *g*=0,08 *s*=0,47 olarak hesaplanmıştır. Bu maddelerde *g* parametrelerinin oldukça düşük *s* parametrelerinin ise beklenen değerin üzerinde olduğu görülmüştür. İki koşul için parametre değişmezliği incelendiğinde bir madde dışında bütün maddeler için parametreler arasında istatistiksel olarak anlamlı bir fark gözlenmemiştir. Araştırma sonucunda elde edilen bulgular incelendiğinde tahmin parametresinin yapılandırılmış yanıtlı maddelerde çoktan seçmeli maddelere göre manidar düzeyde düşük olduğu gözlenmiştir. Kaydırma parametresi için ise tam tersi bir durum söz konusudur. Bu durum parametrelerin üretilme koşulları madde formatlarının temel özellikleri göz önüne alındığında literatür ile uyumludur. Bununla birlikte yapılandırılmış yanıtlı maddelerin test içinde ve testten ayrı analizleri sonuçlarında elde edilen bulgular karma madde formatı kullanılan PISA, TIMSS gibi testlerde BTM analizleri yapmanın parametre değişmezliği açısından herhangi bir soruna yol açmayacağına işaret etmektedir (Basokcu & Ceylan, 2018).

4.3.4 Uygulanan Testlere Yönelik Tutumlar ile İlgili Bulgular

Araştırmada izleme-2 testinden itibaren her test için öğrencilere, uygulamaya ilişkin tutumları hakkında dört soru yöneltilmiştir. Bu sorulara ilişkin yanıtlar her kitapçığın sonunda Şekil 25'te gösterilen form aracılığı ile toplanmıştır.

Lütfen aşağıdaki ifadelere katılma düzeyinizi işaretleyiniz.	HİÇ	ÇOK AZ	ORTA	ÇOK	TAMAMEN
Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum.	1	2	3	4	5
Bu testteki sorulara benzer sorularla daha sık karşılaşmak isterim.	1	2	3	4	5
Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi.	1	2	3	4	5
Bu testte yer alan soruların öğrenme isteğimi artırdığını düşünüyorum.	1	2	3	4	5

Şekil 25. Uygulamaya İlişkin Görüş Formu

Şekil 25'te görüldüğü gibi dört soru 5'li dereceleme ölçeğinde öğrencilerin ifadelere katılma düzeylerini "hiç" ile "tamamen" aralığında ölçmektedir. Öğrencilere formda yöneltilen ifadeler sırasıyla "Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum (M1)", "Bu testteki sorulara benzer sorular ile daha sık karşılaşmak isterim (M2)", "Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi (M3)" ve "Bu testte yer alan soruların öğrenme isteğimi arttırdığını düşünüyorum(M4) " şeklindedir. Bu ifadelere üç izleme testinde verilen cevaplar ilişkin betimsel istatistikler Tablo 62'de verilmiştir.

Tablo 62. Testle İlgili Görüşlerin İzleme Testlerine göre Betimsel İstatistikleri

			İzleme-2		İzleme-3		İzleme-4	
Grup	Madde	N	Ort.	ss.	Ort.	ss.	Ort.	ss.
Deney-1	M1	1501	2,963	1,222	2,994	1,379	3,060	1,345
	M2	1501	2,668	1,381	2,699	1,518	3,002	1,469
	M3	1501	3,093	1,379	3,235	1,514	3,194	1,507
	M4	1501	3,086	1,403	3,135	1,544	3,211	1,522
Deney-2	M1	1443	2,950	1,207	2,721	1,358	2,781	1,398
	M2	1443	2,461	1,357	2,458	1,466	2,685	1,491
	M3	1443	3,014	1,347	2,970	1,430	2,942	1,484
	M4	1443	3,039	1,445	2,930	1,486	2,902	1,544
Toplam	M1	2944	2,957	1,215	2,868	1,369	2,931	1,384
	M2	2944	2,574	1,374	2,588	1,500	2,855	1,485
	M3	2944	3,057	1,365	3,113	1,516	3,077	1,460
	M4	2944	3,065	1,422	3,041	1,537	3,067	1,520

İzleme testleri arasında öğrencilerin verilen ifadelere katılma düzeylerindeki değişim tekrar ölçümler ANOVA analiz ile test edilmiştir. Analizde faktör üç izleme testi ve grup, Deney-1 ve Deney-2 gruplarıdır (faktör x grup 3 x 2).

İlk ifade olan “Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum (M1)” için ölçümler arasında gruplara göre farklılaşmalar için yapılan ANOVA sonuçları Tablo 63’de verilmiştir.

Tablo 63. “Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum” ifadesi için ANOVA Sonuçları

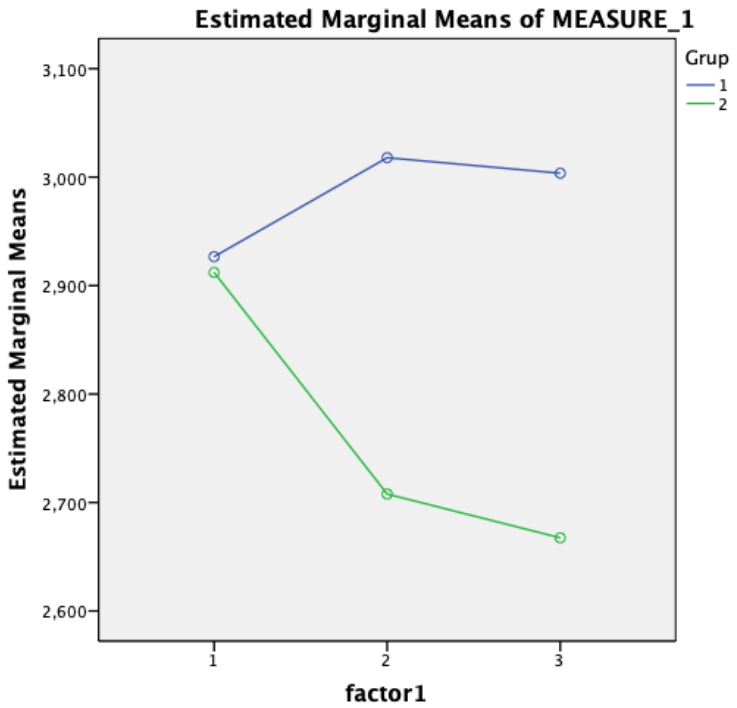
Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Denekler arası	26745,981	979			
Grup	34,944	1	34,944	11,656	0,001
Hata	2931,906	978	2,998		
Denekler içi	2208,994	1960			
Ölçüm	3,514	2	1,757	1,569	0,208
Grup*Ölçüm	15,323	2	7,661	6,842	0,001
Hata	2190,157	1956	1,12		
Toplam	28954	2938			

Analiz sonuçları incelendiğinde ilk ifade için grup ölçüm etkileşiminin istatistiksel olarak anlamlı olduğu görülmektedir; $F(2, 1956) = 6,842$; $p = 0,00$. Hangi gruplar arası farkların anlamlı olduğunu belirlemek için Post Hoc testi olarak ilişkisiz örneklemler t testi kullanılmıştır. Tablo 64 ilk ifade de her bir izleme için grup ortalamaları arasındaki farkın anlamlılığını göstermektedir.

Tablo 64. “Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum” ifadesi için T Testi Sonuçları

	Grup	N	Ort.	ss.	t	sd	p
İzleme-2	Deney-1	961	2,96	1,22	0,219	1756	0,827
	Deney-2	797	2,95	1,21			
İzleme-3	Deney-1	986	2,99	1,38	4,272	1830	0,00
	Deney-2	846	2,72	1,34			
İzleme-4	Deney-1	1098	3,06	1,36	4,569	2041	0,00
	Deney-2	945	2,78	1,40			

Tablo 64’de yer alan sonuçlar incelendiğinde ifadenin ilk uygulamasında Deney-1 ve Deney-2 grupları ortalamaları arasında anlamlı bir fark gözlenmemiştir. Bununla birlikte ikinci ve üçüncü uygulamalarda Deney-1 grubunun *“Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum”* ifadesine katılma düzeyi Deney-2 grubundan istatistiksel olarak anlamlı düzeyde büyüktür. Şekil 26 bu madde için grup ölçüm etkileşimini göstermektedir.



Şekil 26. Deney-1 ve Deney-2 Gruplarının "Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum" ifadesi için Profil Grafiği

Şekil 26'da görüldüğü gibi Deney-1 grubu öğrencilerinin *"Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum"* ifadesine katılma düzeyleri ikinci ve üçüncü ölçümde Deney-2 grubuna göre önemli düzeyde yükselmektedir.

İkinci ifade olan "Bu testteki sorulara benzer sorularla daha sık karşılaşmak isterim (M2)" için ölçümler arasında gruplara göre farklılaşmalar için yapılan ANOVA sonuçları Tablo 65'te verilmiştir.

Tablo 65."Bu testteki sorulara benzer sorularla daha sık karşılaşmak isterim" ifadesi için ANOVA Sonuçları

Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Denekler arası	22731,438	976			
Grup	52,499	1	52,499	14,568	0,000
Hata	3510,013	974	3,604		
Denekler içi	2570,029	1952			
Ölçüm	44,070	2	22,035	17,016	0,000
Grup*Ölçüm	3,439	2	1,719	1,328	0,265
Hata	2522,520	1948	1,295		
Toplam	25301,467	2928			

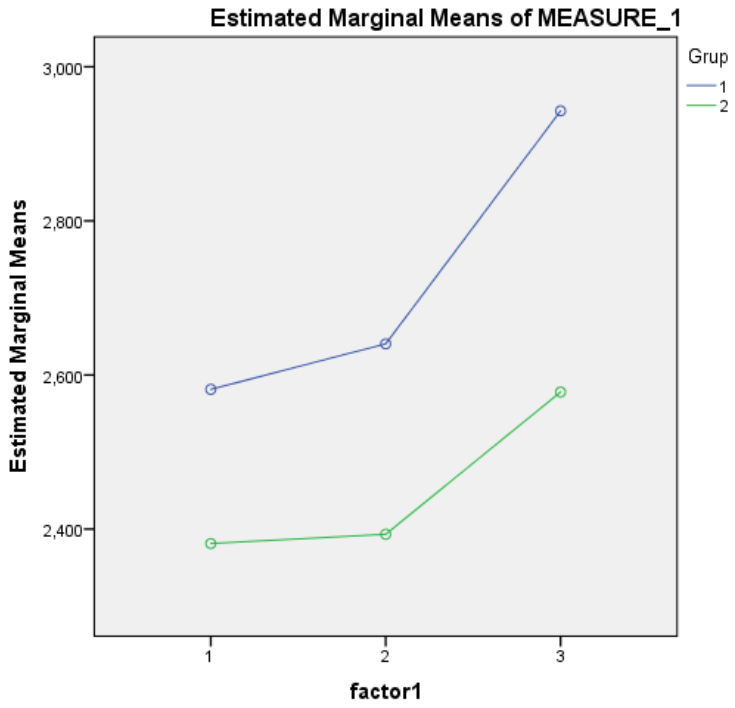
Analiz sonuçları incelendiğinde ikinci ifade için grup ölçüm etkileşiminin istatistiksel olarak anlamlı olduğu görülmektedir; $F(2, 1948) = 1,719$; $p = 0,265$. Hangi gruplar arası farkların

anlamlı olduğunu belirlemek için Post Hoc testi olarak ilişkisiz örneklemler t testi kullanılmıştır. Tablo 66 ikinci ifadede her bir izleme için grup ortalamaları arasındaki farkın anlamlılığını göstermektedir.

Tablo 66.“Bu testteki sorulara benzer sorularla daha sık karşılaşmak isterim” ifadesi için T Testi Sonuçları

	Grup	N	Ort.	ss.	t	sd	p
İzleme-2	Deney-1	956	2,67	1,38	3,152	1746	0,002
	Deney-2	792	2,46	1,36			
İzleme-3	Deney-1	985	2,70	1,52	3,443	1826	0,001
	Deney-2	843	2,46	1,47			
İzleme-4	Deney-1	1091	3,00	1,47	4,814	2030	0,000
	Deney-2	941	2,69	1,49			

Tablo 66’da yer alan sonuçlar incelendiğinde ifadenin bütün uygulamalarda Deney-1 ve Deney-2 grupları ortalamaları arasında anlamlı bir fark gözlenmiştir. Bu fark bütün izleme testlerinde Deney-1 grubu lehinedir. Şekil 27 bu madde için grup ölçüm etkileşimini göstermektedir.



Şekil 27.Deney-1 ve Deney-2 Gruplarının “Bu testteki sorulara benzer sorularla daha sık karşılaşmak isterim” ifadesi için Profil Grafiği

Şekil 27’de görüldüğü gibi Deney-1 grubu öğrencilerinin *“Bu testteki sorulara benzer sorularla daha sık karşılaşmak isterim”* ifadesine katılma düzeyleri üç ölçümde de Deney-2 grubuna göre önemli düzeyde yükselmektedir.

Üçüncü ifade olan “Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi (M3)” için ölçümler arasında gruplara göre farklılaşmalar için yapılan ANOVA sonuçları Tablo 67’de verilmiştir.

Tablo 67. “Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi" ifadesi için ANOVA Sonuçları

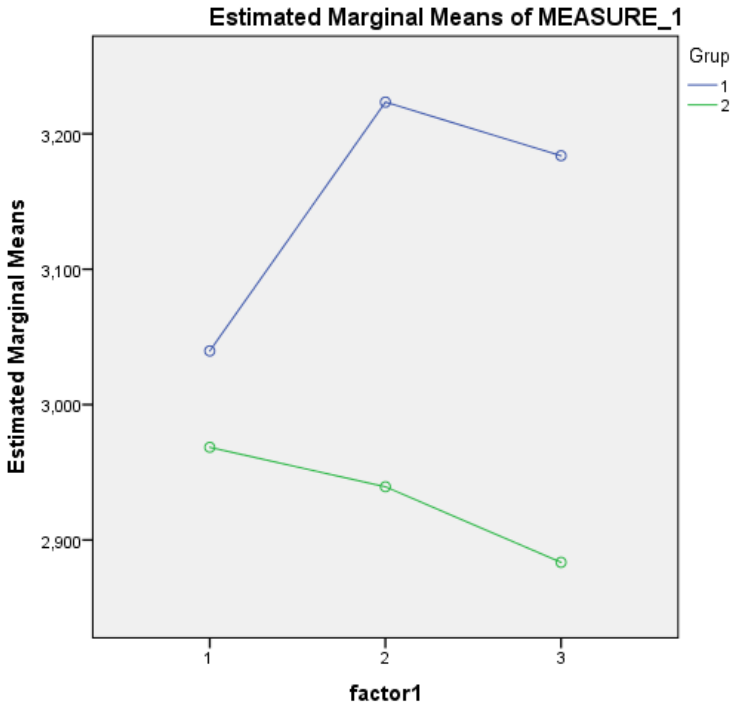
Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Denekler arası	30003,106	967			
Grup	33,877	1	33,877	8,715	0,003
Hata	3751,077	965	3,887		
Denekler içi	2342,692	1934			
Ölçüm	2,880	2	1,440	1,192	0,304
Grup*Ölçüm	7,731	2	3,865	3,199	0,041
Hata	2332,081	1930	1,208		
Toplam	32345,797	2928			

Analiz sonuçları incelendiğinde üçüncü ifade için grup ölçüm etkileşiminin istatistiksel olarak anlamlı olduğu görülmektedir; $F(2, 1930) = 3,865$; $p = 0,041$. Hangi gruplar arası farkların anlamlı olduğunu belirlemek için Post Hoc testi olarak ilişkisiz örneklemler t testi kullanılmıştır. Tablo 68 üçüncü ifadede her bir izleme için grup ortalamaları arasındaki farkın anlamlılığını göstermektedir.

Tablo 68. "Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi" ifadesi için T Testi Sonuçları

	Grup	N	Ort.	ss.	t	sd	p
İzleme-2	Deney-1	954	3,09	1,38	1,207	1740	0,227
	Deney-2	788	3,01	1,35			
İzleme-3	Deney-1	980	3,23	1,51	3,722	1818	0,000
	Deney-2	840	2,97	1,51			
İzleme-4	Deney-1	1083	3,19	1,43	3,874	2018	0,000
	Deney-2	937	2,94	1,48			

Tablo 68’de yer alan sonuçlar incelendiğinde ifadenin bütün uygulamalarda Deney-1 ve Deney-2 grupları ortalamaları arasında anlamlı bir fark gözlenmiştir. Bu fark bütün izleme testlerinde Deney-1 grubu lehinedir. Şekil 28 bu madde için grup ölçüm etkileşimini göstermektedir.



Şekil 28. Deney-1 ve Deney-2 Gruplarının "Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi" ifadesi için Profil Grafiği

Şekil 28'te görüldüğü gibi Deney-1 grubu öğrencilerinin *"Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi"* ifadesine katılma düzeyleri üç ölçümde de Deney-2 grubuna göre önemli düzeyde yükselmektedir.

Dördüncü ifade olan "Bu testte yer alan soruların öğrenme isteğimi artırdığını düşünüyorum (M4)" için ölçümler arasında gruplara göre farklılaşmalar için yapılan ANOVA sonuçları Tablo 69'da verilmiştir.

Tablo 69. "Bu testte yer alan soruların öğrenme isteğimi artırdığını düşünüyorum" ifadesi için ANOVA Sonuçları

Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Denekler arası	28568,444	936			
Grup	14,188	1	14,188	3,346	0,068
Hata	3960,102	934	4,240		
Denekler içi	2352,229	1872			
Ölçüm	0,325	2	0,163	0,129	0,879
Grup*Ölçüm	4,300	2	2,150	1,711	0,181
Hata	2347,603	1868	1,257		
Toplam	30920,673	2928			

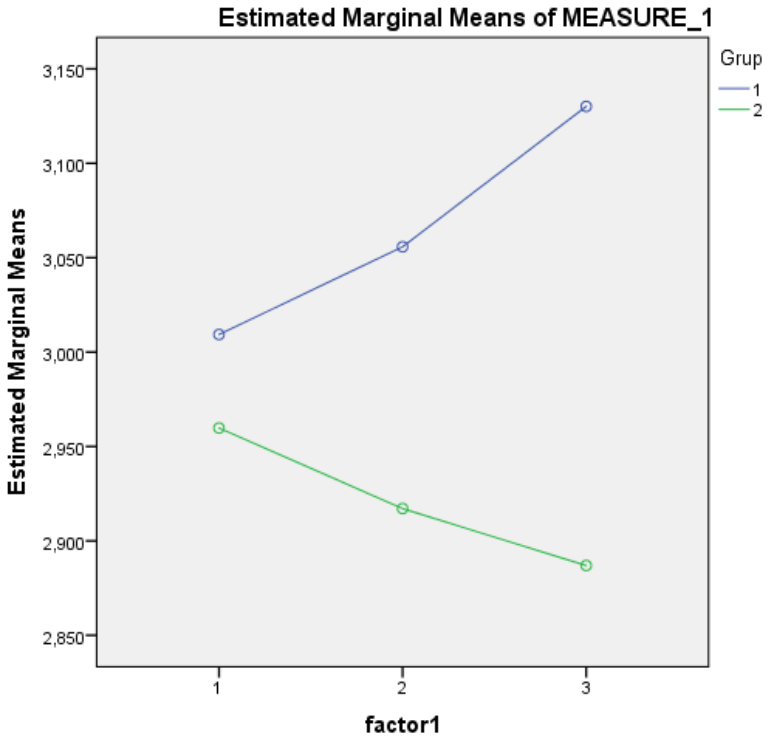
Analiz sonuçları incelendiğinde dördüncü ifade için grup ölçüm etkileşiminin istatistiksel olarak anlamlı olduğu görülmektedir; $F(2, 1868) = 2,150$; $p = 0,181$. Hangi gruplar arası farkların

anlamlı olduğunu belirlemek için Post Hoc testi olarak ilişkisiz örneklemler t testi kullanılmıştır. Tablo 70 dördüncü ifadede her bir izleme için grup ortalamaları arasındaki farkın anlamlılığını göstermektedir.

Tablo 70. "Bu testte yer alan soruların öğrenme isteğimi artırdığını düşünüyorum" ifadesi için T Testi Sonuçları

	Grup	N	Ort.	ss.	t	sd	p
İzleme-2	Deney-1	951	3,09	1,40	0,684	1737	0,494
	Deney-2	788	3,04	1,45			
İzleme-3	Deney-1	970	3,14	1,54	2,826	1800	0,005
	Deney-2	832	2,93	1,52			
İzleme-4	Deney-1	1072	3,21	1,49	4,552	2001	0,000
	Deney-2	931	2,90	1,54			

Tablo 70'de yer alan sonuçlar incelendiğinde ifadenin bütün uygulamalarda Deney-1 ve Deney-2 grupları ortalamaları arasında anlamlı bir fark gözlenmiştir. Bu fark bütün izleme testlerinde Deney-1 grubu lehinedir. Şekil 29 bu madde için grup ölçüm etkileşimini göstermektedir.



Şekil 29. Deney-1 ve Deney-2 Gruplarının "Bu testte yer alan soruların öğrenme isteğimi artırdığını düşünüyorum" ifadesi için Profil Grafiği

Şekil 29'da görüldüğü gibi Deney-1 grubu öğrencilerinin *"Bu testte yer alan soruların öğrenme isteğimi artırdığını düşünüyorum"* ifadesine katılma düzeyleri üç ölçümde de Deney-2 grubuna göre önemli düzeyde yükselmektedir.

5 SONUÇ ve TARTIŞMA

Bu bölümde proje sonucunda elde edilen bulgulara ilişkin değerlendirmeler yapılmış ve sonraki araştırmalar için önerilerde bulunulmuştur.

5.1 Deneysel Süreç Sonuçları

Projenin deney sürecinde temel iki hipotez test edilmiştir. Bunlardan birincisi öğrencilerin PISA, TIMSS gibi testlerin ölçtüğü özellikler ile ilişkili problem durumları ile öğrencileri daha sık karşılaştırmanın öğrencilerin bu testlerdeki başarı düzeyini arttıracağıdır. İkinci temel hipotez ise öğrencilerin karşılaştıkları problem durumları ile ilgili performansları hakkında BTM analizleri ile belirlenmiş detaylı geri bildirimler vermenin PISA, TIMSS testlerindeki başarı düzeyini arttıracağına yöneliktir.

Projede Deney-1 ve Deney-2 grupları arasındaki temel manipülasyon geri bildirim üzerinden kurgulanmıştır. Bu amaç doğrultusunda öğrenciye uygulanan izleme testlerinin hemen ardından Deney-1 gruplarına detaylı geri bildirim formları verilmiştir. Deney-2 gruplarına ise sadece doğru ve yanlış sayılarını gösteren bir geri bildirim formu verilmiştir. Detaylı geri bildirim formu testten elde edilen bir öğrenci profilidir. Bu profil içinde öğrencinin hangi özelliklere ne olasılıkla sahip olduğu, hangi üst düzey becerileri gösterebildiği ve en önemlisi hangi tür hataları tekrar ettiği görülmüştür. Deney-1 ve Deney-2 grupları arasındaki bu geri bildirim farklılığı aynı zamanda test etkisini görmeye de imkan vermiştir. BTM modellerine uygun olarak tasarlanan yazılım ile öğrencilerin kazanım, yeterlik ve hata profilleri dijital olarak üretilebilmiş aynı zamanda da bu sistem özellikle açık uçlu sorular için oldukça yeni ve esnek bir geribildirim modeli oluşturmuştur. Sonuçlar, gruplara ait ön-test ortalamaları arasında istatistiksel olarak anlamlı bir fark olmadığını, ancak son-testte istatistiksel olarak anlamlı bir farkın ortaya çıktığını göstermiştir. Deney-1 grubunun son test ortalaması Deney-2 ve Kontrol grubunun son test ortalamalarından; Deney-2 grubunun son test ortalaması da kontrol grubunun son test ortalamasından istatistiksel olarak anlamlı bir fark göstermiştir. Ayrıca deneysel işlem süreci sonuçlarına göre Deney-1, Deney-2'den ve Deney-2 de Kontrol grubundan ön-son test ortalamaları arasında daha yüksek düzeyde gelişim göstermiştir. Deneysel süreç tamamlandıktan sonra elde edilen bulgularının kalıcılığının incelenebilmesi amacı son testten yaklaşık 8 ay sonra uygulanan kalıcılık testi ortalamaları bütün gruplarda son testle aynı düzeyde farklılaşmış, ancak her grup için son ve kalıcılık testi arasında anlamlı bir fark gözlenmemiştir. Yine de Deney-1 grubunun son teste göre kalıcılık testindeki başarı düzeyi artmış, kontrol grubunda ise bu ölçümlerde belirgin düşüş gözlenmiştir.

Proje bulguları incelendiğinde araştırmanın her iki hipotezinin de istatistiksel olarak doğrulandığı görülmüştür. Bu konuda üst düzey düşünme becerileri gelişimi ile ilgili kuramlar

(S M Brookhart, 2010; Conklin, 2012; Goodson & Rohani, 1998) göz önüne alındığında maruz bırakma ve değerlendirme basamakları dikkat çekmektedir. Projenin önemli bulgularından biri olarak öğrencileri geliştirilmek istenen beceri ile ilgili problem durumlarına maruz kalmasının başarıyı önemli düzeyde etkilediğidir. Bununla birlikte BTM göz önüne alınarak geliştirilen soruların tanımlanma ve yapılandırılma aşamalarının daha bilimsel ve işlevsel olduğu görülmektedir. Bu sorular kullanılarak belirlenen öğrenci profilleri ise öğrencilerin sadece performansları değil aynı zamanda bilgilerinin kaynakları, kullandıkları veya kullanmadıkları stratejiler hakkında da önemli bilgilere ulaşılmayı sağlamıştır. Proje bulguları incelendiğinde geri bildirim verilen Deney-1 grubunun detaylı geri bildirim almayan Deney-2 grubuna göre istatistiksel olarak yüksek bir başarı düzeyine ulaştığı görülmektedir.

5.2 Öğrenci Yetenek Düzeyleri ve BTM Analizleri ile İlgili Sonuçlar

Projenin temel hipotezi göz önüne alındığında öğrencilerin başarı düzeylerinin uluslararası geniş ölçekli sınavlar standartlarında belirlenmesi gerektiği görülmektedir. Projede geliştirilen soruların PISA ve TIMSS soruları ile eşdeğer olduğu ve kapsam açısından da uyumlu olduğu ortaya konulmuştur. Bu bağlamda elde edilen bir diğer sonuç da PISA 2015 uygulaması ile araştırmada elde edilen yetenek düzeylerinin birbirlerine yakın olduğudur (Basokcu & Ceylan, 2017). Aynı zamanda proje ön ve son test puanları PISA uygulamalarında kullanılan yetenek belirleme yöntemleri ile hesaplanmış ve projede MTK analizleri ile belirlenen öğrenci yetenek puanları daha sonra PISA standart puanlarına dönüştürülmüş ve bu şekilde PISA matematiksel yeterliklerin dağılımlarına ulaşılmıştır. Bu dağılımlar kullanılarak öğrenci yetenek düzeylerinde değişimler hesaplanmıştır. Buna göre Deney-1, 2 ve kontrol grubu öğrencilerin üçte birinden fazlasının son-testte, ön-teste göre daha yüksek seviyeye çıktığı belirlenmiştir. Ön testte düşük sayıda özelliğe sahip olan öğrenci sayısı son testte giderek azalmıştır. Deney-1 grubundaki öğrencilerin %51'i, Deney-2 grubundaki %40'ı, Kontrol grubu öğrencilerinin ise %29'u son testte ön teste göre daha yüksek bir düzeydedir. Diğer bir anlatımla projede kullanılan sorular öğrencilerin yeterlik alanlarının iyileşmesine katkıda bulunmuş ve deneysel süreç öğrencilerin daha yüksek yeterlik düzeylerine geçmesinde etkili olmuştur.

Bu bulgular göz önüne alındığında proje deneysel sürecinde uygulanan işlemin öğrencilerin PISA test sonuçlarında en az %10'luk bir seviye artışına yol açabileceği söylenebilir. Bu rakam PISA araştırmaları(OECD, 2015a) göz önüne alındığında dikkate değer bir artışı temsil ettiği söylenebilir.

Projenin deneysel sürecinin önemli bir bölümünü oluşturan ve özgün değerine katkıda bulunan BTM yaklaşımları ile belirlenen öğrenci örtük sınıflarına ilişkin bulgular incelendiğinde Deney-1 grubunun %51 düzeyinde ön testte göre son testte daha yüksek bir örtük sınıfa atandığı görülmüştür. Bu yüzde Deney-2 için %40, Kontrol grubu için ise %29 olarak hesaplanmıştır.

Bu iki bulgu göz önüne alındığında Deney sürecinin öğrencilerin üst düzey düşünme becerilerini önemli düzeyde arttırdığı anlaşılmaktadır. Aynı zamanda bu artışın öğrencilerin PISA yeterlik düzeylerinde de bir artışın göstergesi olduğu görülmektedir.

Süreçle ilgili bir diğer önemli bulgu ise BTM yaklaşımına uyumlu madde yazımı ve test geliştirme prosedürlerinin proje kapsamında başarı ile gerçekleştirildiğine ilişkin ölçüt dayanaklı kanıtlar sunmasıdır. Uygulanan testlerin ve yapılan analizlerin hem BTM hem de MTK ve KTK analizleri ile benzer sonuçlar üretmesi geliştirilen testlerin BTM açısından yeterli psikometrik özellikleri taşıdığına işaret etmektedir. Bu sonuç, Türkiye'de de birden çok niteliği aynı anda ölçebilecek üst düzey soruların yazılabildiğini göstermektedir.

5.3 Proje Sürecinde Ulaşılan Diğer Sonuçlar

Proje kapsamındaki 4. izleme çalışmasında öğrencilerin PISA performansları yanında, üstbilişsel izleme süreçleri de Tip-2 STK temelinde ölçülmüştür. Sonuçlar, test performansı yüksek olan öğrencilerin, doğru ve yanlış yanıtlarını ayırt etme becerisi bakımından, düşük test performansı gösteren öğrencilere göre daha yüksek olduğunu göstermiştir. Bu durum özetle, test sorularında yüksek performans (bilişsel beceri) sergilemenin, aynı zamanda başka tür yüksek bilişsel beceriyle (üstbilişsel izleme becerisi ile) de paralellik gösterebileceğine işaret etmektedir. Alanyazında bilişsel ve üstbilişsel süreçlerin farklı süreçler olduğu işaret edilmiş olsa da, çalışma bilişsel becerisi yüksek olan öğrencilerin doğru yanıtları verirken, doğru yanıtlarının doğru (yanlış yanıtlarının da yanlış olduğunu) konusundaki farkındalıklarının bilişsel performansı düşük olan öğrencilere göre daha yüksek olabileceğine işaret etmektedir. Çalışma, aynı zamanda sonsal bir değerlendirme içererek, öğrencilere test sonrasında testte ne kadar doğru yapmış olabileceklerine dair bir tahminde bulunmalarını da incelemiştir. Sonuçlar, sonsal kestirimlerin (test bitimi sonrası yapılan tahminin) yüksek performans sergileyen öğrencilerde gerçek puanlarından farklılaşmadığını göstermiştir. Ancak, düşük performans sergileyen öğrenciler hem ilksel (test öncesi tahmin) hem sonsal tahminlerinde gerçekte aldıkları puandan daha yüksek bir tahmin yapmışlardır. Özetle, düşük test performansı hem üstbilişsel izleme becerisi bakımından hem test puanı tahminleri yapma bakımından, yüksek test puanı alan öğrencilere göre düşük performans sergilemeleriyle paralel görünmektedir.

Projede, örneğin üstbilişsel izleme becerisi PISA testinden daha önce kullanılmamış olan bir yöntemle (Tip-2 Sinyal Tespit Kuramı'na ait yöntemle) incelenmiştir. Alanyazında Tip-2 STK yöntemi yalnızca SAT'de (Scholastic Aptitute Test) kullanılmış ve diğer olası alternatiflerine göre daha etkin hesaplamalar yapabildiği saptanmıştır (Higham, 2007). PISA ve diğer uluslararası geçerliği olan testlerde, test maddelerinin doğrudan farklı becerilerle ilişkilendirilerek bu becerileri ölçmelerinin yanında, farklı ve görece kolay şekilde yöntemlerle

de öğrenci performanslarının belirlenebilmesi mümkün görünmektedir. Örneğin, çoktan seçmeli ya da açık uçlu fark etmeksizin ne türde soru olursa olsun, öğrencilere soruya yanıt verme veya vermeme seçenekleri sunularak (serbest-raporlama), sonrasında da verilen yanıt tahmin de olsa verilen yanıtın ne derece doğru olduğunu düşündüklerini derecelendirmeleri, doğru ve yanlış yanıtlarını ne düzeyde ayırt edebildiklerini (üstbilişsel izleme becerisini) belirleyebilmemizi de sağlayabilmektedir. Özetle, bu becerinin de ölçümü PISA gibi uluslararası kullanılan bir testte üstbiliş becerisinin de sorunun kendisinden ziyade –ya da bunun yanı sıra- soruya verilen yanıt hakkındaki farkındalık düzeyinin ölçülmesi bakımından da önemli bir değerlendirme olduğu düşünülmektedir. Bununla birlikte, gelecekteki çalışmaların, kağıt kalem testleri yerine soru çözümü için ek kağıt ve kalemlerin verilmesinin yanında bilgisayar temelli olarak yapılmasının örneğin olası veri kayıplarını en aza indirebilmek için etkili bir yöntem olabileceği düşünülmektedir. Sonuçlar, üst bilişsel izleme performansları açısından öğrencilerin test performansı daha iyi olduğunda, doğru ve yanlış yanıtlarını saptama yeteneklerinin daha yüksek olduğunu göstermiştir. Düşük performans grubunun kendi test performansları ile ilgili olduğundan daha yüksek, Yüksek performans grubunun ise olduğundan daha düşük bir beklenti ile teste başladıkları, testi tamamladıktan sonra Düşük performans grubunun test performanslarını olduğundan çok yüksek, Yüksek grubun ise test performanslarını gerçek performanslarına çok yakın tahminledikleri görülmüştür.

Projede, soru yazarları ile hakemler arasında görüş alış verişinin etkin bir şekilde yürütülmesini sağlayan bir çevrimiçi editörlük sistemi geliştirilmiştir. Bu sistem, editörlük sürecinde yaşanan zorluklar göz önüne alarak değerlendirme sürecini iyileştirmek, hızlandırmak ve daha sonraki aşamalarda kullanılabilecek veriler toplamak amaçlarına hizmet etmiştir. Geliştirilen editörlük sistemini diğer modellerden ayıran en önemli özellikleri ise sürecin çevrimiçi olması, hakem yazar arasında anlık iletişime izin vermesi ve bütün süreç boyunca hakem yazar anonimliğini korumasıdır. Geliştirilen sistem sayesinde hakemler birbirinden bağımsız olarak soru yazarlarıyla değerlendirme sürecini yürütmüştür. Hakem – yazar arasındaki anlık bilgi ve doküman alışverişi süreci geleneksel soru değerlendirme yaklaşımından soruyu geliştirme anlayışına çevirmiştir(Basokcu & Duran, 2016). Bu süreç soru değerlendirme aşaması için alternatif bir yöntem uygulaması olmuştur.

Projede hem çoktan seçmeli hem de açık uçlu sorular birlikte kullanılmıştır. Projenin temel amaçlarından biri açık uçlu soruların genel sınavlarda kullanılabilmesi ve puanlanmasına ilişkin dönütler alabilmektir. Açık uçlu olarak tanımlanan sorular PISA ve TIMSS içeriğine benzer şekilde kısa cevaplı, uygulamalı ve doğru yanlış türü maddelerden oluşmuştur. Projede açık uçlu ve çoktan seçmeli maddelerin bir arada okutulabilmesi için özel bir optik form geliştirilmiştir. Öğrenciler sınav esnasında çoktan seçmeli maddeleri optik form üzerine

işaretlemişler, açık uçlu soruları ise kitapçıkta çözmüşlerdir. Öğrencilerin açık uçlu sorulara verdiği cevapların tamamının dijital olarak kaydedilmesi ve optik okuyucu tarafından bir seferde okunması sağlanmıştır. Bu şekilde elde edilen veri farklı analizler ile en doğru puanlama anahtarını oluşturmaya uygun hale getirilmiş, aynı zamanda da öğrencilerin öğrenme eksikleri genel hata ve yanılgılarını analiz edebilmeyi mümkün kılmıştır. Bu süreçte açık uçlu soruyu değerlendiren hakemin katılık düzeyi puanlama sürecinden arındırılmaya çalışılmıştır. Bunun için Muraki'nin puanlayıcı etkisi modeli ve Yao'nun puanlayıcı modeli kullanılmıştır (Patz et al., 2002). Böylelikle hakem katılığının puanlama üzerindeki etkisi belirlenmiştir (Basokcu & Gökçe, 2016).

Araştırmada geliştirilen izleme testlerinde hem çoktan seçmeli hem de yapılandırılmış yanıtlı maddeler birlikte kullanılmıştır. Aynı test içinde farklı madde formatlarının birlikte kullanıldığı geniş ölçekli sınav sayısı oldukça azdır (PISA, TIMSS, PEARLS). Bununla birlikte değişik madde formatlarının kullanıldığı ve BTM modellerine göre geliştirilmiş ve analiz edilmiş bir test uygulanması literatürde bulunmamaktadır. Projede yapılandırılmış yanıtlı maddelerin test içinde ve testten ayrı analizleri sonuçlarında karma madde formatı kullanılan PISA, TIMSS gibi testlerde BTM analizleri yapmanın parametre değişmezliği açısından herhangi bir soruna yol açmadığı görülmüştür. Bu durum doğru organize edilmiş testlerde farklı madde formatlarının doğru analiz yöntemleri ile değerlendirildiğinde testin psikometrik yapısını olumsuz yönde etkilemediğini göstermektedir.

Araştırmada Deney-1 ve Deney-2 grubu öğrencilerine izleme testleri sonunda sınav uygulaması ve sorulara ilişkin görüşlerine yönelik sorular sorulmuştur. Bu sorular "Açık uçlu soruları cevaplarken bildiklerimi daha iyi gösterebiliyorum", "Bu testteki sorulara benzer sorular ile daha sık karşılaşmak isterim", "Bu testteki sorular matematiğin günlük yaşamdaki önemini bana gösterdi" ve "Bu testte yer alan soruların öğrenme isteğimi arttırdığını düşünüyorum" şeklindedir. Analiz sonuçlarına göre Deney-1 grubu öğrencilerinin bu dört soruya da katılma düzeyleri uygulamalar ilerledikçe artmıştır. Özellikle son uygulamada Deney-1 grubu katılma düzeyleri Deney-2'ye istatistiksel olarak anlamlı düzeyde yüksek bulunmuştur. Bu durum geri bildirimin öğrencilerin görüşleri üzerinde de olumlu bir etkisi olduğunu göstermektedir (Basokcu & Ozel, 2018).

5.4 Öneriler

Proje sürecinde elde edilen deneyimler ve projenin amaçları doğrultusunda elde edilen sonuçlardan hareketle aşağıdaki öneriler getirilebilir.

Ulusal sınavlarda, uluslararası geniş ölçekli sınavların kapsamı ve bu sınavların test ettiği nitelikler, projenin hazırlık sürecindeki çalışmalar örnek alınarak hazırlanabilir. Böylelikle Türkiye'nin hem uluslararası sınavlardaki başarısı tahminlenebilir hem de öğrencilerin üst düzey düşünme ve öğrenme yeteneklerinin gelişmesine yönelik öğretim programı ve öğretim programı dışı çalışmalar tetiklenebilir.

Ulusal sınavlarda hem çoktan seçmeli hem de açık uçlu sorular birlikte kullanılabilir. Açık uçlu sorular PISA ve TIMSS içeriğine benzer şekilde kısa cevaplı, uygulamalı ve doğru yanlış türü maddelerden oluşabilir. Açık uçlu ve çoktan seçmeli maddelerin bir arada okutulabilmesi için bu projede olduğu gibi özel optik formlar geliştirilebilir ve cevapların tamamı dijital olarak kaydedilebilir. Bu süreçte açık uçlu soruyu değerlendiren hakemlerin katılık düzeyi puanlama sürecinden arındırabilecek modeller de kullanılabilir. Değişik madde formatlarının kullanıldığı ve BTM modellerine göre geliştirilen ve analiz edilen testlerin geliştirilmesinde projede gerçekleştirilen çalışmalardan yararlanılabilir ve ortaya çıkan sonuçlar kullanılabilir.

Ulusal sınav sonuçlarını öğrencilere bildirmede, öğrencilerin sadece başarı puanları raporlanmamalıdır. Bu projede gruplar arasındaki temel manipülasyon geri bildirim üzerinden kurgulandığından ve başarı elde edildiğinden, öğrencilerin öğrenme eksikleri, genel hata ve yanılgılarını içeren daha ayrıntılı bir geribildirim yaklaşımının kullanılması eğitim sisteminin program ve uygulamalarını değerlendirebilmesi için de bilgi sağlayabilir.

Ulusal düzey soru yazarları ve program uygulamalarında öğretmenler, birden çok niteliği aynı anda ölçebilecek üst düzey düşünme becerilerini içeren sorular sormaları için cesaretlendirilebilir, bu konuda yeterlik kazanmaları için mesleki gelişim programları düzenlenebilir. Böylelikle, öğrencilerin uluslararası geniş ölçekli sınavlardaki yeterlik düzeyleri yükseltilebilir.

Öğrencilerin sahip oldukları özellik sayısı arttıkça üst bilişsel izleme becerisinde de bir artış gözlendiğinden öğretim süreçlerinde öğrencilerin özellikle akıl yürütme ve strateji geliştirmede daha yetkin hale gelmelerini sağlayacak program faaliyetlerine yer verilebilir. Sonuçlara göre akıl yürütme ve strateji geliştirme üstbilişsel izleme üzerinde etkili olduğundan, okullarda özellikle üstbilişsel strateji öğretimi ve üstbilişsel izlemeyi geliştiren sosyal ortamların yaratılması önerilebilir. Bu uygulamalarda alana özgü bilgiler yanında öğrencinin kendisiyle ilgili bilgisi içerik olarak kullanılabilir.

6 KAYNAKÇA

Aydın, E., & Önder, O. (2010). Sınava Hazırlık Biçiminin Farklı Sınav Türlerinde Ölçülen Matematik Sınav Başarı Düzeylerine Etkisi. *M.Ü. Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, (31), 5–24.

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review*. https://doi.org/10.1037/h0029099

Barahal, S. L. (2008). Thinking about thinking: Preservice teachers strengthen their thinking artfully. *Phi Delta Kappan*. https://doi.org/10.1177/003172170809000412

Basokcu, T. O. (2010). Öğrenme Eksiklerinin Belirlenmesinde Klasik Test Teorisine Dayalı Yöntemler ve DINA Modelin Karşılaştırılması TheComparison of DINA Model andClassical Test Theory in Determiningthe Learning Gaps. *Ege Eğitim Dergisi*, *1*(1), 59–83. https://doi.org/10.1016/B978-1-59749-661-2.00010-3

Basokcu, T. O., & Canpolat, A. (2016). Study in Local Equating of Cognitively Diagnostic Modeled Observed Scores with Anchor-test Design. In *The Association of Measurement and Evaluation in Education and Psychology* (pp. 26–34).

Basokcu, T. O., Canpolat, A., & Gönülal, G. (2016). DINA Modelde Madde Doygunluğunun ve Madde Uyum Düzeyinin Model Veri Uyumuna Etkisi. In *IIIrd International Eurasian Educational Research Congress* (pp. 1593–1594). Mugla: Anı Yayıncılık.

Basokcu, T. O., & Ceylan, S. (2017). Generation Study of PISA Maths Proficiency Levels in Turkish 6th Grade Students. In M. Shelley, M. Pehlivan, & M. T. Hebeci (Eds.), *International Conference on Education in Mathematics, Science & Technology ICEMST* (pp. 14–23). Kuşadası: ISRES Publishing.

Basokcu, T. O., & Ceylan, S. (2018). Yapılandırılmış ve Çoktan Seçmeli Maddelerde Parametre Değişmezliğinin İncelenmesi: Bilişsel Tanı Modellerinde Madde Formatının Etkisi. In *International Cogress on Measurement and Evaluation in Education and Psychology* (p. 389). Prizen/Kosovo: Cmeep.

Basokcu, T. O., & Duran, N. (2016). Çevrimiçi Editörlük Sisteminin Test Geliştirme Süreçlerine Uyarlanması. In *3rd. International Conference on New Trends in Education*.

Basokcu, T. O., & Gökçe, S. (2016). Hakem Katılığının YapılandırılmıĢ-Yanıtlı Maddelerde Yetenek ve Parametre Kestirimine Etkisi. In *V. Ululsal Egitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi* (p. 237). Antalya: EPOD.

Basokcu, T. O., & Güngör, D. (2016). Comparison of DINA Model and Exploratory Latent

Class Analysis. In *International Meeting of the Psychometric Society (IMPS)* (p. 75). Asheville, NC: Psychometric Society.

Basokcu, T. O., & Güzel, M. A. (2017). A Cognitive Diagnostic Model: High-order Thinking and Metacognition involved in Mathematics. In *EARLI 17th Biennial Conference European Assocition for Research on Learning and Instruction* (p. 57). Tampere/Finlandiya.

Basokcu, T. O., & Güzel, M. A. (2018). Beyond Counting the Correct Responses: Metacognitive Monitoring and Estimations about Test Scores. In *8th International Biennial Conference of EARLI SIG 16 Metacognition* (p. 73). Zurich.

Basokcu, T. O., & Ozel, S. (2018). Longitudinal Analysis of The Students' Attitudes Towards Real Life Problems Testing Their High-Level Thinking Skills in Mathematics. In *International Conference on Education in Mathematics, Science & Technology (ICEMST2018)* (p. 175). Muğla.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. D. McKay.

Boaler, J. (1998). Open and Closed Mathematics: Student Experiences and Understandings. *Journal for Research in Mathematics Education*, *29*(1), 41–62. https://doi.org/10.2307/749717

Bridgeman, B. (1992). A Comparison of Quantitative Questions in Open-Ended and Multiple-Choice Formats. *Journal of Educational Measurement*. https://doi.org/10.1111/j.1745-3984.1992.tb00377.x

Brookhart, S. M. (1994). Teachers' Grading: Practice and Theory. *Applied Measurement in Education*, *7*(4), 279–301. https://doi.org/10.1207/s15324818ame0704_2

Brookhart, S. M. (2010). *How to Assess Higher-order Thinking Skills in Your Classroom*. ASCD.

Bümen, N. T. (2006). Program Geliştirmede Bir Dönüm Noktası: Yenilenmiş Bloom Taksonomisi. *Eğitim ve Bilim*, *32*(142).

Buxton, L. (1981). *Do You Panic about Maths?: Coping with Maths Anxiety*. Heinemann Educational.

Büyüköztürk, Ş. (2010). *Sosyal bilimler için veri analizi el kitabı: istatistik, araştırma deseni, SPSS uygulamaları ve yorum*. Pegem Akademi.

Carrol, T. M. (1989). Critical thinking : Promoting it in the Classroom. *ERIC-Education*

Resources Information Center, (ERIC Document Reproduction Service No: ED306554)., 0–3.

Collins, L. M., & Lanza, S. T. (2013). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Wiley. Retrieved from https://books.google.com.tr/books?id=gPJQWKsgh3YC

Collins, R. (2014). Skills for the 21st Century: teaching higher-order thinking. *Curriculum & Leadership Journal*, *12*(14).

Conklin, W. (2012). *Strategies for Developing Higher-Order Thinking Skills*. Shell Educational Publishing.

Council, N. R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. (J. W. Pellegrino, N. Chudowsky, & R. Glaser, Eds.). Washington, DC: The National Academies Press. https://doi.org/10.17226/10019

Dayton, C. M., & Macready, G. B. (2006). 13 Latent Class Analysis in Psychometrics. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. Volume 26, pp. 421–446). Elsevier. https://doi.org/http://dx.doi.org/10.1016/S0169-7161(06)26013-9

de la Torre, J. (2011). The Generalized DINA Model Framework. *Psychometrika*, *76*(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7

de la Torre, J. (2008). An Empirically Based Method of Q-Matrix Validation for the DINA Model: Development and Applications. *Journal of Educational Measurement*, *45*(4), 343–362. https://doi.org/10.1111/j.1745-3984.2008.00069.x

de la Torre, J. (2009). A Cognitive Diagnosis Model for Cognitively Based Multiple-Choice Options. *Applied Psychological Measurement*, *33*(3), 163–183. https://doi.org/10.1177/0146621608320523

de la Torre, J., & Chiu, C.-Y. (2016). A General Method of Empirical Q-matrix Validation. *Psychometrika*, *81*(2), 253–273. https://doi.org/10.1007/s11336-015-9467-8

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333–353. https://doi.org/10.1007/BF02295640

de la Torre, J., & Douglas, J. (2008). Model Evaluation and Multiple Strategies in Cognitive Diagnosis: An Analysis of Fraction Subtraction Data. *Psychometrika*, *73*(4), 595–624. https://doi.org/10.1007/s11336-008-9063-2

de la Torre, J., & Lee, Y.-S. (2010). A Note on the Invariance of the DINA Model Parameters. *Journal of Educational Measurement*, *47*(1), 115–127. https://doi.org/10.1111/j.1745-

3984.2009.00102.x

DeCarlo, L. T. (2011). On the Analysis of Fraction Subtraction Data: The DINA Model, Classification, Latent Class Sizes, and the Q-Matrix. *Applied Psychological Measurement*, *35*(1), 8–26. https://doi.org/10.1177/0146621610377081

Edition, S. (2009). *PISA Data Analysis Manual: SPSS, Second Edition*. *Analysis*. https://doi.org/10.1787/9789264056275-en

Embretson, S. E. (1999). Cognitive psychology applied to testing. In F. T. Durso (Ed.), *Handbook of applied cognition* (p. xxii, 881-xxii, 881). New York, NY, US: John Wiley & Sons Ltd.

Embretson, S. E., & Reise, S. P. (2013). *Item Response Theory*. Taylor & Francis. Retrieved from https://books.google.com.tr/books?id=AcWQtfGjCawC

Ennis, R. H., & Norris, S. P. (1990). Critical thinking assessment: Status, issues, needs. *Cognitive Assessment of Language and Math Outcomes Legg*. https://doi.org/10.1371/journal.pone.0075214

Ersözlü, Z. N., & Kazu, H. (2011). İlköğretim Beşinci Sınıf Sosyal Bilgiler Dersinde Uygulanan Yansıtıcı Düşünmeyi Geliştirme Etkinliklerinin Akademik Başarıya Etkisi. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, *24*(1), 141–159.

Fischbach, A., Keller, U., Preckel, F., & Brunner, M. (2013). PISA proficiency scores predict educational outcomes. *Learning and Individual Differences*, *24*, 63–72. https://doi.org/10.1016/j.lindif.2012.10.012

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *37*(6), 359–374. https://doi.org/10.1016/0001-6918(73)90003-6

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*. https://doi.org/10.1037/0003-066X.34.10.906

Foong, P. Y. (2000). Open-ended problems for higher-order thinking in mathematics. *Teaching and Learning*, *20*(2), 49–57.

Frederiksen, N. (1984). Implications of Cognitive Theory for Instruction in Problem Solving. *Review of Educational Research*. https://doi.org/10.3102/00346543054003363

Garofalo, J. (2018). Beliefs, Responses, and Mathematics Education: Observations From the Back of the Classroom. *School Science and Mathematics*, *89*(6), 451–455. https://doi.org/10.1111/j.1949-8594.1989.tb11947.x

Goodson, L., & Rohani, F. (1998). Higher order thinking skills: Definition, teaching strategies, assessment. *Center for Advancement of Learning and Assessment*. https://doi.org/10.1063/1.3155135

Gough, D. (1991). *Thinking About Thinking*. Alexandria, VA.: National Association of Elementary School Principals. https://doi.org/10.1080/0266736910060407

Green, D. G., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. *Wiley & Sons, Inc.* https://doi.org/10.1901/jeab.1969.12-475

Greenwood, J. (1984). My Anxieties About Math Anxiety. *The Mathematics Teacher*, *77*(9), 662–663.

Haertel, E. H. (1989). Using Restricted Latent Class Models to Map the Skill Structure of Achievement Items. *Journal of Educational Measurement*, *26*(4), 301–321. https://doi.org/10.1111/j.1745-3984.1989.tb00336.x

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and Validating Test Items*. Taylor & Francis.

Hambleton, R. K., & Swaminathan, H. (2013). *Item Response Theory: Principles and Applications*. Springer Netherlands.

Haney, W., & Madaus, G. (1989). Searching for Alternatives to Standardized Tests: Whys, Whats, and Whithers. *Phi Delta Kappan*.

Hartz, S. M. C. (2002). *A Bayesian Framework for the Unified Model for Assessing Cognitive Abilities: Blending Theory With Practicality*. Retrieved from https://books.google.com.tr/books?id=5zomjwEACAAJ

Henningsen, M., & Stein, M. K. (1997). Mathematical Tasks and Student Cognition: Classroom-Based Factors That Support and Inhibit High-Level Mathematical Thinking and Reasoning. *Journal for Research in Mathematics Education*, *28*(5), 524–549. https://doi.org/10.2307/749690

Henson, R., & Douglas, J. (2005). Test Construction for Cognitive Diagnosis. *Applied Psychological Measurement*, *29*(4), 262–277. https://doi.org/10.1177/0146621604272623

Henson, R., Roussos, L., Douglas, J., & Xuming He. (2008). Cognitive Diagnostic Attribute-Level Discrimination Indices. *Applied Psychological Measurement*, *32*(4), 275–288. https://doi.org/10.1177/0146621607302478

Herman, J. L., Klein, D. C. D., & Wakai, S. T. (1997). American Students' Perspectives on

Alternative Assessment: do they know it's different? *Assessment in Education: Principles, Policy & Practice*, *4*(3), 339–352. https://doi.org/10.1080/0969594970040302

Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory and Cognition*. https://doi.org/10.3758/BF03195266

Higham, P. A. (2007). No special K! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.136.1.1

Higham, P. A., & Arnold, M. M. (2007). How many questions should I answer? Using bias profiles to estimate optimal bias and maximum score on formula-scored tests. *European Journal of Cognitive Psychology*. https://doi.org/10.1080/09541440701326121

Higham, P. A., & Tam, H. (2005). Generation failure: Estimating metacognition in cued recall. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2005.01.015

Hirose, S. (1992). Critical Thinking in Community Colleges. *ERIC-Education Resources Information Center, (ERIC Document Reproduction Service No: 348128.*, 1–7.

Holland, P. W., Dorans, N. J., & Petersen, N. S. (2006). 6 Equating Test Scores. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. Volume 26, pp. 169–203). Elsevier. https://doi.org/http://dx.doi.org/10.1016/S0169-7161(06)26006-1

Jaworski, B. (2006). Theory and Practice in Mathematics Teaching Development: Critical Inquiry as a Mode of Learning in Teaching. *Journal of Mathematics Teacher Education*, *9*(2), 187–211. https://doi.org/10.1007/s10857-005-1223-z

Kato, K. (2009). *Improving efficiency of cognitive diagnosis by using diagnostic items and adaptive testing*. University of Minnesota.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Normal (Ed.), *Models of Human Memory* (pp. 333–370). New York, NY, US: Academic Press.

Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices*. Springer. Retrieved from https://books.google.com.tr/books?id=tusOwjb7LWsC

Koray, Ö., Altunçekiç, A., & Yaman, S. (2005). Fen Bilgisi Öğretmenlerinin Soru Sorma Becerilerinin Bloom Taksonomisine Göre Değerlendirilmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, *17*(17), 33–39.

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into*

Practice, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Krosnick, J. A., & Presser, S. (2010). Question and Questionnaire Design. In P. V Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (Vol. 2, pp. 263–314). Bingley, UK: Emerald Group Publishing Limited. https://doi.org/10.1111/j.1432-1033.1976.tb10115.x

Lave, J., Murtaugh, M., & de la Rocha, O. (1984). The dialectic of arithmetic in grocery shopping. In *Everyday cognition: Its development in social context.* (pp. 67–94). Cambridge, MA, US: Harvard University Press.

Leech, N., Barrett, K., & Morgan, G. A. (2007). *SPSS for Intermediate Statistics: Use and Interpretation, Third Edition*. Taylor & Francis.

Leighton, J., & Gierl, M. (2007). *Cognitive Diagnostic Assessment for Education: Theory and Applications*. Cambridge University Press. Retrieved from http://books.google.com.tr/books?id=E13E9lOFUMYC

Lenzner, T., Neuert, C., & Otto, W. (2016). Cognitive Pretesting. In *GESIS Survey Guidelines*. Mannheim, Germeny: GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_010

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory Into Practice*, *32*(3), 131–137. https://doi.org/10.1080/00405849309543588

Libman, Z. (2010). Integrating Real-Life Data Analysis in Teaching Descriptive Statistics: A Constructivist Approach. *Journal of Statistics Education*, *18*(1). https://doi.org/10.1080/10691898.2010.11889477

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2006). *Methods in Educational Research: From Theory to Practice*. Wiley. Retrieved from https://books.google.com.tr/books?id=G9D81mh9xCAC

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Erlbaum Associates. Retrieved from http://books.google.com.tr/books?id=7YhqAAAAMAAJ

Lord, T., & Baviskar, S. (2007). Moving Students From Information Recitation to Information Understanding: Exploiting Bloom's Taxonomy in Creating Science Questions. *Journal of College Science Teaching*, *36*(5), 40–44.

Luce, R. D., & Krumhansl, C. L. (1988). Measurement, scaling, and psychophysics. In *Stevens' handbook of experimental psychology: Perception and motivation; Learning and cognition, Vols. 1-2, 2nd ed.* (pp. 3–74). Oxford, England: John Wiley & Sons.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the Relative Value of Multiple-Choice, Constructed Response, and Examinee-Selected Items on Two Achievement Tests. *Journal of Educational Measurement*. https://doi.org/10.1111/j.1745-3984.1994.tb00445.x

Maier, N. R. F. (1937). Reasoning in rats and human beings. *Psychological Review*. https://doi.org/10.1037/h0062900

MAIER, N. R. F. (1933). AN ASPECT OF HUMAN REASONING. *British Journal of Psychology. General Section*. https://doi.org/10.1111/j.2044-8295.1933.tb00692.x

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*(2), 187–212. https://doi.org/10.1007/bf02294535

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*. https://doi.org/10.1207/s15326985ep3404_2

Marzano, R. J. (1998). *A Theory-Based Meta-Analysis of Research on Instruction*. Aurora, Colorado: Mid-continent Regional Educational Laboratory.

McCutcheon, A. L. (1987). *Latent Class Analysis*. SAGE Publications.

McDonald, R. P. (2013). *Test Theory: A Unified Treatment*. Taylor & Francis.

McKeachie, W. J. (1987a). Can evaluating instruction improve teaching? *New Directions for Teaching and Learning*. https://doi.org/10.1002/tl.37219873103

McKeachie, W. J. (1987b). Cognitive skills and their transfer: Discussion. *International Journal of Educational Research*. https://doi.org/10.1016/0883-0355(87)90010-3

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics.* (pp. 575–596). New York, NY, England: Macmillan Publishing Co, Inc.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist*. https://doi.org/10.1037//0003-066X.50.9.741

Miri, B., David, B.-C., & Uri, Z. (2007). Purposely Teaching for the Promotion of Higher-order Thinking Skills: A Case of Critical Thinking. *Research in Science Education*, *37*(4), 353–369. https://doi.org/10.1007/s11165-006-9029-2

Mislevy, R. J., & Levy, R. (2006). 26 Bayesian Psychometric Modeling From An Evidence-Centered Design Perspective. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. Volume 26, pp. 839–865). Elsevier. https://doi.org/http://dx.doi.org/10.1016/S0169-

7161(06)26026-7

Mullis, I. V. S., Martin, M. O. (Eds. . (2013). *TIMSS 2015 Assessment Frameworks*. *TIMSS 2015 Assessment Frameworks*. https://doi.org/10.1108/AJEMS-03-2012-0017

Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: principles and applications*. Pearson/Prentice Hall.

Naveh-Benjamin, M., McKeachie, W. J., Lin, Y. G., & Tucker, D. G. (1986). Inferring Students' Cognitive Structures and Their Development Using the "Ordered Tree Technique." *Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.78.2.130

Nichols, P. D. (1994). A Framework for Developing Cognitively Diagnostic Assessments. *Review of Educational Research*, *64*(4), 575–603. https://doi.org/10.3102/00346543064004575

Nickerson, R. S. (1989). New Directions in Educational Assessment. *Educational Researcher*, *18*(9), 3–7. https://doi.org/10.3102/0013189X018009003

Nitko, A. J., & Brookhart, S. M. (2007). *Educational Assessment of Students*. Pearson Merrill Prentice Hall.

Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics.* New York, NY, US: Cambridge University Press.

Nuthall, G., & Alton-Lee, A. (1995). Assessing Classroom Learning: How Students Use Their Knowledge and Experience to Answer Classroom Achievement Test Questions in Science and Social Studies. *American Educational Research Journal*. https://doi.org/10.3102/00028312032001185

O'Neil Jr., H. F., & Brown, R. S. (1998). Differential Effects of Question Formats in Math Assessment on Metacognition and Affect. *Applied Measurement in Education*, *11*(4), 331–351. https://doi.org/10.1207/s15324818ame1104_3

Oecd. (2013). PISA 2015 Draft Mathematics Framework. *Oecd*, (March 2013), 52. https://doi.org/10.1177/0022146512469014

OECD. (n.d.). *PISA 2012 Assessment and Analytical Framework*. Retrieved from file:///content/book/9789264190511-en

OECD. (2009a). *PISA Data Analysis Manual: SPSS*. *Oecd*. https://doi.org/10.1787/9789264056275-en

OECD. (2009b). *PISA PISA Data Analysis Manual: SAS, Second Edition*. OECD Publishing.

Retrieved from https://books.google.com.tr/books?id=idUjxoSfnZIC

OECD. (2015a). *PISA 2015: Results in focus*. *Pisa 2015*. https://doi.org/10.1787/9789264266490-en

OECD. (2015b). *PISA 2015 Assessment and Analytical Framework*. Retrieved from file:///content/book/9789264255425-en

Oulte, D. (2011). *Sampling Methods*. GRIN Verlag.

Patz, R. J., Junker, B. W., Johnson, M. S., & Mariano, L. T. (2002). The Hierarchical Rater Model for Rated Test Items and Its Application to Large-Scale Educational Assessment Data. *Journal of Educational and Behavioral Statistics*, *27*(4), 341–384. Retrieved from http://www.jstor.org/stable/3648122

Popham, W. J. (2003). *Test Better, Teach Better: The Instructional Role of Assessment*. Association for Supervision and Curriculum Development.

Porst, R. (2011). *Fragebogen: Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Prawat, R. S. (1996). Constructivisms, modern and postmodern. *Educational Psychologist*, *31*(3–4), 215–225. https://doi.org/10.1207/s15326985ep3103&4_6

Reckase, M. D. (2006). 18 Multidimensional Item Response Theory. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. Volume 26, pp. 607–642). Elsevier. https://doi.org/http://dx.doi.org/10.1016/S0169-7161(06)26018-8

Roussos, L. A., Stout, W., DiBello, L. V, Hartz, S. M., Templin, J. L., & Henson, R. A. (2007). The Fusion Model Skills Diagnosis System. In J. P. L. M. J. Gierl (Ed.), *Cognitive Diagnostic Assessment for Education Theory and Applications* (pp. 275–318). New York: Cambridge University Press.

Rupp, A. A., & Mislevy, R. J. (2007). Cognitive psychology as it applies to diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment in education: Theory and practice.* New York, NY, US: Cambridge University Press. https://doi.org/http://dx.doi.org/10.1017/CBO9780511611186

Smith, M. S., & Stein, M. K. (1998). Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School*.

Sousa, D. A. (2008). *How the Brain Learns Mathematics*. SAGE Publications.

Stenmark, J. K., EQUALS., & Committee, C. M. C. C. for M. A. (1989). *Assessment Alternatives in Mathematics: An Overview of Assessment Techniques that Promote*

Learning. EQUALS. Retrieved from https://books.google.com.tr/books?id=9QOOq6V2HF8C

Stodolsky, S. S. (1985). Telling Math: Origins of Math Aversion and Anxiety. *Educational Psychologist*, *20*(3), 125–133. https://doi.org/10.1207/s15326985ep2003_2

Stout, W. (2002). Psychometrics: From practice to theory and back. *Psychometrika*, *67*(4), 485–518. https://doi.org/10.1007/BF02295128

Tatsuoka, K. K. (1983). Rule Space: An Approach for Dealing with Misconceptions Based on Item Response Theory. *Journal of Educational Measurement*, *20*(4), 345–354. Retrieved from http://www.jstor.org/stable/1434951

Tatsuoka, K. K. (1995). Architecture of Knowledge Structures and Cognitive. In R. L. B. Paul D. Nichols, Susan F. Chipman (Ed.), *Cognitively Diagnostic Assessment*. L. Erlbaum. Retrieved from https://books.google.com.tr/books?id=ZnPUnp_nNy0C

Tatsuoka, K. K., & Linn, R. L. (1983). Indices for Detecting Unusual Patterns: Links Between Two General Approaches and Potential Applications. *Applied Psychological Measurement*, *7*(1), 81–96. https://doi.org/10.1177/014662168300700111

The White House. (2003). An Overview of No Child Left Behind Act. Retrieved from http://www.whitehouse.gov/news/reports/no-child-left-behind.html

Thompson, S. K. (2012). *Sampling*. Wiley. Retrieved from https://books.google.com.tr/books?id=-sFtXLIdDiIC

Traub, R. E., & Rowley, G. L. (2018). Understanding Reliability. *Educational Measurement: Issues and Practice*, *10*(1), 37–45. https://doi.org/10.1111/j.1745-3992.1991.tb00183.x

van der Linden, W. J., & Wiberg, M. (2010). Local Observed-Score Equating With Anchor-Test Designs. *Applied Psychological Measurement*, *34*(8), 620–640. https://doi.org/10.1177/0146621609349803

Walstad, W. B., & Becker, W. E. (1994). Achievement Differences on Multiple-Choice and Essay Tests in Economics. *The American Economic Review*.

Watkins, M. J., & Gardiner, J. M. (1979). An appreciation of generate-recognize theory of recall. *Journal of Verbal Learning and Verbal Behavior*. https://doi.org/10.1016/S0022-5371(79)90397-9

Xin, T., & Zhang, J. (2015). Local Equating of Cognitively Diagnostic Modeled Observed Scores. *Applied Psychological Measurement*, *39*(1), 44–61. https://doi.org/10.1177/0146621614542427

Zohar, A., Degani, A., & Vaaknin, E. (2001). Teachers' beliefs about low-achieving students and higher order thinking. *Teaching and Teacher Education*, *17*(4), 469–485. https://doi.org/https://doi.org/10.1016/S0742-051X(01)00007-5

Zoller, U. (1993). Are lecture and learning compatible? Maybe for LOCS: Unlikely for HOCS. *Journal of Chemical Education*. https://doi.org/10.1021/ed070p195

Zoller, U. (1999). Scaling-up of higher-order cognitive skills-oriented college chemistry teaching: An action-oriented research. *Journal of Research in Science Teaching*. https://doi.org/10.1002/(SICI)1098-2736(199905)36:5<583::AID-TEA5>3.0.CO;2-M

Zoller, U., & Pushkin, D. (2007). Matching Higher-Order Cognitive Skills (HOGS) promotion goals with problem-based laboratory practice in a freshman organic chemistry course. *Chemistry Education Research and Practice*. https://doi.org/10.1039/B6RP90028C

Zoller, U., & Tsaparlis, G. (1997). Higher and lower-order cognitive skills: The case of chemistry. *Research in Science Education*. https://doi.org/10.1007/BF02463036

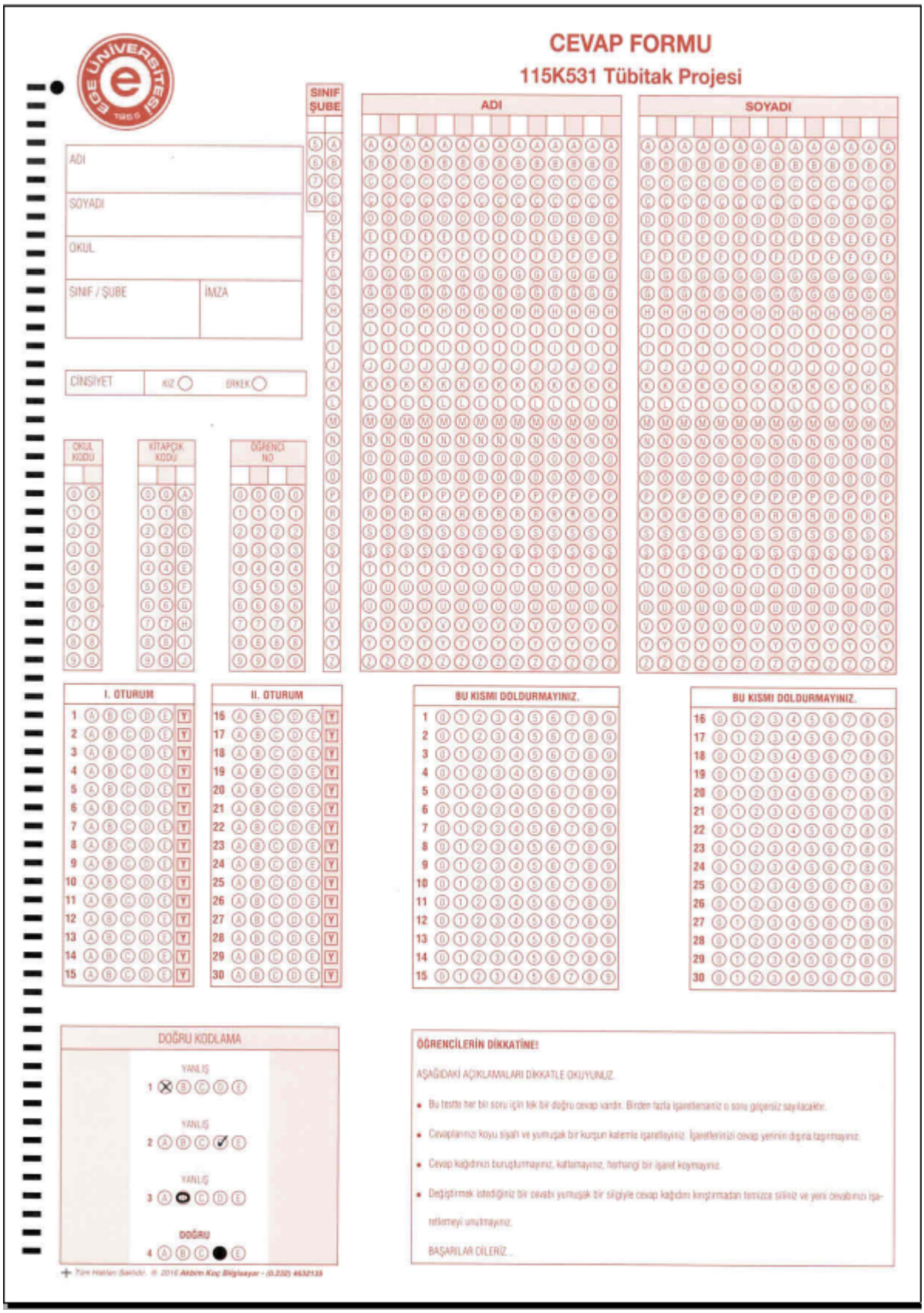
Züll, C. (2016). Open-Ended Questions. In *GESIS Survey Guidelines*. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_002

EKLER

EK 1. Soru Değerlendirme Formu

	DEĞERLENDİRME ÖLÇÜTLERİ	0	1	2	3	4
1	Soruda tanımlanan kazanımlar soruyla uyumludur.					
2	Sorunun çözümü birden çok matematiksel yeterlilik/davranış/beceri gerektirmektedir.					
3	Sorunun anlatım dili 6. sınıf öğrencileri için açık ve anlaşılırdır.					
4	Soru üst düzey düşünme becerilerini ölçebilecek niteliktedir.					
5	Soru kültürel, coğrafi ve etnik yanlılık içermemektedir.					
6	Tercih edilen soru formatı (kısa cevaplı- açık uçlu vs.) uygundur.					
7	Soru sadece ezber bilgisi kullanarak çözülebilir.					
8	Soru PISA ve TIMSS sorularına benzer niteliktedir.					
9	Soru proje kapsamında ölçülmek istenen ilgili kazanımları ölçebilecek niteliktedir.					
10	Soru bağlamı gerçek yaşam problemlerine yönelik hazırlanmıştır.					
11	Soruda kullanılan görsel materyaller soru için gereklidir.					
12	Soruda kullanılan senaryo/tablo/resim/şekil vb. uygundur.					
13	Soru bağlamı matematiksel becerileri kullanmaya uygundur.					
14	Soru açık, belirgin ve cevaplanabilir niteliktedir.					
15	Soru özgündür.					

EK 2a. Proje Ön-Son Test Optik Formu



CEVAP FORMU

115K531 Tübitak Projesi

ADI

SOYADI

SINIF ŞUBE

OKUL

SINIF / ŞUBE

İMZA

CİNSİYET

KIZ

ERKEK

OKUL KODU

KİTAPÇIK KODU

ÖĞRENCİ NO

I. OTURUM

II. OTURUM

BU KISMI DOLDURMAYINIZ.

BU KISMI DOLDURMAYINIZ.

DOĞRU KODLAMA

YANLIŞ

YANLIŞ

YANLIŞ

DOĞRU

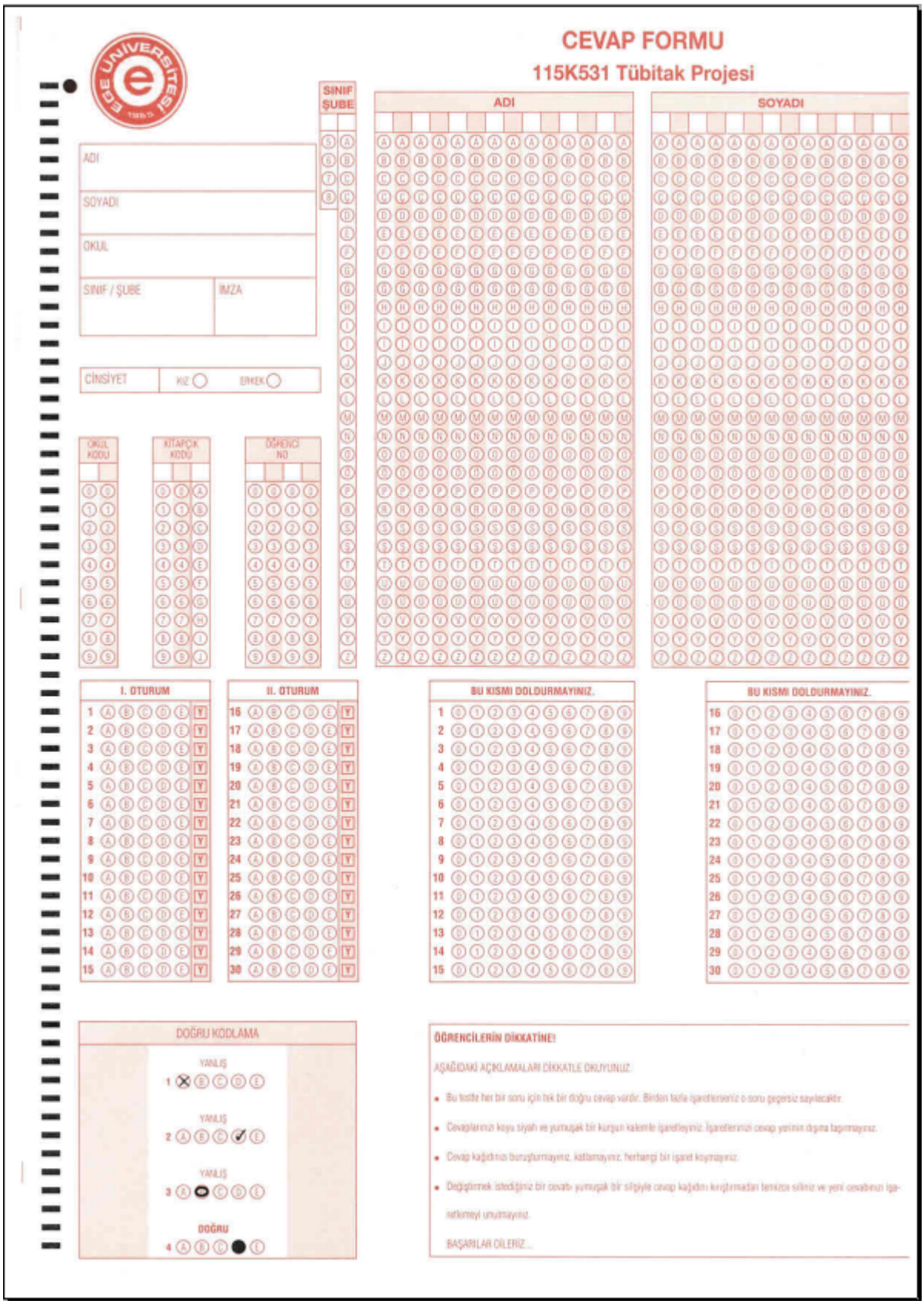
ÖĞRENCİLERİN DİKKATİNE!

AŞAĞIDAKİ AÇIKLAMALARI DİKKATLE OKUYUNUZ.

- Bu testte her bir soru için tek bir doğru cevap vardır. Birden fazla işaretlenseniz o soru geçersiz sayılacaktır.
- Cevaplarınızı koyu siyah ve yumuşak bir kurşun kalemle işaretleyiniz. İşaretlerinizi cevap yerinin dışına taşırmayınız.
- Cevap kağıdınızı buruşturmayınız, katlamayınız, herhangi bir işaret koymayınız.
- Değiştirmek istediğiniz bir cevabı yumuşak bir silgiyle cevap kağıdını kırıştırmadan temizce siliniz ve yeni cevabınızı işaretlemeyi unutmayınız.

BAŞARILAR DİLERİZ...

EK 2b. Proje İzleme Testleri Optik Formu



CEVAP FORMU

115K531 Tübitak Projesi

ADI

SOYADI

OKUL

SINIF / ŞUBE

İMZA

CİNSİYET

KIZ

ERKEK

SINIF ŞUBE

ADI

SOYADI

OKUL KODU

KİTAPÇIK KODU

ÖĞRENCİ NO

I. OTURUM

II. OTURUM

BU KISMI DOLDURMAYINIZ.

BU KISMI DOLDURMAYINIZ.

DOĞRU KODLAMA

YANLIŞ

YANLIŞ

YANLIŞ

DOĞRU

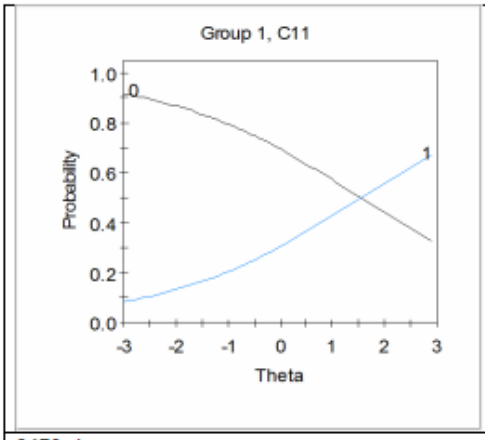
ÖĞRENCİLERİN DİKKATİNE!

AŞAĞIDAKİ AÇIKLAMALARI DİKKATLE OKUYUNUZ.

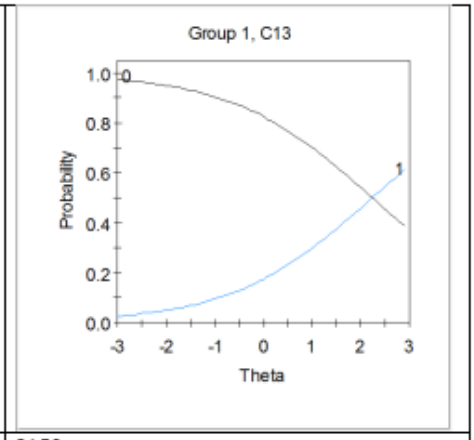
- Bu testte her bir soru için tek bir doğru cevap vardır. Birden fazla işaretlenseniz o soru geçersiz sayılacaktır.
- Cevaplarınızı koyu siyah ve yumuşak bir kurşun kalemle işaretleyiniz. İşaretlemenizi cevap yerinin dışına taşırmayınız.
- Cevap kağıdınızı buruşturmayınız, katlamayınız, herhangi bir işaret koymayınız.
- Değiştirmek istediğiniz bir cevabı yumuşak bir silgiyle cevap kağıdını kırıştırmadan temizce siliniz ve yeni cevabınızı işaretlemeyi unutmayınız.

BAŞARILAR DİLERİZ...

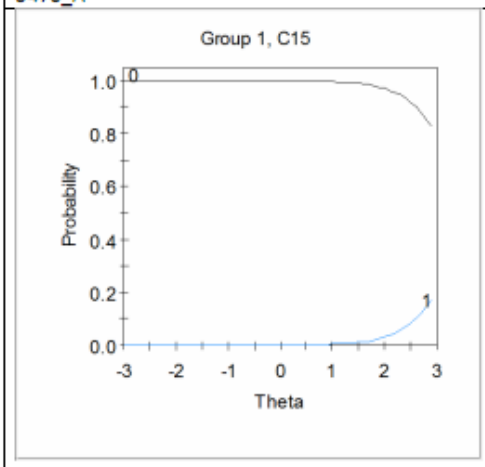
EK 3. Pilot Uygulama Madde Karakteristik Eğrileri



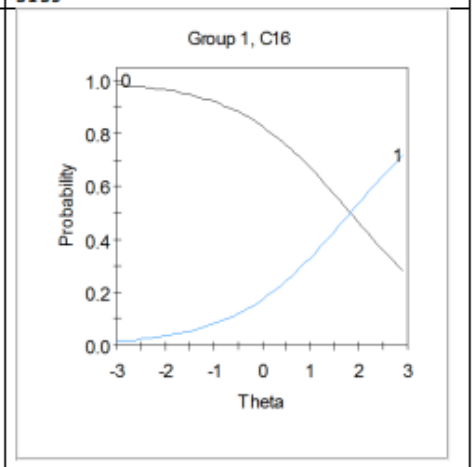
6470_A



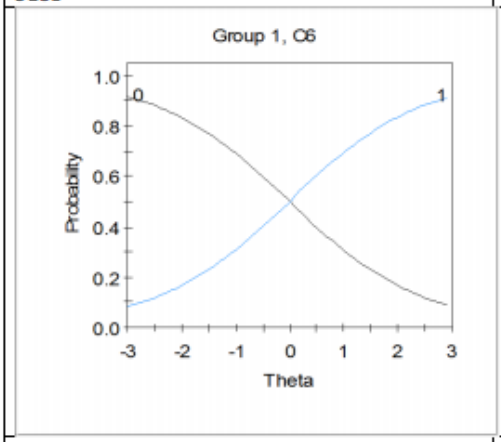
3159



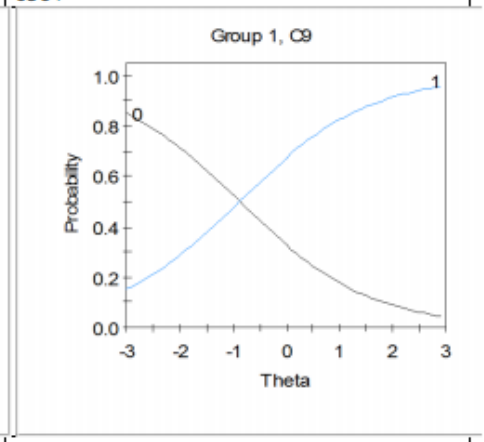
3281



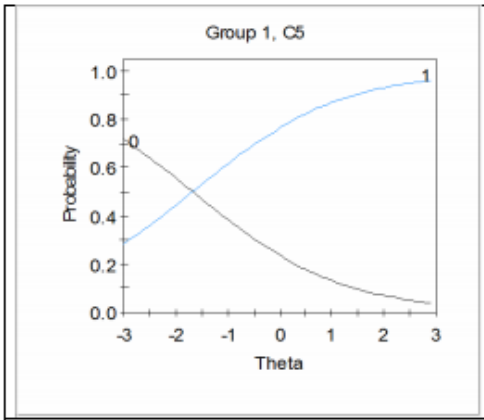
0504



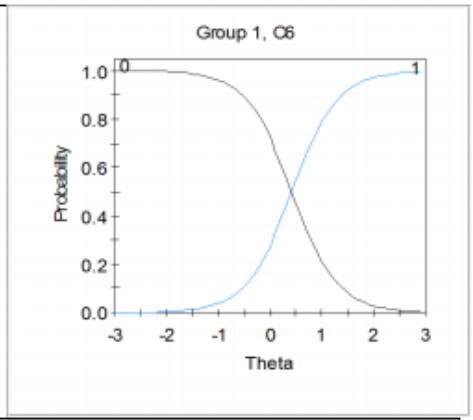
0205



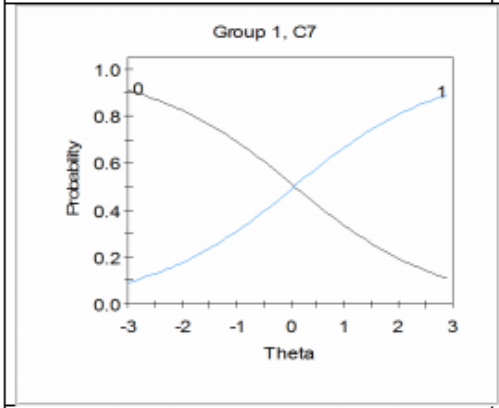
7305_A



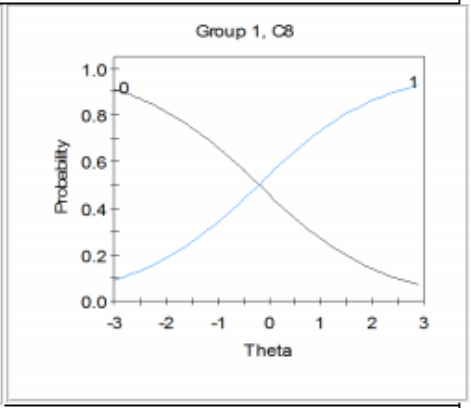
6188



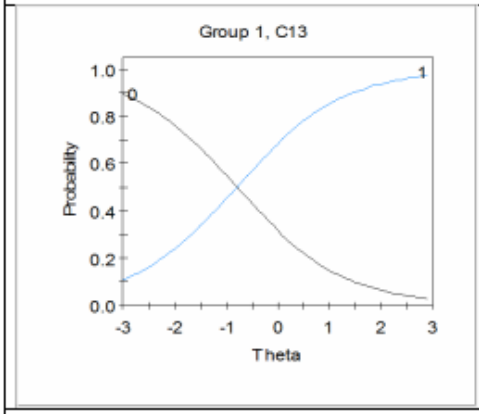
0501



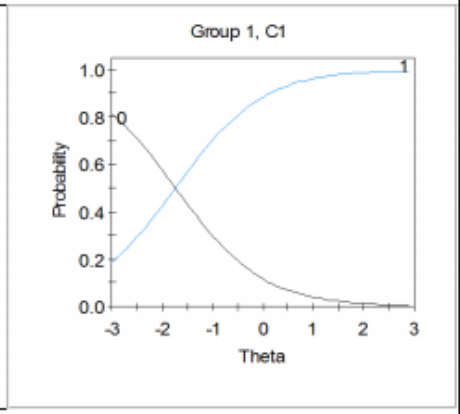
4427



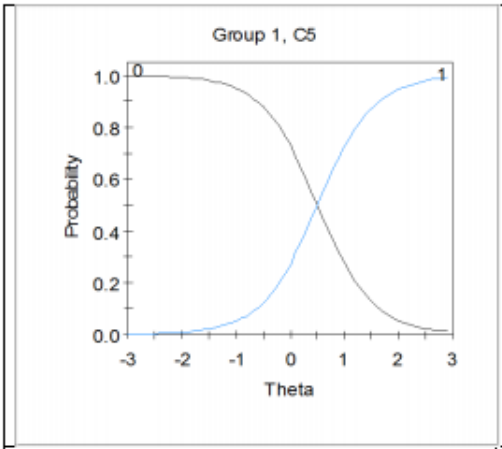
0505



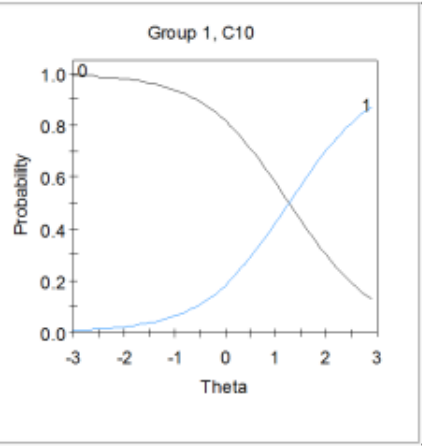
6471



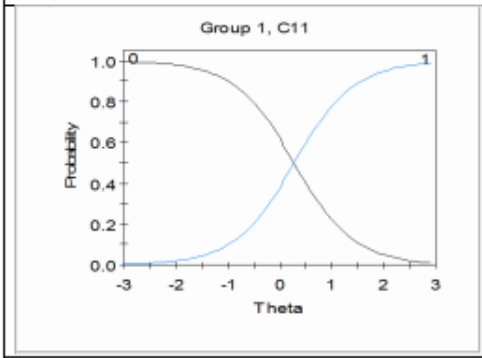
0203



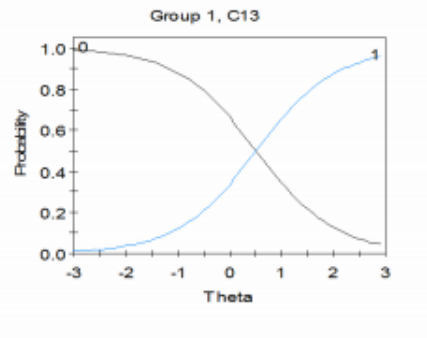
1119



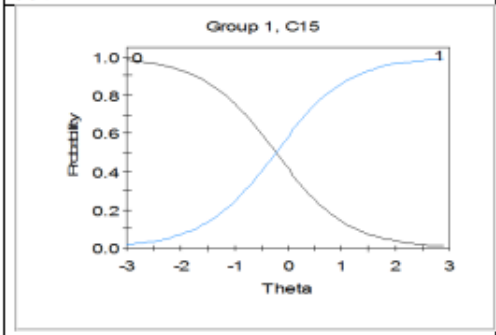
6370



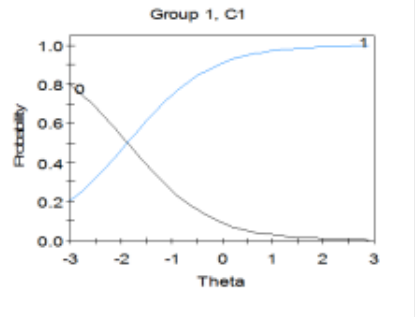
9342



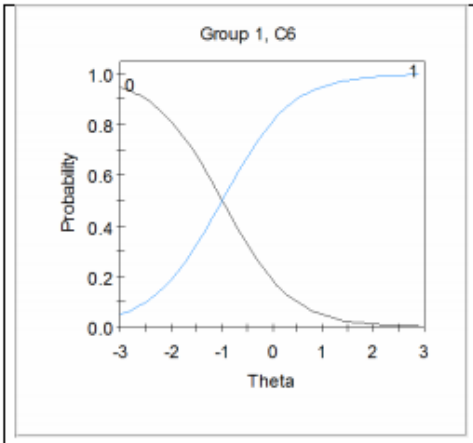
6323_1



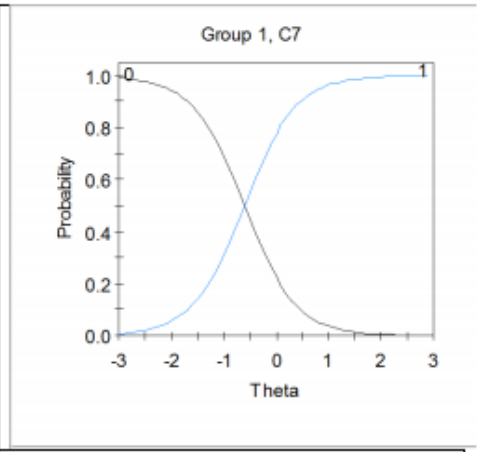
9074



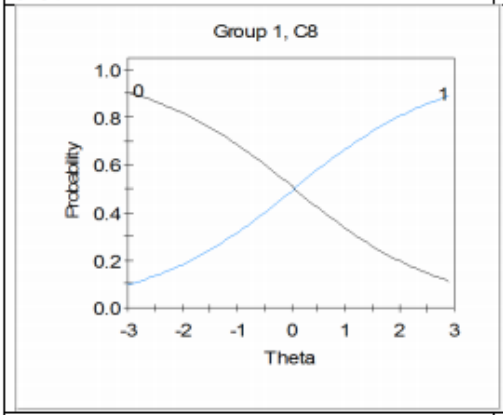
0507



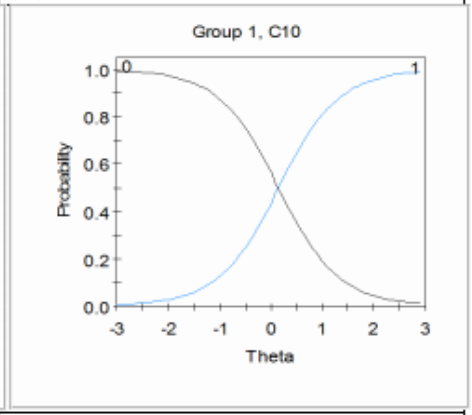
0204



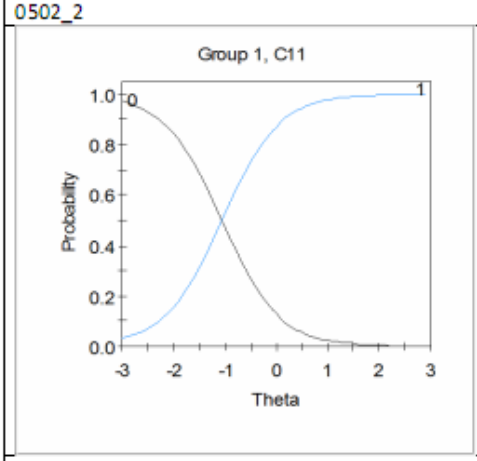
0201



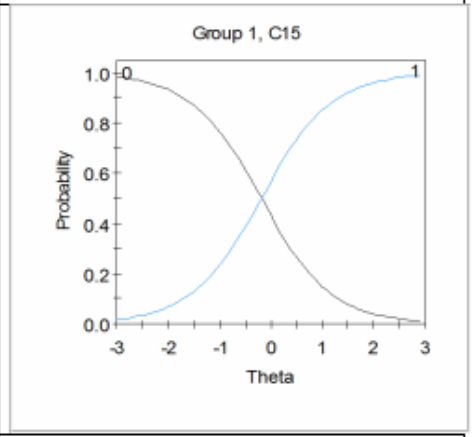
0502_2



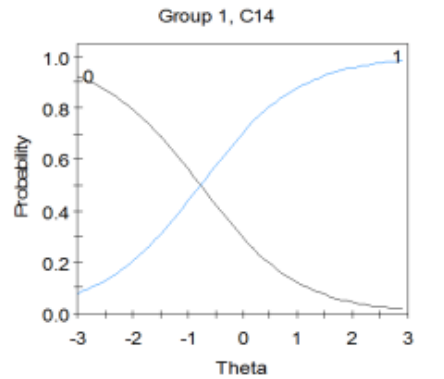
8867



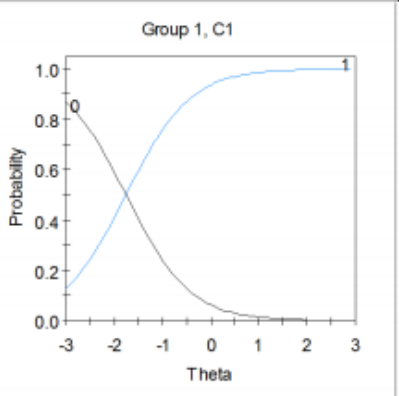
5752_3



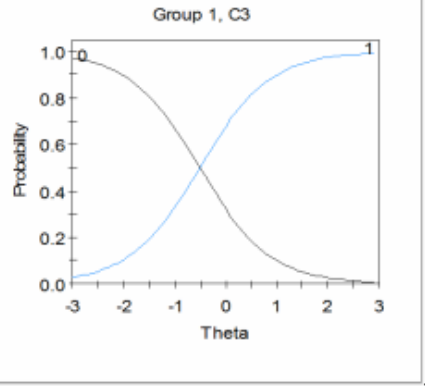
6963_B



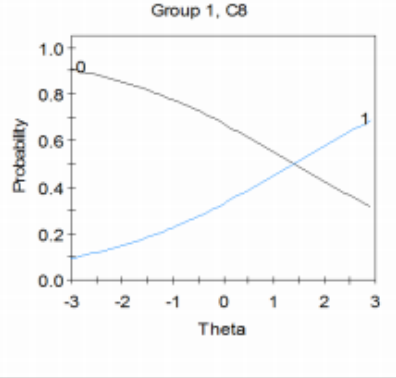
6963_A



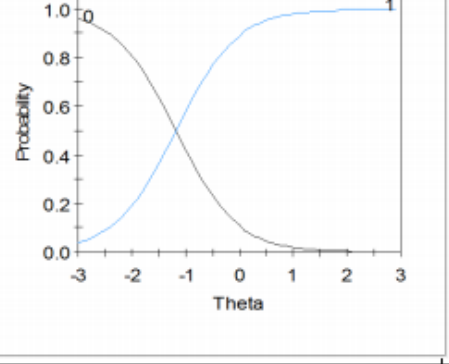
0206



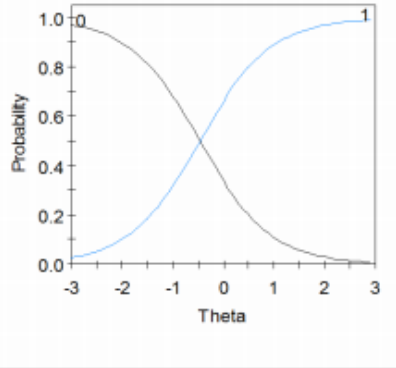
0503



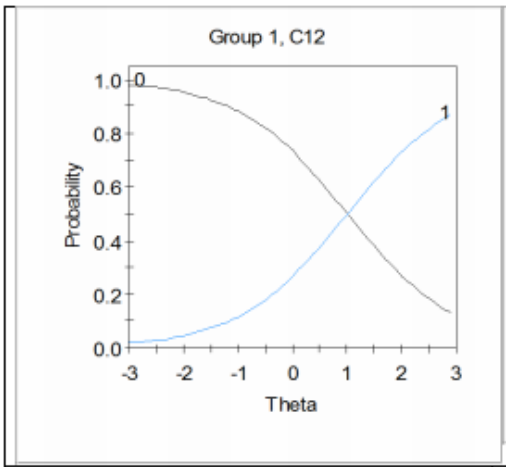
9739



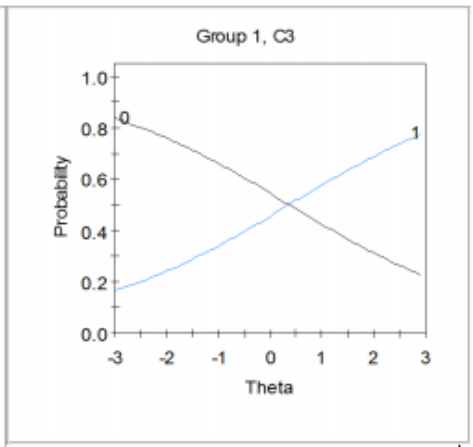
6251



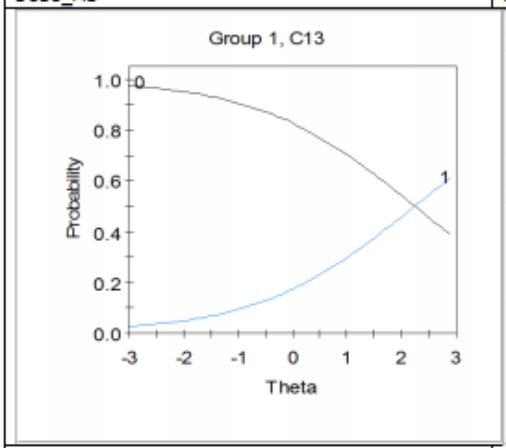
9270



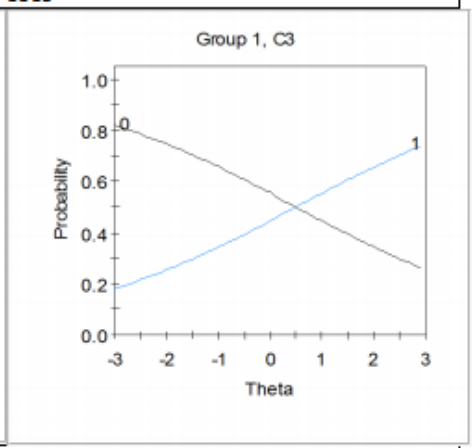
5050_A1



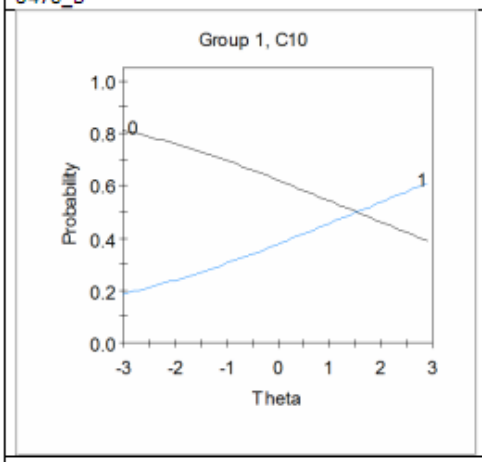
1113



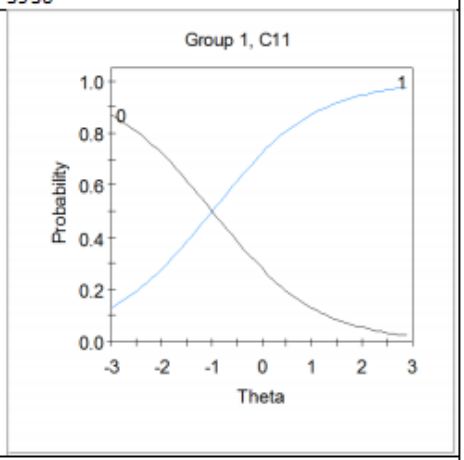
6470_B



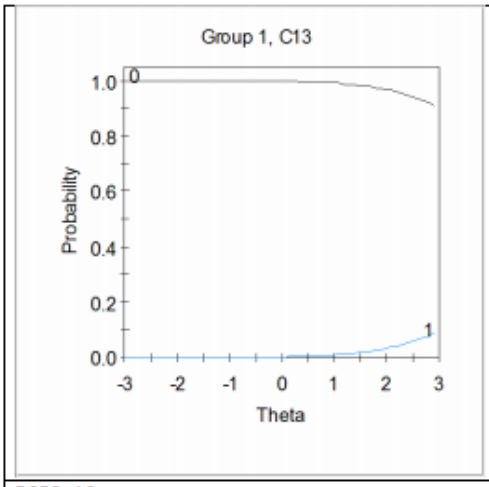
5936



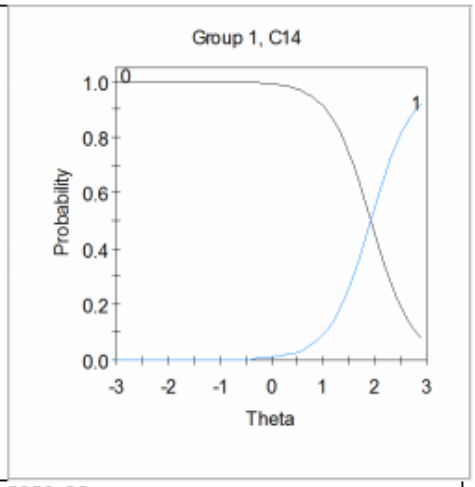
7305_B



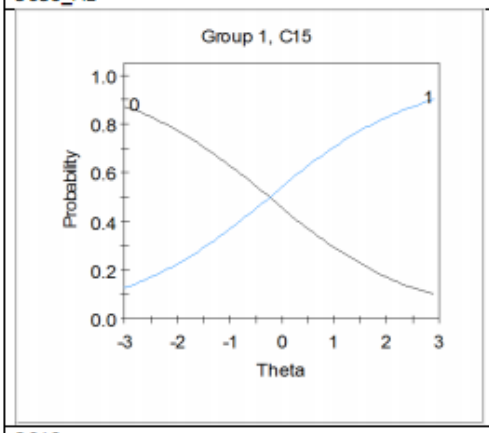
0202



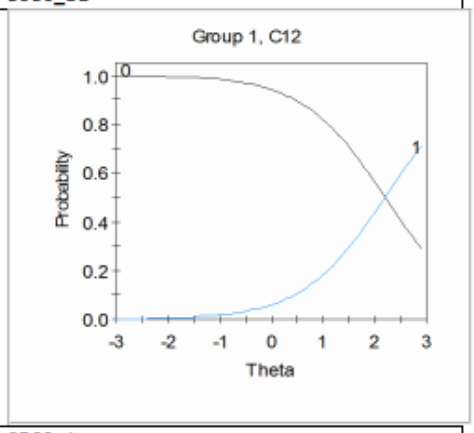
5050_A2



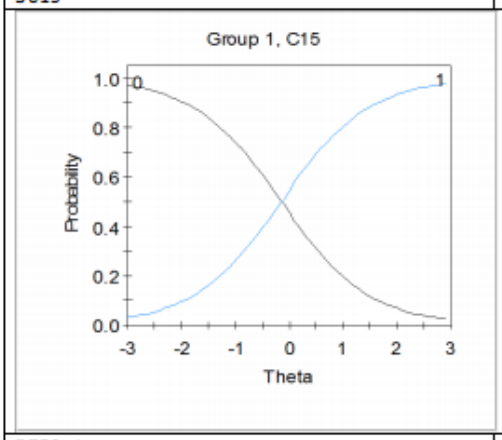
5050_B2



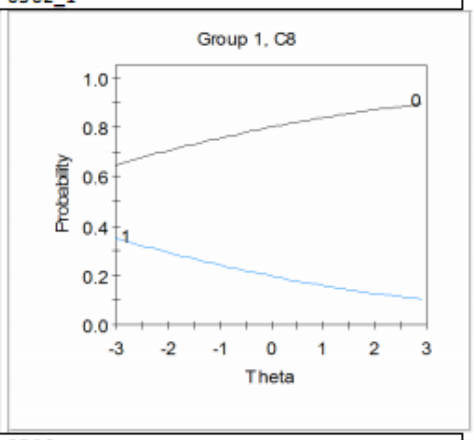
3019



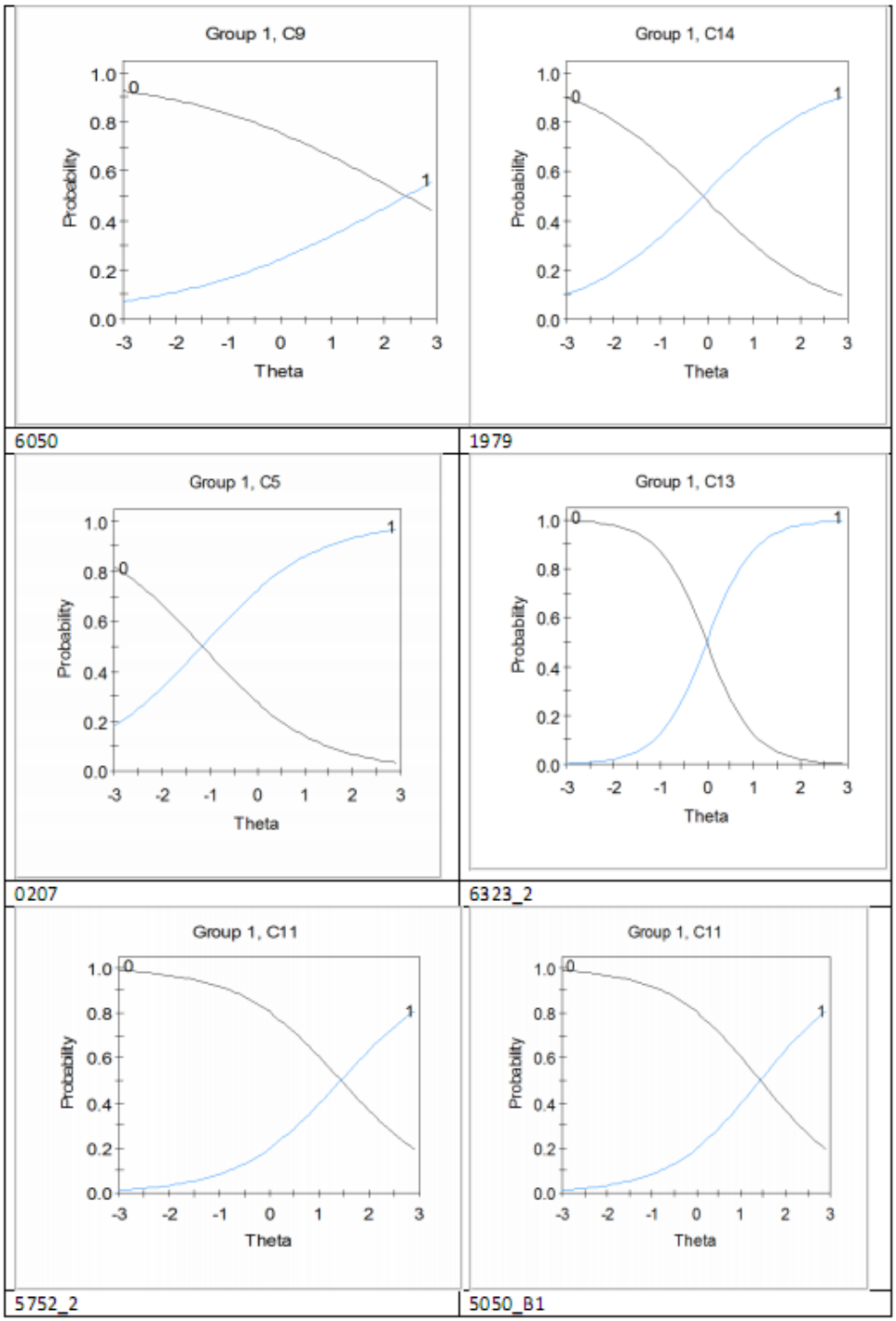
0502_1



5752_1



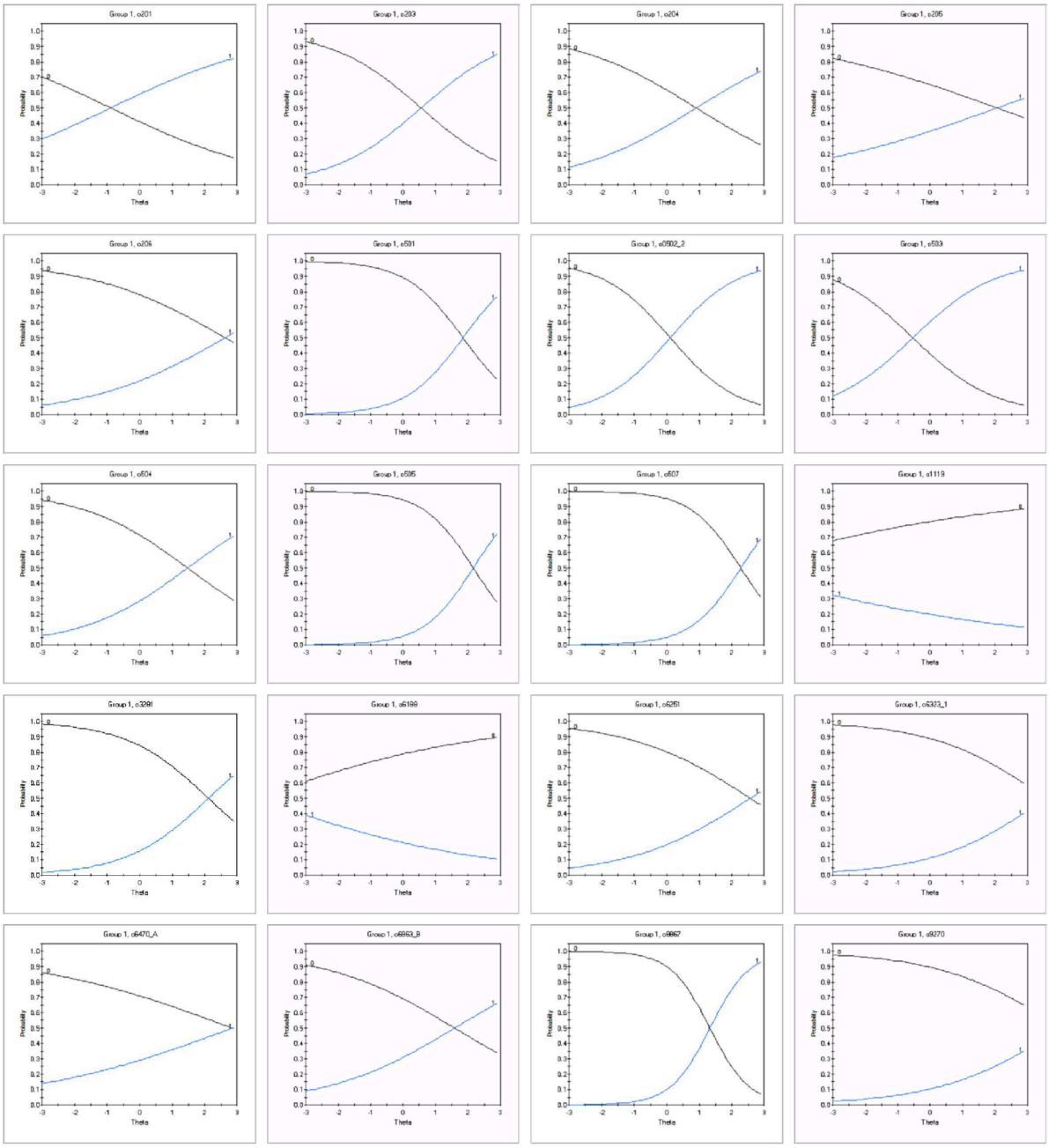
0506

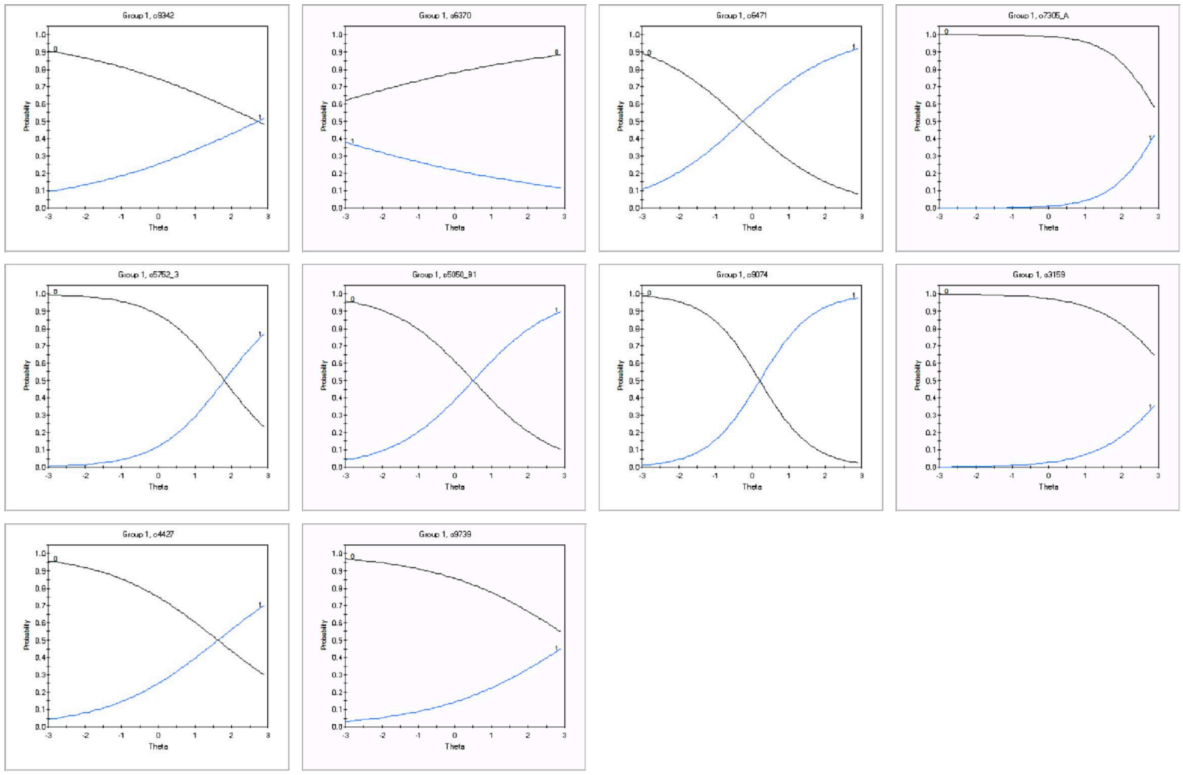

Group 1, C9
Probability
Theta
6050
Group 1, C14
Probability
Theta
1979
Group 1, C5
Probability
Theta
0207
Group 1, C13
Probability
Theta
6323_2
Group 1, C11
Probability
Theta
5752_2
Group 1, C11
Probability
Theta
5050_B1

EK 4. Ön test uygulaması MTK madde parametreleri

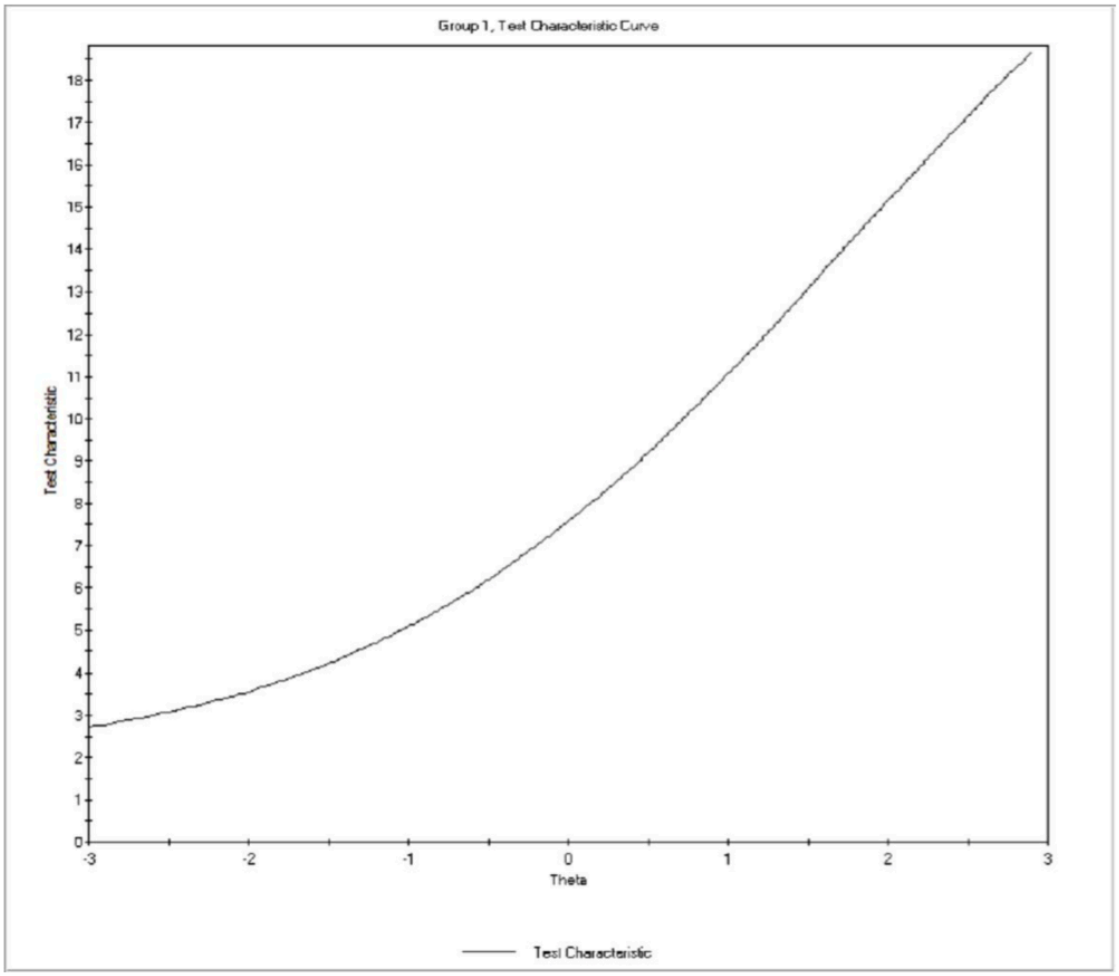
Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	o201	2	0.41	0.04	1	0.36	0.03	-0.89	0.12
2	o203	4	0.73	0.05	3	-0.41	0.04	0.56	0.06
3	o204	6	0.52	0.04	5	-0.48	0.04	0.91	0.10
4	o205	8	0.30	0.04	7	-0.62	0.04	2.06	0.30
5	o206	10	0.48	0.05	9	-1.27	0.04	2.63	0.26
6	o501	12	1.14	0.07	11	-2.11	0.07	1.85	0.09
7	o0502_2	14	0.97	0.06	13	-0.11	0.04	0.11	0.04
8	o503	16	0.80	0.05	15	0.43	0.04	-0.53	0.05
9	o504	18	0.62	0.05	17	-0.92	0.04	1.47	0.11
10	o505	20	1.30	0.09	19	-2.82	0.09	2.17	0.10
11	o507	22	1.30	0.09	21	-2.97	0.10	2.28	0.11
12	o1119	24	-0.22	0.05	23	-1.40	0.04	-6.40	1.44
13	o3281	26	0.79	0.06	25	-1.68	0.05	2.12	0.14
14	o6188	28	-0.29	0.05	27	-1.32	0.04	-4.56	0.77
15	o6251	30	0.54	0.05	29	-1.40	0.04	2.57	0.23
16	o6323_1	32	0.58	0.06	31	-2.08	0.06	3.56	0.34
17	o6470_A	34	0.31	0.04	33	-0.89	0.04	2.85	0.40
18	o6963_B	36	0.51	0.05	35	-0.81	0.04	1.59	0.15
19	o8867	38	1.65	0.09	37	-2.20	0.08	1.33	0.05
20	o9270	40	0.53	0.06	39	-2.16	0.06	4.04	0.44
21	o9342	42	0.40	0.05	41	-1.08	0.04	2.73	0.31
22	o6370	44	-0.26	0.05	43	-1.28	0.04	-4.91	0.90
23	o6471	46	0.77	0.05	45	0.19	0.04	-0.25	0.05
24	o7305_A	48	1.48	0.14	47	-4.59	0.20	3.11	0.20
25	o5752_3	50	1.11	0.07	49	-2.00	0.06	1.81	0.09
26	o5050_B1	52	0.91	0.05	51	-0.45	0.04	0.50	0.05
27	o9074	54	1.39	0.07	53	-0.29	0.04	0.21	0.03
28	o3159	56	1.04	0.10	55	-3.60	0.12	3.47	0.28
29	o4427	58	0.67	0.05	57	-1.10	0.04	1.63	0.12
30	o9739	60	0.55	0.06	59	-1.79	0.05	3.26	0.31

EK 5. Ön test uygulaması Madde Karakteristik eğrileri

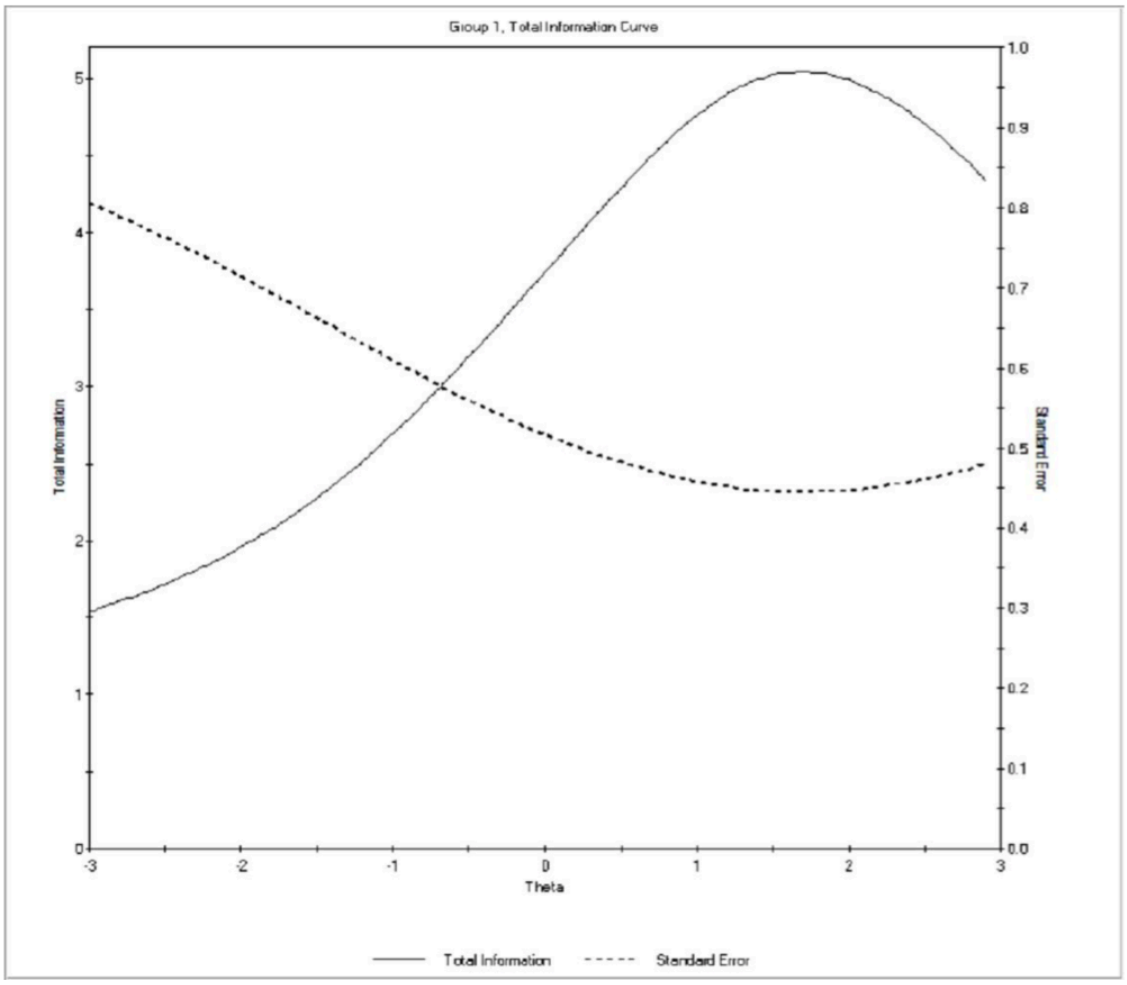



Group 1, o9342
Group 1, o6370
Group 1, o6471
Group 1, o7305_A
Group 1, o5752_3
Group 1, o5950_91
Group 1, o9074
Group 1, o3159
Group 1, o4427
Group 1, o9739
Probability
Theta

EK 6. Ön test uygulaması Test karakteristik Eğrisi



EK 7. Ön test uygulaması Test Bilgi Fonksiyonu



EK 8. Deney-1 grupları detaylı geri bildirim formları

Uluslararası Geniş Ölçekli Sınavlarda Türkiye'nin Matematik Başarısını Arttırabilmek İçin Bir Model Önerisi: Bilişsel Tanıya Dayalı İzleme Modelinin Etkililiği-115K531

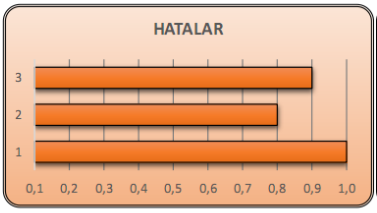
1.İZLEME TESTİ

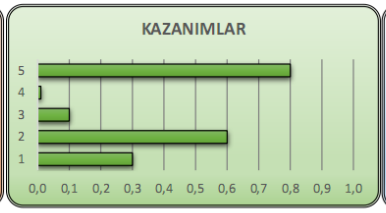
ÖĞRENCİ PROFİL DOSYASI

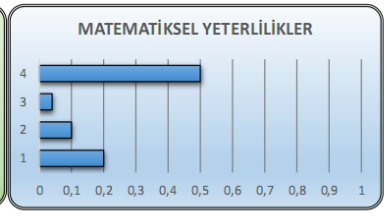
OKUL :	28-ASİL NADİR ORTAOKULU						
SINIF/ŞUBE :	6 C	**AD :**	[redacted]	**SOYAD:**	[redacted]	**NO:**	[redacted]

SORULAR	Bölünme kuralları (9796)	Otopark (3690)	Snapchat takipçi (1027)	Dergi aboneliği (1025)	Candy Crush Soda (1033)	Yunus'un Pateni (1021)	İzdeniz seferleri (1032)	Eratosten (7339)	Hiyeroglif sayı (6728)	Hanoi -1 (5158_B)	Hanoi -2 (5158_C)	**TOPLAM**
CEVAPLAR	1	0	0	0	0	1	1	0	1	0	0	**4**
KAZANIMLAR	K1	K1,K4	K1,K2	K2,K3	K1	K4	K1	K2	K2	K3	K5	
MATEMATİKSEL YETERLİKLER	M1	M2,M3	M2,M3,M4	M1,M4	M1	M1	M3	M1,M3	M1,M4	M1,M3	M1,M3,M4	

HATALAR		**KAZANIMLAR**		**MATEMATİKSEL YETERLİLİKLER**	
E1	Matematiksel Durumu Anlama Güçlüğü	**K1**	Doğal Sayılarda Dört İşlem	**M1**	İletişim ve İlişkilendirme
E2	Dönüştürme Güçlüğü	**K2**	Üslü Sayılarla İşlemler	**M2**	Matematikleştirme
E3	İşlem Güçlüğü	**K3**	Asal Sayılar ve Özellikleri	**M3**	Akıl Yürütme ve Strateji Geliştirme
		K4	Çarpanlar ve Katlar	**M4**	Sembolik ve Teknik Dil Kullanımı
		K5	Bölme ve Bölünebilme		







Not: Öğrenci isimleri gizlenmiştir

EK 9. Deney-2 grupları geri bildirim formları

07-EREN ŞAHİN ERONAT ORTAOKULU																
SINIF ŞUBE	NO	AD	SOYAD	Bölünme kuralları (9796)	Otopark (3690)	Snapchat takipçi (1027)	Dergi aboneliği (1025)	Candy Crush Soda (1033)	Yunus'un Pateni (1021)	İzdeniz seferleri (1032)	Eratosten (7339)	Hiyeroglif sayı (6728)	Hanoi -1 (5158_B)	Hanoi -2 (5158_C)	TOPLAM PUAN
6/B	[illegible]	[illegible]	[illegible]	0	0	0	0	0	0	0	0	0	0	0	0
6/B	781	[illegible]	[illegible]	1	0	0	0	0	0	0	0	0	0	0	1
6/B	796	[illegible]	[illegible]	1	0	0	0	1	0	0	0	0	0	0	2
6/B	1130	[illegible]	[illegible]	0	0	0	0	1	0	1	0	0	0	0	2
6/B	1168	[illegible]	[illegible]	0	0	0	0	0	0	0	0	0	0	1	1
6/B	761	[illegible]	[illegible]	1	0	0	0	1	1	1	0	0	1	1	6
6/B	2387	[illegible]	[illegible]	1	0	1	0	1	1	0	0	0	1	0	5
6/B	792	[illegible]	[illegible]	1	0	0	0	0	1	1	0	0	0	0	3
6/B	789	[illegible]	[illegible]	1	1	0	0	0	1	1	0	0	0	0	4
6/B	1202	[illegible]	[illegible]	1	0	0	1	0	0	0	0	0	0	1	3
6/B	769	[illegible]	[illegible]	1	0	0	1	1	1	0	0	0	1	1	6
6/B	784	[illegible]	[illegible]	0	0	0	0	1	1	1	0	1	0	0	4
6/B	797	[illegible]	[illegible]	1	0	0	1	1	1	1	0	1	1	0	7
6/B	1213	[illegible]	[illegible]	1	0	0	1	1	1	0	0	0	0	0	4
6/B	782	[illegible]	[illegible]	1	0	1	0	1	1	0	0	0	1	0	5
6/B	584	[illegible]	[illegible]	1	1	1	1	1	1	1	0	1	1	1	10
6/B	767	[illegible]	[illegible]	1	0	0	0	1	1	1	0	1	0	0	5
6/B	795	[illegible]	[illegible]	1	0	1	1	1	1	0	0	0	1	1	7
6/B	1083	[illegible]	[illegible]	1	1	0	0	1	1	0	0	1	1	1	7
6/B	793	[illegible]	[illegible]	1	0	0	0	0	0	0	0	0	0	0	1
6/B	788	[illegible]	[illegible]	1	0	0	0	1	0	1	0	0	0	0	3
6/B	41	[illegible]	[illegible]	1	0	0	1	0	1	0	0	1	1	1	6
6/B	900	[illegible]	[illegible]	1	1	0	1	1	1	0	1	1	1	1	9
6/B	87	[illegible]	[illegible]	0	0	0	1	1	1	1	0	1	1	0	6
6/B	987	[illegible]	[illegible]	1	0	0	0	0	1	0	0	0	1	1	4
6/B	72	[illegible]	[illegible]	0	0	1	0	0	1	0	0	0	0	1	3
6/B	790	[illegible]	[illegible]	1	0	0	0	0	1	0	0	0	0	1	3
6/B	775	[illegible]	[illegible]	0	0	0	0	1	1	0	0	0	0	0	2
6/B	787	[illegible]	[illegible]	1	0	0	1	1	1	1	0	0	1	0	6
6/B	771	[illegible]	[illegible]	1	0	1	1	1	1	1	0	0	1	1	8
6/B	778	[illegible]	[illegible]	0	0	1	1	1	1	0	0	0	1	1	6
6/B	785	[illegible]	[illegible]	0	1	0	0	0	0	0	0	0	0	0	1
6/B	214	[illegible]	[illegible]	1	1	1	1	1	1	1	0	1	1	1	10
6/B	774	[illegible]	[illegible]	0	0	0	1	1	1	1	0	1	1	1	7
6/B	825	[illegible]	[illegible]	1	1	1	1	0	1	0	0	1	0	0	6
6/B	[illegible]	[illegible]	[illegible]	1	1	1	0	1	1	0	0	1	1	0	7

Not: Öğrenci isimleri gizlenmiştir

EK 10. İzleme-1Testi Pilot Uygulaması(20) Madde Parametreleri

No	Soru Kodu	Madde Güçlüğü (Pj)	Madde Ayırt Ediciliği (rbis)
1	P9796	0,78	0,60
2	P3690	0,19	0,15
3	P1027	0,23	0,65
4	P1025	0,16	0,24
5	P1033	0,35	0,57
6	P1021	0,40	0,53
7	P1032	0,39	0,73
8	P7339	0,01	0,53
9	P6728	0,35	0,71
10	P5158_B	0,10	0,81
11	P5158_C	0,40	0,68
	Ortalama	0,31	0,56

EK 11. İzleme-1 Testi(21) Okul Şube İstatistikleri

Okul	Şube	Kişi Sayısı	Ortalama	Standart Sapma
03	A	37	3,70	1,79
	B	35	4,74	2,42
	C	38	4,21	2,16
	D	34	3,85	2,39
	E	34	4,29	2,20
05	A	43	3,77	2,08
	B	45	3,20	1,66
	C	44	3,61	1,82
	D	43	3,98	2,04
06	A	23	3,48	2,19
	B	20	3,75	2,00
	C	24	3,71	1,76
	D	24	3,46	1,96
	E	24	4,04	1,88
	F	24	3,71	1,83
	G	22	3,82	1,71
	H	21	4,81	2,04
	I	23	2,91	1,65
	İ	22	3,18	1,56
07	A	37	4,78	2,25
	B	36	4,72	2,61
	C	38	4,39	2,19
	D	37	4,14	2,03
	E	38	4,11	1,83
	F	35	4,17	1,77
	G	33	3,91	2,08
16	A	27	2,85	1,81
	B	27	2,96	1,83
	C	26	2,96	1,61
	D	21	3,14	1,59
	E	29	3,55	2,11
	F	24	3,42	2,10
17	A	32	3,50	1,97
	B	38	3,95	2,36
	C	32	3,13	1,64
	D	40	3,40	2,00
	E	32	3,00	2,02
	F	32	3,41	1,90
	G	38	3,63	2,20
	H	33	2,85	1,46
18	A	26	4,23	1,92
	B	27	3,48	1,97
	C	28	3,79	2,01
	D	24	2,79	1,64
	E	27	4,07	1,92
	F	27	3,41	2,15
32	A	19	6,31	2,11

Okul	Şube	Kişi Sayısı	Ortalama	Standart Sapma
19	A	26	3,38	2,37
	B	29	3,69	2,35
	C	29	3,59	2,28
	D	32	3,13	2,08
	E	30	3,70	1,56
	F	24	4,00	1,87
	G	31	3,29	2,19
20	A	29	2,52	1,79
	B	25	2,48	1,61
	C	24	3,38	2,10
	D	25	3,24	2,09
	E	24	2,92	1,86
	F	23	2,70	1,46
	G	22	2,95	1,94
22	A	23	2,65	2,06
	B	25	2,96	1,72
	C	29	3,59	1,66
	D	23	2,65	1,53
	E	24	2,58	1,50
	F	21	3,19	1,97
	G	20	2,40	1,50
	H	20	2,80	1,51
	I	25	2,68	1,22
23	A	32	3,13	1,95
	B	30	3,00	1,29
	C	31	2,77	1,36
	D	32	3,31	1,69
	E	31	4,00	1,55
	F	23	2,22	1,35
27	A	32	3,66	1,96
	B	29	3,03	1,48
	C	31	3,29	2,08
	D	32	3,28	1,69
	E	32	3,00	2,13
	F	32	3,72	2,19
	G	30	3,23	1,94
	H	31	3,71	1,90
	İ	28	3,46	2,01
	J	26	3,81	1,90
	K	30	3,37	2,11
	L	30	3,53	1,80
	M	29	4,14	2,63
	N	31	3,48	2,05
28	A	30	3,20	1,67
	B	33	4,42	2,08
	C	31	4,13	1,93
	D	32	4,50	2,00

B	18	7,72	2,36	33	A	22	6,14	2,42
C	17	3,52	2,74		B	24	7,25	2,27
D	18	6,38	3,11		C	24	5,54	2,45
E	17	5,74	2,87		D	23	5,78	2,49

EK 12. İzleme-1(21) Q Matrisler

	Matematiksel Yeterlikler				Kazanımlar					Hatalar		
	M1	**M2**	**M3**	**M4**	**K1**	**K2**	**K3**	**K4**	**K5**	**E1**	**E2**	**E3**
9796	0	0	0	1	1	0	0	0	0	1	1	1
3690	1	1	0	0	0	0	0	1	1	0	0	1
1027	1	1	1	0	0	0	1	1	0	1	0	0
1025	0	0	1	1	0	0	0	1	0	0	0	1
1033	0	0	0	1	0	0	0	1	0	1	1	0
1021	0	0	0	1	0	0	0	1	0	1	0	0
1032	1	0	0	0	0	0	0	0	1	0	1	1
7339	1	0	0	1	0	1	0	0	0	1	0	0
6728	0	0	1	1	0	1	1	0	0	1	1	1
5158_B	1	0	0	1	0	0	1	0	0	1	0	0
5158_C	1	0	1	1	0	0	1	0	0	1	0	0

M1, M2, M3 ve M4 olarak isimlendirilen matematiksel yeterlikler sırasıyla “İletişim ve İlişkilendirme”, “Matematikleştirme”, “Akıl Yürütme ve Strateji Geliştirme” ve “Sembolik ve Teknik Dil Kullanımı” şeklindedir. K1, K2, K3, K4 ve K5 olarak isimlendirilen kazanımlar sırasıyla “Doğal Sayılarda Dört İşlem”, “Üslü Sayılarla İşlemler ”, “Asal Sayılar ve Özellikleri ”, “Çarpanlar ve Katlar” ve “Bölme ve Bölünebilme” şeklindedir. Son olarak E1, E2 ve E3 olarak isimlendirilen hatalar ise “Matematiksel Durumu Anlama Güçlüğü”, “Dönüştürme Güçlüğü” ve “İşlem Güçlüğü” şeklindedir.

EK 13. İzleme-2 Testi Pilot Uygulaması(30) Madde Parametreleri

No	Soru Kodu	Madde Güçlüğü (Pj)	Madde Ayırt Ediciliği (rbis)
1	P1011_A	0,28	0,54
2	P1011_B	0,12	0,75
3	P1011_C	0,02	0,43
4	P1012_A	0,37	0,69
5	P1012_B	0,33	0,44
6	P1013_A	0,11	0,82
7	P1013_B	0,25	0,60
8	P1014_A	0,28	0,55
9	P1014_B	0,04	0,40
10	P1014_C	0,18	0,40
11	P1015_A	0,07	0,61
12	P1015_B	0,35	0,55
	Ortalama	0,20	0,57

EK 14. İzleme-2 Testi(31) Okul Şube İstatistikleri

Okul	Şube	Kişi Sayısı	Ortalama	Standart Sapma
03	A	37	3,43	1,86
	B	35	3,40	1,70
	C	37	2,81	1,24
	D	36	3,33	1,84
	E	33	2,70	1,33
05	A	45	2,91	2,04
	B	45	2,93	1,42
	C	41	2,49	1,50
	D	46	2,76	1,61
06	A	22	3,86	1,78
	B	21	2,62	1,47
	C	23	3,00	2,00
	D	24	2,25	1,59
	E	18	3,22	2,24
	F	22	2,82	1,37
	G	22	3,41	1,50
	H	18	2,72	1,64
	I	23	2,39	1,27
	İ	22	2,55	1,37
07	A	36	3,53	1,65
	B	34	2,94	1,56
	C	38	3,66	1,77
	D	34	3,03	1,45
	E	38	2,92	1,55
	F	34	3,06	1,81
	G	35	2,34	1,57
16	A	25	2,64	1,41
	B	23	2,91	1,35
	C	26	2,23	1,45
	D	20	2,40	1,50
	E	24	2,63	1,35
	F	27	2,37	1,76
17	A	35	2,34	1,43
	B	33	2,21	1,49
	C	33	2,15	1,33
	D	36	2,72	1,80
	E	33	2,45	1,39
	F	32	2,47	1,39
	G	33	2,30	1,38
	H	28	2,14	1,21
18	A	27	2,19	1,42
	B	26	2,85	1,95
	C	25	3,40	1,35
	D	23	1,96	1,11
	E	27	2,67	1,41
	F	26	1,96	1,22
32	A	18	4,33	1,75
	B	19	3,11	1,33
19	A	26	2,15	1,46
	B	27	2,15	1,63
	C	26	2,46	1,33
	D	28	2,29	1,56
	E	24	2,54	1,41
	F	27	2,74	2,03
	G	27	1,85	1,51
20	A	27	1,96	1,83
	B	24	2,17	1,31
	C	24	3,38	1,13
	D	25	2,24	1,39
	E	24	2,54	1,59
	F	23	1,57	1,62
	G	24	1,42	1,25
22	A	25	2,12	,88
	B	26	1,73	1,04
	C	26	2,54	1,73
	D	24	2,25	1,26
	E	24	2,08	1,35
	F	21	2,38	1,56
	G	20	2,50	1,15
	H	22	2,32	1,73
	I	25	2,28	1,40
23	A	28	1,93	1,30
	B	26	1,96	1,51
	C	27	2,30	1,27
	D	26	2,65	1,52
	E	29	2,86	1,55
	F	19	1,42	1,02
27	A	31	2,97	1,33
	B	33	2,15	1,18
	C	29	1,97	1,18
	D	31	2,77	1,38
	E	32	2,22	1,07
	F	31	2,39	1,56
	G	32	2,72	1,35
	H	32	2,03	1,33
	İ	33	3,15	1,86
	J	24	2,92	1,95
	K	30	2,37	1,77
	L	31	2,06	1,55
	M	33	2,52	1,60
	N	32	2,19	1,73
28	A	29	2,41	1,35
	B	33	2,85	1,52
	C	27	2,67	1,39
	D	29	2,55	1,12
33	A	21	5,33	1,46

C	17	3,65	2,21	B	23	5,30	2,16
D	16	4,69	1,58	C	22	5,00	2,07
E	17	4,53	2,18	D	21	4,14	1,59

EK 15. İzleme-2(31) Q Matrisler

	Matematiksel Yeterlikler				Kazanımlar					Hatalar		
	M1	**M2**	**M3**	**M4**	**K1**	**K2**	**K3**	**K4**	**K5**	**E1**	**E2**	**E3**
1011_A	0	0	1	0	1	1	0	0	0	1	1	0
1011_B	1	0	1	0	1	1	1	0	0	1	0	1
1012_A	1	0	0	1	1	0	0	0	0	1	0	1
1011_C	1	0	1	0	1	1	1	0	0	1	0	1
1012_B	1	0	0	1	1	0	0	0	0	0	1	1
1014_A	1	0	0	0	1	1	0	0	0	1	0	1
1015_A	1	1	0	0	0	0	0	1	1	1	1	0
1014_B	1	0	0	0	1	0	0	1	0	0	1	1
1015_B	0	1	1	0	0	0	0	0	1	0	1	0
1014_C	1	0	0	0	0	0	1	0	0	1	0	0

Ek 15'de M1, M2, M3 ve M4 olarak isimlendirilen matematiksel yeterlikler sırasıyla "İletişim ve İlişkilendirme", "Matematikleştirme", "Akıl Yürütme ve Strateji Geliştirme" ve "Sembolik ve Teknik Dil Kullanımı" şeklindedir. K6, K7, K8, K9 ve K10 olarak isimlendirilen kazanımlar sırasıyla "Kavramsal Kesir Bilgisi", "Kesirlerle İşlemler", "Kesirlerde Problemler", "Kavramsal Oran Bilgisi" ve "Açı ve Özellikleri" şeklindedir. Son olarak E1, E2 ve E3 olarak isimlendirilen hatalar ise "Matematiksel Durumu Anlama Güçlüğü", "Dönüştürme Güçlüğü" ve "İşlem Güçlüğü" şeklindedir. Q Matristeki her bir hücre ilişkili olduğu soruyu doğru cevaplayabilmek için matematiksel yeterliklere ve kazanımlara sahip olup olmamayı göstermektedir.

EK 16. İzleme-3 Testi Pilot Uygulaması(40) Madde Parametreleri

No	Soru Kodu	Madde Güçlüğü (Pj)	Madde Ayırt Ediciliği (rbis)
1	P1023_A	0,00	0,00
2	P1023_B	0,23	0,31
3	P1024	0,28	0,45
4	P1028	0,38	0,58
5	P1022_A	0,05	0,17
6	P1022_B	0,39	0,52
7	P1026	0,43	0,62
8	P1020_A	0,41	0,43
9	P1020_B	0,22	0,39
10	P1020_C	0,21	0,40
11	P1021_A	0,05	0,28
12	P1021_B	0,45	0,40
13	P1021_C	0,18	0,24
	Ortalama	0,25	0,37

EK 17. İzleme-3 Testi(41) Okul Şube İstatistikleri

Okul	Şube	Kişi Sayısı	Ortalama	Standart Sapma
03	A	36	4,81	2,04
	B	37	4,19	1,96
	C	38	3,89	1,71
	D	34	4,15	1,84
	E	33	5,06	2,28
05	A	46	3,98	1,84
	B	44	3,25	1,60
	C	43	3,77	1,74
	D	44	3,20	1,41
06	A	22	3,95	1,62
	B	24	3,38	1,69
	C	22	4,36	2,26
	D	22	3,68	1,91
	E	24	3,79	1,82
	F	24	2,83	1,69
	G	23	3,13	2,47
	H	22	3,77	1,82
	I	23	4,13	1,25
	İ	22	3,14	1,28
07	A	37	4,27	1,92
	B	33	4,67	2,15
	C	39	3,92	1,97
	D	28	3,68	2,04
	E	35	4,14	1,78
	F	39	4,36	2,35
	G	33	3,42	1,50
16	A	26	3,42	1,90
	B	28	3,25	1,55
	C	24	3,33	1,43
	D	21	3,38	1,75
	E	26	3,96	2,09
	F	27	3,48	1,99
17	A	30	3,90	1,94
	B	37	3,00	1,72
	C	34	3,09	1,83
	D	40	3,38	1,72
	E	32	3,22	1,90
	F	33	3,55	1,52
	G	36	3,50	1,58
	H	31	3,13	1,77
18	A	26	3,54	2,70
	B	27	3,63	1,78
	C	27	3,85	1,88
	D	23	3,22	1,51
	E	26	4,12	2,30
	F	25	3,76	2,13
32	A	18	6,28	1,90
	B	19	5,58	2,14
	C	16	4,81	1,64
	D	18	6,44	1,72
19	A	29	3,31	1,69
	B	27	2,78	1,60
	C	23	3,09	1,62
	D	30	3,67	1,56
	E	30	2,87	1,70
	F	24	3,13	1,87
	G	27	3,44	1,48
20	A	27	2,96	1,60
	B	27	3,22	1,65
	C	26	3,65	1,90
	D	24	3,54	1,59
	E	22	3,64	1,68
	F	23	3,09	1,12
	G	24	2,63	1,06
22	A	24	3,13	1,54
	B	24	2,79	1,25
	C	28	3,68	1,47
	D	24	2,67	1,37
	E	25	3,12	1,81
	F	24	2,58	1,64
	G	18	3,33	1,37
	H	24	2,96	1,68
	I	25	3,24	1,67
23	A	29	3,24	1,72
	B	29	2,72	1,46
	C	32	2,56	1,34
	D	30	3,10	1,67
	E	32	4,13	1,43
	F	21	2,48	1,57
27	A	33	3,33	1,59
	B	29	3,59	1,70
	C	28	3,43	2,08
	D	32	3,47	2,02
	E	28	3,50	1,64
	F	31	3,35	1,78
	G	30	3,60	1,52
	H	33	3,21	1,76
	İ	30	3,50	2,35
	J	29	3,72	2,05
	K	29	3,62	1,63
	L	16	3,50	1,90
	M	31	4,55	1,91
	N	30	3,70	1,66
28	A	31	4,13	2,25
	B	31	4,39	1,52
	C	31	4,61	1,80
	D	31	4,35	2,03
33	A	23	6,48	2,64
	B	24	6,75	2,35
	C	25	5,44	2,08

E	17	4,35	2,57	D	24	4,88	2,05

EK 18. İzleme-3(41) Q Matrisler

	Matematiksel Yeterlikler				Kazanımlar					Hatalar		
	M1	**M2**	**M3**	**M4**	**K1**	**K2**	**K3**	**K4**	**K5**	**E1**	**E2**	**E3**
1023_A	1	0	1	0	0	0	1	1	0	1	1	0
1023_B	0	1	0	0	0	1	0	0	0	0	1	0
1024	1	0	0	1	0	1	0	0	0	1	0	0
1028	0	1	0	0	0	0	1	1	0	1	0	0
1022_A	0	0	1	0	0	0	0	0	1	1	1	1
1022_B	1	0	0	0	0	0	1	1	0	0	0	1
1026	1	0	1	0	0	1	0	0	0	0	1	0
1020_A	1	0	1	0	0	0	1	1	0	0	0	1
1020_B	1	1	1	0	1	1	1	1	0	1	1	0
1020_C	1	0	1	0	1	0	0	0	0	0	1	1
1021_A	1	1	1	0	1	0	0	0	0	0	1	0
1021_B	0	0	1	0	0	0	0	0	1	0	1	0
1021_C	1	0	0	1	1	0	0	0	0	1	0	1

EK 18'de M1, M2, M3 ve M4 olarak isimlendirilen matematiksel yeterlikler sırasıyla "İletişim ve İlişkilendirme", "Matematikleştirme", "Akıl Yürütme ve Strateji Geliştirme" ve "Sembolik ve Teknik Dil Kullanımı" şeklindedir. K11, K12, K13, K14 ve K15 olarak isimlendirilen kazanımlar sırasıyla "Kavramsal Tam Sayı Bilgisi", "Tam Sayı Bilgisi Uygulama", "Kavramsal Ondalık Sayı Bilgisi", "Ondalık Sayılarda İşlemler" ve "Veri Analizi" şeklindedir. Son olarak E1, E2 ve E3 olarak isimlendirilen hatalar ise "Matematiksel Durumu Anlama Güçlüğü", "Dönüştürme Güçlüğü" ve "İşlem Güçlüğü" şeklindedir.

EK 19. İzleme-4 Testi Pilot Uygulaması(50) Madde Parametreleri

No	Soru Kodu	Madde Güçlüğü (P_j)	Madde Ayırt Ediciliği (r_{bis})
1	P5017_A	0,32	0,45
2	P5017_B	0,04	0,12
3	P5036_A	0,39	0,65
4	P5036_B	0,23	0,27
5	P5036_C	0,17	0,32
6	P5058_A	0,78	0,53
7	P5058_B	0,62	0,64
8	P5059	0,34	0,43
9	P5057	0,78	0,54
	Ortalama	0,41	0,44

EK 20. İzleme-4 Testi(51) Okul Şube İstatistikleri

Okul	Şube	Kişi Sayısı	Ortalama	Standart Sapma
3	A	37	3,86	1,25
	B	35	3,49	1,74
	C	36	3,64	1,91
	D	34	3,71	1,82
	E	30	3,90	1,49
5	A	46	3,04	1,35
	B	43	2,79	1,41
	C	42	3,19	1,45
	D	42	2,76	1,51
6	A	21	3,52	1,57
	B	22	3,91	1,54
	C	17	3,12	1,58
	D	22	3,82	1,14
	E	23	3,61	1,73
	F	22	3,91	1,44
	G	25	3,96	1,59
	H	21	3,95	1,72
	I	23	3,00	1,62
	İ	23	4,13	1,25
7	A	33	3,76	1,35
	B	35	3,63	1,57
	C	19	4,68	1,53
	D	33	4,09	1,40
	E	38	3,61	1,46
	F	39	3,56	1,41
	G	33	3,30	1,53
16	A	24	2,54	1,53
	B	24	3,08	1,61
	C	24	1,83	1,24
	D	21	2,05	1,60
	E	27	2,96	1,45
	F	25	3,00	2,04
17	A	31	3,26	1,26
	B	33	3,06	1,90
	C	33	2,67	1,36
	D	40	3,10	1,69
	E	28	3,14	1,30
	F	29	2,62	1,32
	G	15	3,07	1,58
	H	30	2,73	1,57
18	A	27	4,70	1,81
	B	28	3,57	1,40
	C	27	4,26	1,29
	D	22	3,09	1,57
	E	27	3,89	1,22
	F	25	3,96	1,14
32	A	18	5,50	1,04
	B	19	5,16	1,17
20	A	27	3,22	1,72
	B	28	1,93	1,09
	C	25	3,12	1,48
	D	27	3,11	1,42
	E	28	3,43	1,45
	F	21	2,38	1,28
	G	24	2,92	1,47
22	A	24	2,83	1,09
	B	24	2,83	1,31
	C	27	3,15	1,54
	D	21	3,29	1,45
	E	21	2,90	1,26
	F	21	2,67	,91
	G	18	2,72	1,02
	H	24	3,13	1,51
	I	27	3,15	1,10
23	A	30	3,20	1,27
	B	31	2,84	,97
	C	32	2,47	1,08
	D	30	2,80	1,03
	E	31	2,13	1,28
	F	20	2,15	1,14
27	A	34	2,85	1,64
	B	31	3,00	,97
	C	33	3,06	1,50
	D	31	3,26	1,37
	E	29	3,34	1,37
	F	32	3,41	1,76
	G	28	3,68	1,12
	H	32	3,53	1,41
	İ	29	3,76	1,70
	J	33	3,58	1,46
	K	31	3,26	1,29
	L	32	2,81	1,49
	M	30	3,20	1,92
	N	30	3,07	1,14
28	A	31	3,61	1,23
	B	31	3,55	1,15
	C	31	3,23	1,45
	D	33	3,73	1,33
33	A	22	5,09	1,57
	B	24	6,04	1,52
	C	24	5,00	1,69
	D	20	4,00	1,78
19	A	28	2,29	1,08
	B	29	2,76	1,64
	C	30	3,40	1,30
	D	28	3,04	1,37

C	10	4,70	1,25	E	30	3,03	1,43
D	18	5,11	1,64	F	29	3,62	1,29
E	16	4,56	1,67	G	32	3,06	1,39

EK 21. İzleme-4(51) Q Matrisler

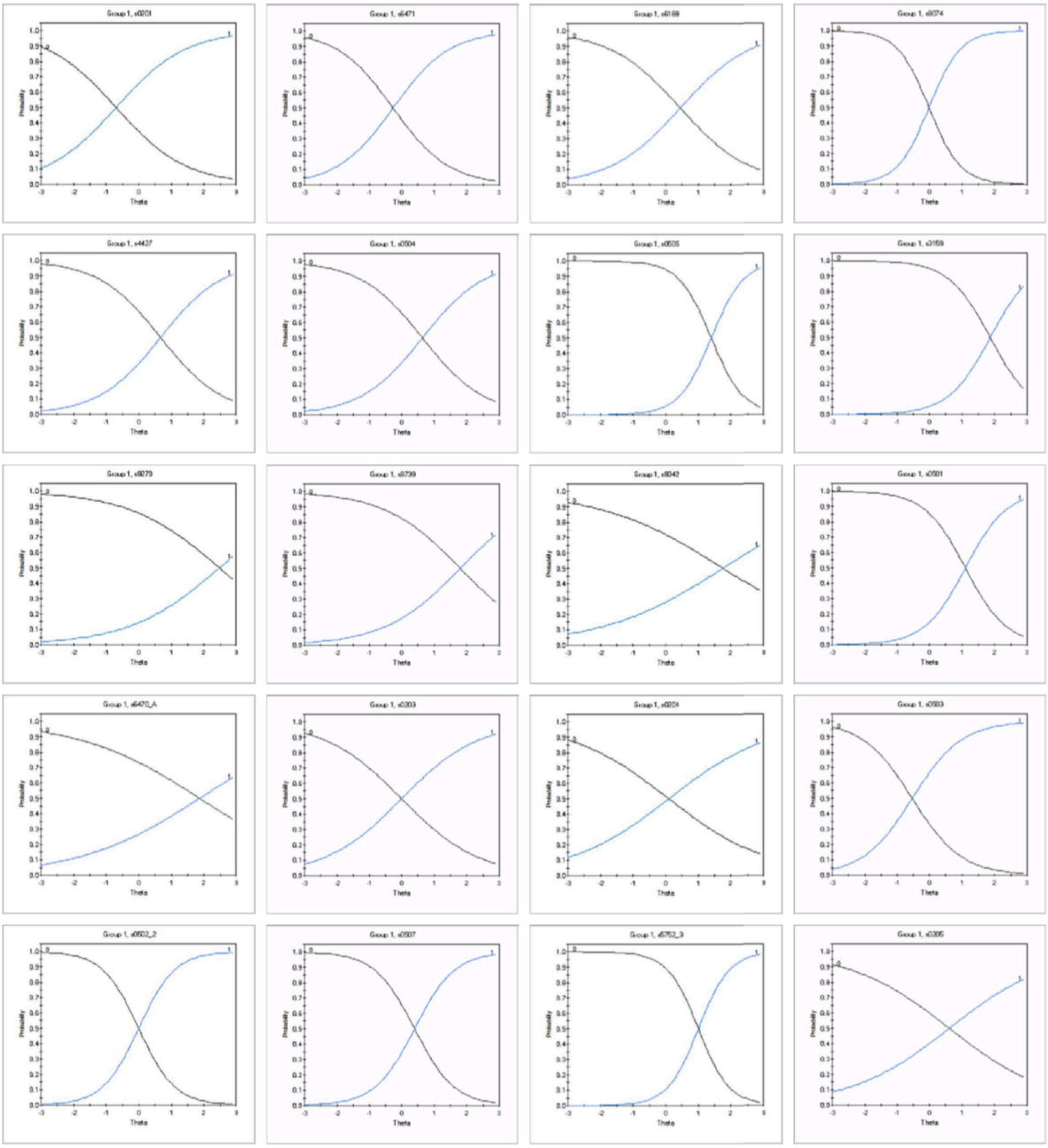
	Matematiksel Yeterlikler				**Kazanımlar**					**Hatalar**		
	M1	**M2**	**M3**	**M4**	**K1**	**K2**	**K3**	**K4**	**K5**	**E1**	**E2**	**E3**
5017_A	1	0	0	1	1	1	0	0	0	0	1	0
5017_B	0	0	1	1	1	0	0	0	0	0	1	1
5036_A	0	0	1	1	0	1	0	0	0	1	1	0
5036_B	0	1	1	0	0	1	0	0	0	1	1	0
5036_C	0	1	1	0	0	1	0	0	0	1	0	0
5058_A	0	0	1	0	0	0	0	0	1	1	0	0
5058_B	0	0	1	0	0	0	1	0	1	1	1	0
5059	0	1	0	0	0	0	0	1	0	0	0	1
5057	0	0	1	1	0	0	0	0	1	1	0	0

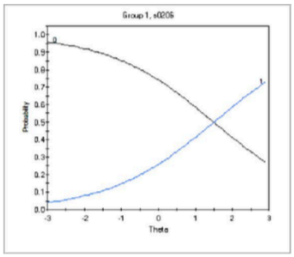
EK 21'de M1, M2, M3 ve M4 olarak isimlendirilen matematiksel yeterlikler sırasıyla "İletişim ve İlişkilendirme", "Matematikleştirme", "Akıl Yürütme ve Strateji Geliştirme" ve "Sembolik ve Teknik Dil Kullanımı" şeklindedir. K1, K2, K3, K4 ve K5 olarak isimlendirilen kazanımlar sırasıyla "Cebirsel İfade oluşturma", "Alan bilgisi uygulama", "Hacim bilgisi uygulama", "Kavramsal Oran Bilgisi" ve "Hacim bilgisi uygulama" şeklindedir. Son olarak E1, E2 ve E3 olarak isimlendirilen hatalar ise "Matematiksel Durumu Anlama Güçlüğü", "Dönüştürme Güçlüğü" ve "İşlem Güçlüğü" şeklindedir.

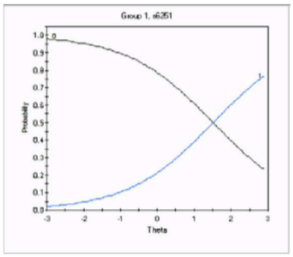
EK 22. Son test uygulaması MTK madde parametreleri

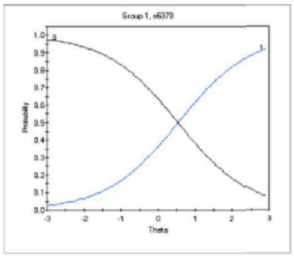
Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	s0201	2	0.93	0.05	1	0.65	0.04	-0.69	0.05
2	s6471	4	1.16	0.06	3	0.31	0.04	-0.27	0.04
3	s6188	6	0.93	0.05	5	-0.42	0.04	0.45	0.05
4	s9074	8	2.06	0.09	7	0.03	0.05	-0.02	0.03
5	s4427	10	1.05	0.05	9	-0.71	0.04	0.67	0.05
6	s0504	12	1.05	0.05	11	-0.67	0.04	0.64	0.04
7	s0505	14	2.02	0.10	13	-2.84	0.10	1.40	0.04
8	s3159	16	1.56	0.08	15	-2.92	0.10	1.87	0.07
9	s9270	18	0.72	0.05	17	-1.79	0.05	2.48	0.16
10	s9739	20	0.86	0.05	19	-1.55	0.05	1.80	0.10
11	s9342	22	0.54	0.04	21	-0.95	0.04	1.76	0.14
12	s0501	24	1.59	0.07	23	-1.76	0.06	1.11	0.04
13	s6470_A	26	0.54	0.04	25	-1.01	0.04	1.87	0.15
14	s0203	28	0.85	0.05	27	0.01	0.04	-0.01	0.04
15	s0204	30	0.65	0.04	29	-0.07	0.04	0.10	0.05
16	s0503	32	1.32	0.06	31	0.70	0.05	-0.53	0.04
17	s0502_2	34	1.80	0.08	33	0.00	0.05	-0.00	0.03
18	s0507	36	1.63	0.07	35	-0.64	0.05	0.40	0.03
19	s5752_3	38	2.12	0.10	37	-2.14	0.08	1.01	0.03
20	s0205	40	0.66	0.04	39	-0.40	0.04	0.61	0.06
21	s0206	42	0.70	0.05	41	-1.05	0.04	1.50	0.10
22	s6251	44	0.86	0.05	43	-1.30	0.05	1.52	0.08
23	s6370	46	1.03	0.05	45	-0.56	0.04	0.54	0.04
24	s6323_1	48	0.60	0.05	47	-1.41	0.04	2.36	0.17
25	s3281	50	1.13	0.06	49	-1.15	0.05	1.01	0.05
26	s1119	52	0.60	0.04	51	-0.88	0.04	1.46	0.11
27	s8867	54	2.17	0.10	53	-2.00	0.08	0.92	0.03
28	s5050_A1	56	1.35	0.06	55	-0.47	0.04	0.35	0.03
29	s6963_B	58	0.77	0.05	57	-0.63	0.04	0.82	0.06
30	s7305_A	60	2.44	0.16	59	-4.97	0.24	2.03	0.06

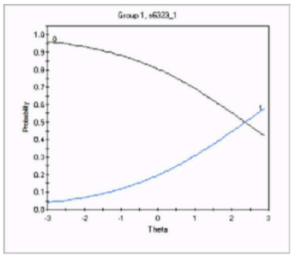
EK 23. Son test uygulaması Madde Karakteristik eğrileri

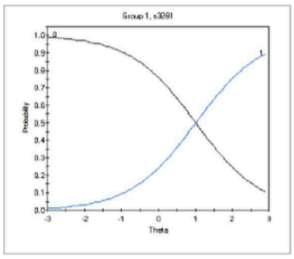


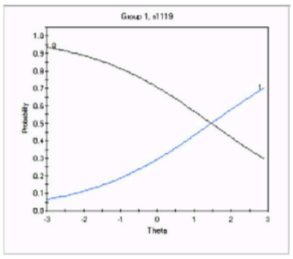

Group 1, s0205
Probability
Theta

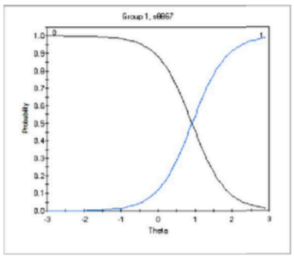

Group 1, s6351
Probability
Theta

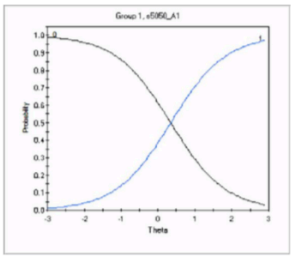

Group 1, s6370
Probability
Theta

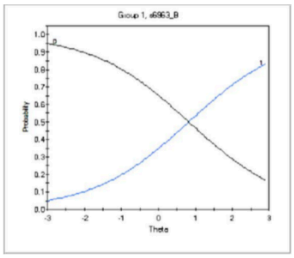

Group 1, s6323_1
Probability
Theta

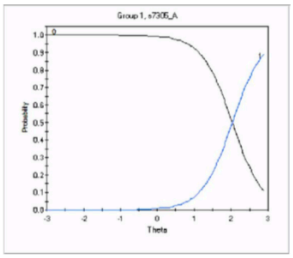

Group 1, s3091
Probability
Theta


Group 1, s1119
Probability
Theta

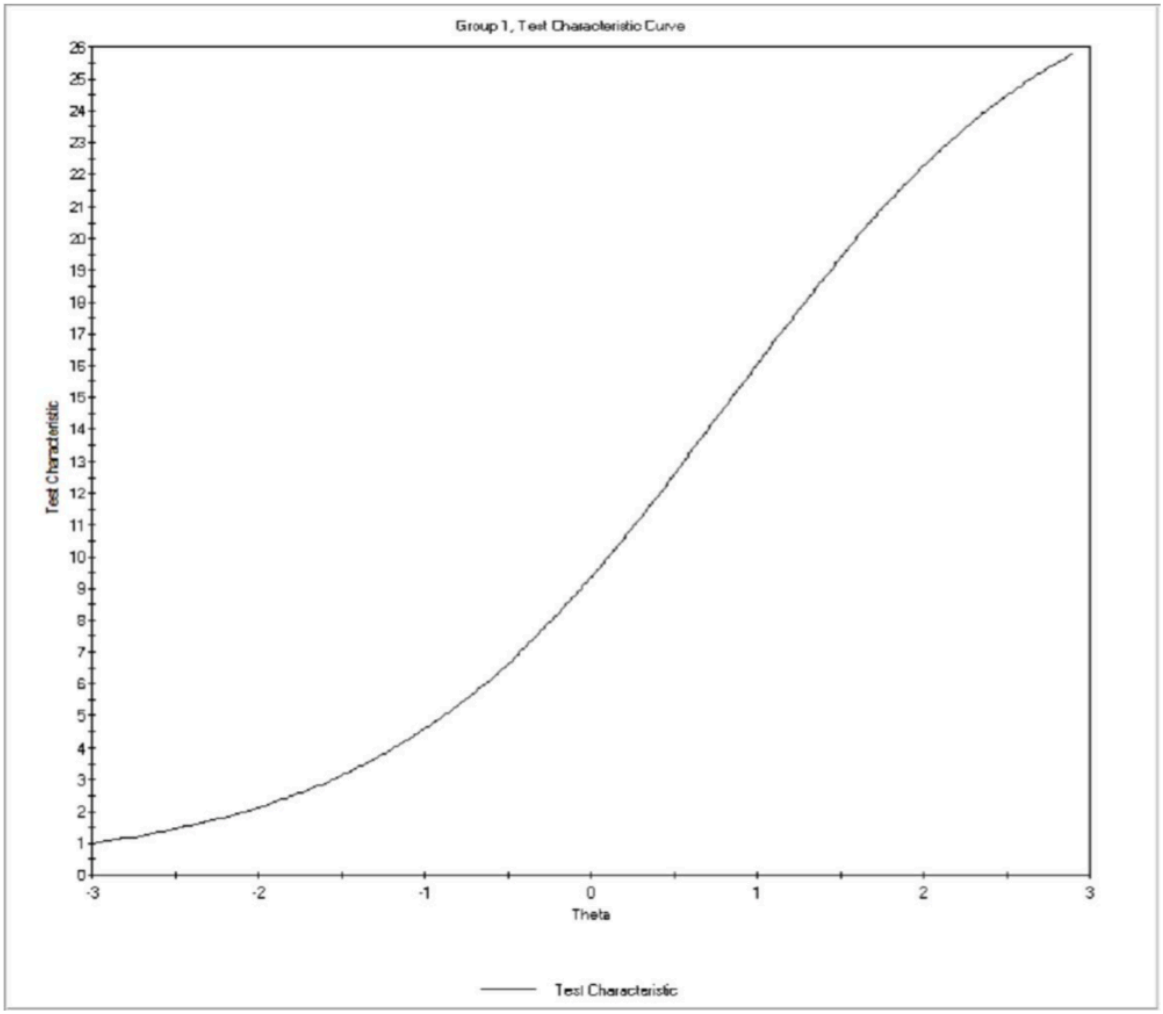

Group 1, s6867
Probability
Theta


Group 1, s5950_A1
Probability
Theta

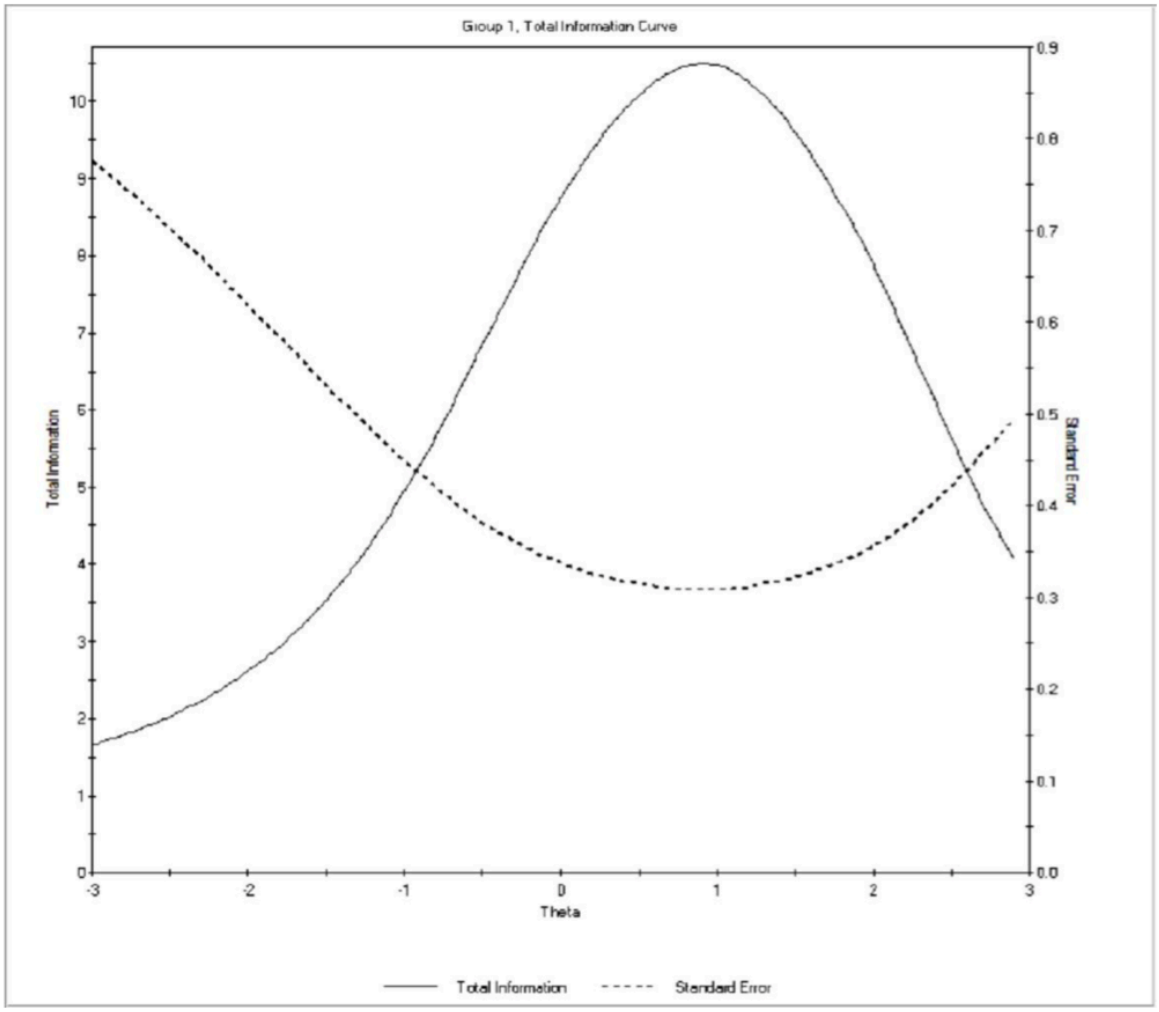

Group 1, s6963_B
Probability
Theta


Group 1, s7305_A
Probability
Theta

EK 24. Son test uygulaması Test karakteristik Eğrisi



EK 25. Son test uygulaması Test Bilgi Fonksiyonu



EK 26. Ön-Son test Matematiksel Yeterlikler Q matrisi

Madde	M1	M2	M3	M4
1	0	1	0	0
2	0	0	0	1
3	0	1	1	0
4	1	0	0	1
5	0	1	0	1
6	0	1	0	1
7	0	1	1	0
8	0	0	0	1
9	0	1	0	1
10	1	1	0	1
11	0	1	0	0

12	1	0	1	0
13	0	1	0	0
14	0	1	0	1
15	1	1	1	0
16	1	1	1	0
17	1	1	0	0
18	0	1	0	0
19	0	1	1	0
20	1	1	1	0
21	1	0	1	0
22	1	0	1	0
23	0	0	1	0
24	1	1	1	1
25	1	0	0	1
26	1	0	0	1
27	0	1	0	1
28	0	1	1	0
29	0	1	1	0
30	0	1	1	0

EK 27. Deney-1 Grubu Ön-son Test Öğrenci Örtük Sınıf Dağılım Matrisi

	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	64	5	6	3	2	6	7	7	1	1	1	2	1	1	1	1
0001	38	9	1	5	2	8	8	3	8	4	1	0	1	1	0	0
0010	15	2	11	0	4	3	2	2	3	1	4	1	1	1	3	1
0011	11	13	48	1	4	3	5	8	2	3	3	0	1	1	2	1
0100	16	4	6	3	4	2	2	3	4	2	1	2	1	2	1	6
0101	1	2	5	1	12	0	0	2	0	1	2	2	0	0	0	3
0110	12	2	7	2	1	6	3	6	6	4	2	5	3	2	1	0
0111	4	1	5	11	0	4	5	11	3	3	6	2	7	1	5	2
1000	22	2	11	2	1	1	2	2	22	5	4	1	3	0	0	1
1001	15	3	7	2	2	1	1	3	3	5	3	0	1	2	1	0
1010	7	0	7	1	1	1	4	2	2	4	1	2	1	2	0	0
1011	5	2	5	5	2	2	1	1	5	2	6	2	0	2	0	6
1100	13	4	4	0	2	0	1	0	3	2	3	1	0	1	1	1
1101	4	0	1	0	2	0	0	1	0	11	1	1	1	0	1	3
1110	8	3	5	1	4	4	1	5	0	11	11	7	7	11	2	4
1111	11	5	8	10	6	14	16	25	7	12	3	38	10	30	27	149

EK 28. Deney-2 Grubu Ön-son Test Öğrenci Örtük Sınıf Dağılım Matrisi

	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	96	10	13	3	12	3	8	6	11	9	9	8	6	4	5	2
0001	27	5	19	7	3	4	3	5	14	5	3	7	1	3	2	1
0010	14	4	10	5	1	3	2	5	3	3	1	3	2	0	1	3
0011	11	2	9	3	1	1	2	5	2	7	3	4	1	3	3	6
0100	17	3	9	1	6	3	3	2	7	4	1	4	0	2	3	1
0101	2	1	5	7	3	2	1	2	1	1	1	0	2	1	0	3
0110	10	2	14	5	1	5	1	7	3	5	1	3	3	8	2	5
0111	4	1	5	3	2	3	3	5	3	3	3	6	0	7	1	12
1000	18	4	10	2	1	0	1	1	4	3	2	0	0	0	2	2
1001	13	4	6	3	1	2	1	4	3	2	5	3	0	1	0	1
1010	6	1	12	4	2	0	2	1	10	0	2	1	1	2	0	3
1011	7	1	3	2	3	3	4	2	11	2	3	3	3	2	1	7
1100	12	0	4	1	1	0	0	0	3	0	0	1	0	1	0	0
1101	2	3	2	1	2	2	0	2	9	0	1	2	0	1	0	1
1110	3	1	3	5	1	4	9	6	13	7	1	4	2	5	1	17
1111	9	6	6	7	9	9	11	26	2	9	3	10	10	11	4	91

EK 29. Kontrol Grubu Ön-son Test Öğrenci Örtük Sınıf Dağılım Matrisi

	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	109	12	13	7	15	10	9	4	10	8	8	7	1	8	4	6
0001	34	5	14	4	9	3	3	6	5	7	3	3	1	5	3	5
0010	21	4	12	2	4	4	2	3	10	7	4	0	2	2	2	4
0011	6	3	5	8	4	2	2	3	2	2	7	2	2	3	2	3
0100	13	4	7	1	9	1	1	1	1	2	0	6	3	1	2	1
0101	3	1	2	2	1	8	2	0	0	1	1	0	0	1	1	1
0110	7	3	3	3	5	7	11	7	0	6	3	3	0	8	2	8
0111	3	3	4	4	2	3	3	16	2	2	2	3	3	5	2	14
1000	10	4	15	1	0	1	2	2	18	0	1	1	1	2	0	0
1001	11	0	8	1	1	0	1	0	3	12	3	1	0	0	1	4
1010	5	3	3	1	4	0	4	1	0	5	6	3	1	1	0	1
1011	3	0	5	4	2	2	1	4	0	3	1	6	0	4	3	9
1100	7	1	1	0	1	1	1	0	0	1	3	0	10	3	0	1
1101	1	0	1	4	0	0	0	2	0	2	2	2	0	0	3	4
1110	2	0	1	2	2	5	4	6	0	1	1	3	0	3	4	8
1111	2	1	7	4	5	5	9	21	2	3	2	7	1	11	7	96

TÜBİTAK
PROJE ÖZET BİLGİ FORMU

Proje Yürütücüsü:	Doç. Dr. TAHSİN OĞUZ BAŞOKÇU
Proje No:	115K531
Proje Başlığı:	Uluslararası Geniş Ölçekli Sınavlarda Türkiye'nin Matematik Başarısını Arttırabilmek İçin Bir Model Önerisi: Bilişsel Tanıya Dayalı İzleme Modelinin Etkililiği
Proje Türü:	3501 - Kariyer
Proje Süresi:	30
Araştırmacılar:	BÜNYAMİN YURDAKUL, GÖZDE AKYÜZ
Danışmanlar:	VEFA BARDAKCI, TUNCAY ÖĞRETMEN, ERDİNÇ ÇAKIROĞLU
Projenin Yürütüldüğü Kuruluş ve Adresi:	EGE Ü.
Projenin Başlangıç ve Bitiş Tarihleri:	15/11/2015 - 15/11/2018
Onaylanan Bütçe:	331575.0
Harcanan Bütçe:	262940.11
Öz:	Bu projenin temel amacı, Uluslararası geniş ölçekli sınavlarda (PISA, TIMSS vb.) Türkiye?nin matematik başarısını arttırabilmek için ortaöğretim 6. sınıf matematik dersindeki alan bilgisi ve bilişsel süreç becerilerinin bir arada ölçülebilmesine dayalı ölçme araçları tasarlamak, öğrencilerin bu bilgi ve becerilere sahip olma düzeylerini etkili istatistiksel yöntemlerle belirlemek, belirlenen öğrenme yetersizlikleri ve kaynakları konusunda öğretmen ve öğrencilere geribildirimler sunarak başarının gelişimini izlemektir. Proje, deneysel desen kullanılarak yürütülmüştür. Tabakalı, tesadüfi küme örnekleme yöntemi kullanılarak İzmir?in 10 ilçesindeki 20 okul ve 148 şubesinden seçilen toplam 4562 öğrenci araştırmanın örneklemi olarak belirlenmiştir. Desen gereği, belirlenen örneklem Deney-1, Deney-2 ve Kontrol gruplarına ayrılmıştır. Deney-1 grubu 8 okulun, 51 şubesindeki toplam 1481 öğrenciden; Deney-2, 5 okulun 42 şubesindeki toplam 1433 öğrenciden; Kontrol grubu ise 7 okulun 55 şubesindeki toplam 1678 öğrenciden oluşmuştur. Deney-1 öğrencilerine izleme testleri yanında BTM analizlerine göre kazanım, bilişsel süreç ve hata konularında detaylı geribildirim verilmiştir. Deney-2, test etkisinin belirlenebilmesi için oluşturulduğundan bu gruptaki öğrencilere izleme testleri uygulanmış ancak sadece toplam puan üzerinden geribildirim verilmiştir. Kontrol grubu öğrencilerine ise sadece ön ve son-testler uygulanmıştır. Deney-1 ve Deney-2, BTM?ye göre geribildirim vermenin; Kontrol grubu hem izleme testlerinin hem de geribildirimin etkisini belirlemek için oluşturulmuştur. Tüm gruplara PISA testi ile eşdeğerliği kanıtlanan ön test, deney süreci sonunda son test olarak tekrar uygulanmıştır. Sonuçlar, araştırmanın temel hipotezinin kanıtlandığını göstermiş; Deney-1 ile Deney-2 ve Deney-2 ile Kontrol grubu arasında son test ortalamalarında anlamlı farklar saptanmıştır. PISA, TIMSS gibi uluslararası geniş ölçekli sınavlarda ölçülen özelliklere sahip olma düzeyinin üst düzey düşünme becerilerini ölçen testlerin kullanılarak arttırılabileceği görülmüştür. Yanı sıra, BTM ile belirlenen öğrenci profilleri kullanılarak sunulan geribildirimlerin başarı düzeyini daha da yükselttiği ortaya çıkmıştır.
Anahtar Kelimeler:	PISA, TIMSS, Bilişsel Tanı Modelleri (BTM), Üst Düzey Düşünme Becerileri, Üstbilişsel İzleme
Fikri Ürün Bildirim Formu Sunuldu Mu?:	Hayır

TÜBİTAK

Projeden Yapılan Yayınlar:	1- Beyond Counting the Correct Responses: Metacognitive Monitoring and Estimations about Test Scores (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 2- A Cognitive Diagnostic Model: High-order Thinking and Metacognition involved in Mathematics (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 3- Student error profile based on Cognitive Diagnosis: Error Q matrix adaptation (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 4- Matematiksel Yeteneklerin Bilişsel Tanı Modellerinde Telafi Edicilik Özelliği Açısından İncelenmesi (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 5- ANALYSIS OF THE DIFFERENTIAL ITEM FUNCTIONS IN THE ITEMS TESTING HIGHORDER THINKING SKILLS IN TERMS OF META-COGNITIVE MONITORING SKILL (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 6- LONGITUDINAL ANALYSIS OF THE STUDENTS' ATTITUDES TOWARDS REAL LIFE PROBLEMS TESTING THEIR HIGH-LEVEL THINKING SKILLS IN MATHEMATICS (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 7- Çevrimiçi Editörlük Sisteminin Test Geliştirme Süreçlerine Uyarlanması (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 8- Comparison of DINA Model and Exploratory Latent Class Analysis (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 9- DINA Modelde Madde Doygunluğunun ve Madde Uyum Düzeyinin Model Veri Uyumuna Etkisi (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 10- Comparison of DINA Model and Exploratory Latent Class Analysis (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 11- DINA Modelde Madde Doygunluğunun ve Madde Uyum Düzeyinin Model Veri Uyumuna Etkisi (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 12- LATENT CLASS EQUATING WITH ANCHOR TEST DESIGN IN COGNITIVE DIAGNOSTIG MODELS (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 13- GENERATION STUDY OF PISA MATHS PROFICIENCYLEVELS IN TURKISH 6TH GRADE STUDENTS (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 14- CLASSIFICATION OF SECONDARY SCHOOL STUDENTS INTERMS OF THEIR LEVELS OF HAVING PISA MATHEMATICAL COMPETENCIES THROUGH COGNITIVE DIAGNOSTICMODELS: A LONGITUDINAL REAL DATA STUDY (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 15- Evaluation of Item Writing Program Related to Higher Order Thinking Skills for Secondary Math Teachers (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 16- INVESTIGATION INVARIANCE OF THE DINA MODEL PARAMETERS WITH REAL DATA (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 17- YAPILANDIRILMIŞ VE ÇOKTAN SEÇMELİ MADDELERDE PARAMETRE DEĞİŞMEZLİĞİNİN İNCELENMESİ: BİLİŞSEL TANI MODELLERİNDE MADDE FORMATININ ETKİSİ (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 18- EXAMINATION OF CDM ATTRIBUTE PREVALENCE FOR HIGHER-ORDER THINKING SKILLS ITEMS (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 19- BTM'de Ankor Test Deseni İle Gözlenen Puanlara Dayalı Lokal Eşitleme Çalışması. (Bildiri - Ulusal Bildiri - Sözlü Sunum), 20- Hakem Katılığının Yapılandırılmış-Yanıtlı Maddelerde Yetenek ve Parametre Kestirimine Etkisi (Bildiri - Ulusal Bildiri - Sözlü Sunum), 21- Test Geliştirme Sürecinde Madde Seçimi İçin Hiyerarşik Puanlayıcı Model: Sinyal Algılama Puanlayıcı Modelinin Uyarlanması (Bildiri - Ulusal Bildiri - Sözlü Sunum), 22- BTM'de Ankor Test Deseni İle Gözlenen Puanlara Dayalı Lokal Eşitleme Çalışması (Bildiri - Ulusal Bildiri - Sözlü Sunum), 23- Hakem Katılığının Yapılandırılmış-Yanıtlı Maddelerde Yetenek ve Parametre Kestirimine Etkisi. (Bildiri - Ulusal Bildiri - Sözlü Sunum), 24- Test Geliştirme Sürecinde Madde Seçimi İçin Hiyerarşik Puanlayıcı Model: Sinyal Algılama Puanlayıcı Modelinin Uyarlanması. (Bildiri - Ulusal Bildiri - Sözlü Sunum), 25- Uluslararası Geniş Ölçekli Sınavlarda Türkiye'nin Matematik Başarısını Arttırabilmek İçin Bir Model Önerisi: Bilişsel Tanıya Dayalı İzleme Modelinin Etkililiği-115K531- TÜBTAK (Bildiri - Ulusal Bildiri - Poster Sunum), 26- Ege Üniversitesi,Tübitak Ve İzmir İl Milli Eğitim Müdürlüğü İşbirliği İle Yapılan Proje (Yayılım - Ulusal - Medyada Haber), 27- "En iyi Soru Nasıl Yazılır" (Yayılım - Ulusal - Medyada Haber), 28- Matematik öğretmenleri Ege'de buluşacak (Yayılım - Ulusal - Medyada Haber), 29- Öğretmenler soru yazma becerilerini geliştirdi (Yayılım - Ulusal - Medyada Haber), 30- Matematik öğretmenlerine Düşünme Becerilerine Yönelik Soru Yazma Eğitimi (Yayılım - Ulusal - Medyada Haber),

TÜBİTAK

	31- Türkiye'nin Matematik Öğretmenleri İzmir'de Buluştu (Yayılım - Ulusal - Medyada Haber), 32- Türkiyenin matematik öğretmenleri İzmirde (Yayılım - Ulusal - Medyada Haber), 33- Türkiye'nin matematik öğretmenleri İzmir'de bir araya geldi (Yayılım - Ulusal - Medyada Haber), 34- Türkiye'nin Matematik Öğretmenleri İzmir'de Bir Araya Geldi (Yayılım - Ulusal - Medyada Haber), 35- PISA ve TIMSS testlerinde başarılı olabilir miyiz? (Yayılım - Ulusal - Medyada Haber), 36- PISA ve TIMSS testlerinde başarılı olabilir miyiz? (Yayılım - Ulusal - Medyada Haber), 37- İşte Türk akademisyenler işte milli proje (Yayılım - Ulusal - Medyada Haber), 38- Türkiye'nin Matematik Başarısını Artırabilmek İçin Bir Model Önerisi (Yayılım - Ulusal - Medyada Haber), 39- Ege Üniversitesi,Tübitak Ve İzmir İl Milli Eğitim Müdürlüğü İşbirliği İle Yapılan Proje (Yayılım - Ulusal - Medyada Haber), 40- Türkiye'nin Matematik Başarısını Artırabilmek İçin Bir Model Önerisi (Yayılım - Ulusal - Medyada Haber), 41- İşte Türk akademisyenler işte milli proje (Yayılım - Ulusal - Medyada Haber), 42- Ortaokul Matematik Öğretmenleri İçin Üst Düzey Düşünme Becerilerine Yönelik Soru Yazma Eğitimi-2 (Yayılım - Ulusal - Eğitim/Kurs Verilmesi), 43- Ortaokul Matematik Öğretmenleri İçin Üst Düzey Düşünme Becerilerine Yönelik Soru Yazma Eğitimi (Yayılım - Ulusal - Eğitim/Kurs Verilmesi), 44- Test Geliştirme Sürecinde Madde Seçimi İçin Hiyerarşik Puanlayıcı Model: Sinyal Algılama Puanlayıcı Modelinin Uyarlaması (Tez (Araştırmacı Yetiştirilmesi) - Yüksek Lisans Tezi),

TÜBİTAK