

**PROBLEM ÇÖZME BECERİLERİNİN DEĞERLENDİRİLMESİNDE
ÖZ, AKRAN VE ÖĞRETMEN PUANLARININ
ÇOK YÜZEYLİ RASCH ÖLÇME MODELİ İLE İNCELENMESİ**

SEYHAN SARITAŞ

**YÜKSEK LİSANS TEZİ
EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

EYLÜL, 2015

TELİF HAKKI ve TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren(....) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı: Seyhan

Soyadı: Sarıtaş

Bölümü: Eğitimde Ölçme ve Değerlendirme

İmza:

Teslim tarihi:

TEZİN

Türkçe Adı: Problem Çözme Becerilerinin Değerlendirilmesinde Öz, Akran Ve Öğretmen Puanlarının Çok Yüzeyle Rasch Ölçme Modeli İle İncelenmesi

İngilizce Adı: Investigation Of Self, Peer And Teacher's Scores With Many-Facet Rasch Measurement Model In Problem Solving Skills Assessment

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduđumu, yararlandıđım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiđimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduđunu beyan ederim.

Yazar Adı Soyadı: Seyhan SARITAŞ

İmza :

Jüri onay sayfası

Seyhan Sarıtaş tarafından hazırlanan “Problem Çözme Becerilerinin Değerlendirilmesinde Öz, Akran Ve Öğretmen Puanlarının Çok Yüzeyle Rasch Ölçme Modeli İle İncelenmesi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

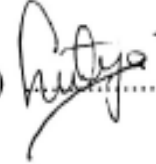
Danışman: Doç. Dr. İsmail KARAKAYA

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi)



Başkan: Prof. Dr. Hülya KELECİOĞLU

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Hacettepe Üniversitesi)



Üye: Doç. Dr. Bayram ÇETİN

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi)



Tez Savunma Tarihi: 17/09/2015

Bu tezin Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Tahir ATICI

Eğitim Bilimleri Enstitüsü Müdürü

.....



Canım Aileme...

TEŞEKKÜR

Ülkemizde yetişmiş uzman ve akademisyen sayısının yetersiz olduğu bu değerli alanda Gazi Üniversitesi'nde Eğitimde Ölçme ve Değerlendirme Anabilim Dalı'nın kurucusu olması ve akademik gelişimimiz için emeklerinden dolayı bu bölümden mezun olan ve olacak olan herkes adına mütevazı hocamız Prof. Dr. Şener BÜYÜKÖZTÜRK' e; tez süreci boyunca desteğini esirgemeyen saygıdeğer danışman hocam Doç. Dr. İsmail KARAKAYA'ya görüş ve önerileri ile çalışmama önemli katkılar sağlayan değerli jüri üyelerim Prof. Dr. Hülya KELLEÇİOĞLU ve Doç. Dr. Bayram ÇETİN'e; okul uygulaması için sınıfını açan Bülent DÜŞMEZ'e; tez yazım sürecinde çalışmama çeşitli şekillerde katkıda bulunan Arş. Gör. Sibel ADA, Arş. Gör. Emine Yavuz, Öğr. Gör. Gülsün YILDIRIM'a; bu vesile ile güzel insanlarla dolu anabilim dalımızda bulunan çalışma arkadaşlarıma en içten teşekkürlerimi sunmak isterim.

Sırada gelelim işin duygusal kısmına ☺ hayatım boyunca elime tabiri caizse ev işi değdirmeyen ellerime okumaya verdiği önemden kalemin varlığını yakıştıran biricik anneme, eğitim hayatım ve çalışma hayatım boyunca aldığım tüm radikal kararların arkasında olan ve desteğini yüreğimde hissettiğim canım babama; ismi gibi çok saygın ve yıllarca istediği hitabı söyleyemediğim ama şimdi yazdığım “ağabeyime” ve tez yazma sürecimde yaşını aşan olgunluk sahibi kardeşime can-ı gönülden teşekkürler...

İsmi sayamadığım güzel insanlar iyi ki varsınız!

**PROBLEM ÇÖZME BECERİLERİNİN DEĞERLENDİRİLMESİNDE
ÖZ, AKRAN VE ÖĞRETMEN PUANLARININ
ÇOK YÜZEYLİ RASCH ÖLÇME MODELİ İLE İNCELENMESİ
(Yüksek Lisans Tezi)**

Seyhan SARITAŞ

**GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

Eylül, 2015

ÖZ

Bu araştırmada, rutin olmayan matematik problemi çözme becerilerinin değerlendirilmesi amacıyla hazırlanan performans görevlerinin dereceli puanlama anahtarıyla öz, akran ve öğretmen puanlamaları çok yüzeysel Rasch ölçme modeli ile incelenmiştir. Araştırma sürecinde öz, akran ve öğretmen puanlarının FACET analizi ile değerlendirilmesi ve öz ile akran puanlamalarının öğretmen puanlarına göre farklılaşma durumu incelenmiştir. Araştırmanın ilk aşamasında, 2014-2015 eğitim öğretim yılı bahar döneminde Ankara ilinde bir devlet okulunda öğrenim gören iki şube olmak üzere 75 altıncı sınıf öğrencisine ilk iki haftası eğitim ve deneme haftası olmak üzere toplamda altı hafta süresinde dört performans görevi uygulanmıştır. Belirtilen 75 öğrenciden üç uygulamanın tamamına katılmayan 18 kişi analizden çıkarılmıştır. Kalan 57 öğrencinin üç performans görevindeki performansları, analitik dereceli puanlama anahtarı ile öz, akran ve üç öğretmen tarafından puanlanmıştır. Araştırmanın ikinci aşamasında, devamındaki üç hafta uygulanan performans görevlerinden elde edilen puanların güvenilirliğine ilişkin güvenilirlik katsayısı her uygulama için ayrı bir şekilde kestirilmiştir. Bu uygulamalara ait veriler yüzey olarak; öğrenci, aşama, puanlayıcı türü ve yapay (dummy) yüzey olarak puanlayıcı yüzeyi tanımlanmış, her yüzeye ilişkin güvenilirlik istatistikleri kestirilmiştir. Devamında ise üç uygulamaya ilişkin performans görevi puanlamaları birleştirilmiş ve görev beşinci yüzey olarak tanımlanmıştır. Dolayısıyla tüm veri setinde öğrenci, aşama, puanlayıcı türü, puanlayıcı ve görev olarak beş yüzeye ilişkin güvenilirlik kestirimleri hesaplanmıştır.

Arařtırma sonucunda, üç hafta uygulanan performans görevinden elde edilen puanlamaların çok yüzeyli Rasch ölçme modeline göre elde edilen güvenilirlik katsayılarının her bir uygulama için yüksek olduđu görölmüřtür. Tüm veride görev yüzeyi de tanımlanarak üç uygulamaya ait öz, akran ve öğretmen puanlamalarının FACET ile analizi sonucunda yine yüksek güvenilirliğe sahip olduđu görölmüřtür. Puanlayıcı türü olarak öz, akran ve öğretmen puanlamalarının katılık/cömertliđi ise haftalara göre farklılaşma gösterdiđi görölmüřtür. Genel olarak öz puanlayıcının en cömert, öğretmenin ise katı puanlayıcı olduđu görölmüřtür. Akran puanlayıcının katılık/cömertliđi ise birinci uygulamadan üçüncü uygulamaya dođru öğretmen puanlayıcıların katılık/cömertliđine yaklařtıđı gözlenmiřtir.

Bilim Kodu :

Anahtar Kelimeler: Çok Yüzeyli Rasch Modeli, Öz Deđerlendirme, Akran Deđerlendirme, Problem Çözme Becerisinin Deđerlendirilmesi, Dereceli Puanlama Anahtarı, Puanlama Güvenirliđi

Sayfa Adedi : 112

Danışman : Doç. Dr. İsmail Karakaya

**INVESTIGATION OF SELF, PEER AND TEACHER’S SCORES
WITH MANY-FACET RASCH MEASUREMENT MODEL IN
PROBLEM SOLVING SKILLS ASSESSMENT
(M.S Thesis)**

Seyhan SARITAŞ

**GAZI UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
Sep, 2015**

ABSTRACT

In this study, self, peer and teacher’s scores of performance tasks which were prepared so as to evaluate solving skills of non-routine math problems have been analyzed by multi-faceted Rasch measurement model. In the research process, the assessment of self, peer and teacher scoring with FACET analysis and differentiation conditions between self and peer scoring and teacher scoring were analyzed. In the first phase of the study, through 2014-2015 academic year spring semester, four performance tasks were administered to 75 sixth grade students in a state school in Ankara province in a six-week period including the first two weeks training and testing week. Among 75 students, 18 students could not attend all three practices, so these students were excluded from analysis. The remaining 57 students’ performance in three performance tasks was scored by themselves, peers and three teachers with analytical rubric. In the second phase of study, reliability coefficients related the reliability of the scores obtained from three-week administration of performance tasks were estimated separately for each administration. Data in these administrations, student, steps (problem-solving), raters type as facet and rater as dummy facet were defined, reliability statistics related each facet were estimated. Subsequently, scorings performance task related three-week were combined and “task” was defined as fifth facet. So using whole the data set, reliability estimations related five facets which were student, steps (problem-solving), raters type, raters and the task were calculated. As a result of the research, it has seen that scores obtained from three-week

administration of performance tasks have high reliability coefficients for each administration according to multi-faceted Rasch measurement model. It has been seen that defining the facet of task in all the data set, self, peer and teacher's scorings of three administration as a result of the FACET analysis have still high reliability coefficients. It has been seen that as a type of raters; self, peer and teacher scoring's lenient/severity differentiate according to the weeks. In general, self-rater has been seen as the most lenient and teacher has been the most severe. It has been observed that severity/lenient of peer-raters get closer to severity/lenient of teachers from the first application to the third application.

Science Code:

Key Words: Many-facets Rasch Measurement Model, Self-Assessment, Peer-Assessment, Assessment of Problem Solving Skills, Rubric, Scoring Reliability

Page Number:112

Supervisor: Associate Professor İsmail KARAKAYA

İÇİNDEKİLER

TELİF HAKKI ve TEZ FOTOKOPİ İZİN FORMU.....	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
ÖZ.....	vi
ABSTRACT.....	viii
İÇİNDEKİLER	x
TABLolar LİSTESİ.....	xiii
ŞEKİLLER LİSTESİ	xiv
SİMGELER VE KISALTMALAR LİSTESİ.....	xv
BÖLÜM I.....	1
GİRİŞ	1
1.1.Problem Durumu.....	1
1.2.Kuramsal Çerçeve	6
1.2.1.Problem Çözme Süreci.....	6
1.2.2.Problem Çözme Becerisinin Değerlendirilmesi	9
1.2.3.Puanlayıcı Güvenirliği.....	11
1.3.Araştırmanın Önemi.....	13
1.4.Araştırmanın Amacı	14
1.5.Araştırma Soruları.....	15
1.6.Sınırlılıklar	16
1.7.Varsayımlar	17
1.8.İlgili Çalışmalar	17
1.8.1.Problem Çözme Becerisi Hakkında Yapılan Çalışmalar.....	17
1.8.2.Öz ve Akran Değerlendirme Hakkında Yapılan Çalışmalar	19
1.8.3.Çok Yüzeyle Rasch Modeli ile İlgili Yapılan Çalışmalar	20
BÖLÜM II	27
YÖNTEM.....	27
2.1.Araştırmanın Modeli	27
2.2.Çalışma Grubu	27

2.3. Veri Toplama Araçları.....	29
2.3.1. Performans Görevleri.....	29
2.4. Verilerin Toplanması	32
2.4.1. Program Tanıtım ve Deneme Uygulaması	32
2.4.2. Performansa Dayalı Durum Belirleme Süreci	33
2.5. Verilerin Analizi	33
2.5.1. Çok Yüzeyle Rasch Ölçme Modeli.....	34
BÖLÜM III.....	37
BULGULAR.....	37
3.1. Birinci Alt Probleme İlişkin Bulgular	37
3.2. İkinci Alt Probleme İlişkin Bulgular	44
3.3. Üçüncü Alt Probleme İlişkin Bulgular	50
3.4. Dördüncü Alt Probleme İlişkin Bulgular.....	56
BÖLÜM IV	65
SONUÇ, TARTIŞMA ve ÖNERİLER.....	65
4.1. Sonuçlar ve Tartışma	65
4.2. Öneriler	69
4.2.1. Araştırma Sonuçlarından Çıkarılan Öneriler	69
4.2.2. İleride Yapılabilecek Çalışmalara İlişkin Öneriler	69
KAYNAKLAR	71
EKLER.....	79
EK 1. Uzman Değerlendirme Formları	80
EK 1.a Performans Görevi Uzman Değerlendirme Formu.....	80
EK 1.b Dereceli Puanlama Anahtarı Uzman Değerlendirme Formu	82
EK 2. Uygulanan Performans Görevleri	84
EK 2.a Deneme Haftasında Uygulanan Performans Görevi.....	84
EK 2.b Birinci Uygulamadaki Performans Görevi.....	87
.....	88
EK 2.c İkinci Uygulamadaki Performans Görevi	89
EK 2.c Üçüncü Uygulamadaki Performans Görevi	91
EK 3. Dereceli Puanlama Anahtarları	94
EK 3.a Öğretmene Yönelik Dereceli Puanlama Anahtarı.....	94
EK 3.b Öğrenciye Yönelik Dereceli Puanlama Anahtarı	95



TABLolar LİSTESİ

Tablo 1. Klasik Test Kuramına Ait Bazı Güvenirlik Hesaplama Yöntemleri.....	12
Tablo 2. Çalışma Grubunun Cinsiyetlerine Göre Frekans ve Yüzdeleri.....	28
Tablo 3. $\alpha= 0,05$ Düzeyinde Kapsam Geçerlik Oranları İçin Minimum Değerler.....	31
Tablo 4. Performans Görevleri Kapsam Geçerliği.....	31
Tablo 5. Dereceli Puanlama Anahtarı Kapsam Geçerliği.....	32
Tablo 6. Birinci Uygulamada Problem Çözme Aşamalarına Ait Ölçüm Raporu.....	42
Tablo 7. Birinci Uygulamada Puanlayıcı Türüne Ait Ölçüm Raporu.....	43
Tablo 8. İkinci Uygulamada Aşamalara Ait Ölçüm Raporu.....	48
Tablo 9. İkinci Uygulamada Puanlayıcı Türüne Ait Ölçüm Raporu.....	49
Tablo 10. Üçüncü Uygulamada Aşamalara Ait Ölçüm Raporu.....	54
Tablo 11. Üçüncü Uygulamada Puanlayıcı Türüne Ait Ölçüm Raporu.....	55
Tablo 12. Üç Uygulamanın Verilerini İçeren Analizde Aşamalara Ait Ölçüm Raporu.....	60
Tablo 13. Üç Uygulamanın Verilerini İçeren Analizde Puanlayıcı Türüne Ait Ölçüm Raporu	61
Tablo 14. Üç Uygulamanın Verilerini İçeren Analizde Görevlere Ait Ölçüm Raporu.....	62

ŞEKİLLER LİSTESİ

Şekil 1. Problem durumları ile yaş ilişkisi	8
Şekil 2. Birinci analizdeki uygulamaya ait kalibrasyon haritası	38
Şekil 3. Birinci uygulamada öğrencilere ait ölçüm raporu	40
Şekil 4. Analizdeki uygulamaya ait kalibrasyon haritası	45
Şekil 5. İkinci uygulamada öğrencilere ait ölçüm raporu	47
Şekil 6. Üçüncü analizdeki uygulamaya ait kalibrasyon haritası	51
Şekil 7. Üçüncü uygulamada öğrencilere ait ölçüm raporu	53
Şekil 8. Üç uygulamanın verilerini içeren kalibrasyon haritası	57
Şekil 9. Üç uygulamanın verilerini içeren analizde öğrencilerin ölçüm raporu	59

SİMGELER VE KISALTMALAR LİSTESİ

ÇYRÖM	Çok Yüzeyle Rasch Ölçme Modeli
DPA	Dereceli Puanlama Anahtarı
F	Frekans
G	Ayırma İndeksi
GK	Genellenebilirlik Kuramı
KTK	Klasik Test Kuramı
KGİ	Kapsam Geçerlik İndeksi
KGO	Kapsam Geçerlik Oranı
KGÖ	Kapsam Geçerlik Ölçütü
MTK	Madde Tepki Kuramı
MEB	Milli Eğitim Bakanlığı

BÖLÜM I

GİRİŞ

Bu bölümde problem durumu, kuramsal çerçeve, araştırmanın önemi, araştırmanın amacı, araştırmanın soruları, sınırlılıkları, varsayımları ve ilgili çalışmalar sunulmuştur.

1.1.Problem Durumu

Bilim ve teknolojinin hızlı bir şekilde gelişmesi, yaşamın çoğu alanında olduğu gibi eğitim alanında da birçok değişime kaynaklık etmiştir. Modern çağın ihtiyaçlarına yanıt verebilecek düzeyde bireylerin yetiştirilebilmesi için eğitim sistemlerinin üst düzey zihinsel becerilere büyük önem vermesi gerekmektedir (Kutlu, Doğan ve Karakaya, 2010, s.15). Buna göre; öğrendiği bilgiyi yeni bilgilere ulaşmak için kullanan, muhakeme eden, olayları derinlemesine analiz eden, eleştirel düşünebilen, üretken, dinamik, özgüveni yüksek, bilimsel düşünme süreçlerini yaşayan, problem çözme becerisi gibi zihinsel becerileri kullanan ve geliştirebilen bireylerin yetiştirilmesi amaçlanmaktadır (Orbeyi ve Güven, 2008; Şanlı, 2005). Evrensel değerler haline gelen bu niteliklerin en önemlilerinden birisi de problem çözme becerisine sahip olmaktır (Şanlı, 2005).

Bireyin ve toplumun ne zaman, hangi güçlüklerle karşılaşacağı ve ihtiyaçların neler olacağı önceden bilinmemektedir. Dolayısıyla, karşılaşacağı sorunların üstesinden gelebilecek bireylerin yani, problem çözme becerisi gelişmiş bireylerin yetiştirilmesi çağdaş eğitim ve öğretimin davranışsal düzeydeki en önemli amaçlarından biri haline gelmiştir (Altun, 2002; Topses, 2003). Problem çözme becerisi gelişmiş birey, karşılaştığı güçlüklerde hemen yardım almak ya da başkalarının karar vermelerini beklemek yerine mevcut bilgi ve becerilerini kullanarak probleme ilişkin olası çözüm yollarını oluşturur (Arslan, 2012).

Öğrenciler düşüncelerini problem çözme sürecine yansıtılmaları için teşvik edildiği takdirde okul dışındaki hayatlarında tanınmayan durumlarla karşılaştıklarında kendine duyduğu güven daha fazla olur. Öğrenciler problem durumlarını çözüme kavuşturdukça farklı düşünme yollarını keşfeder, merak duyar, kendine olan güvenleri artar, daha sabırlı olur, yaratıcı bir tutum içine girer ve bir işin devamını getirme alışkanlığını kazanır (MEB, 2009; NCTM, 2000). Bu açılardan önemi yadsınamayacak olan problem çözme becerisi ile ilişkili olan problem, problem çözme ve problem çözme süreci kavramlarının iyi anlaşılması sürecin değerlendirilmesi açısından da büyük önem taşımaktadır.

Problem çözme sürecinde basamakların bilinmesi ve bunlara uygun çalışma biçimi seçilmesi problem çözmeyi kolaylaştırır. Fakat çözümü sağlamaz. Çünkü sürecin başındaki kişinin birinci basamakta ne istenildiğinin bilmesi, ikinci basamakta hangi stratejiyi seçeceği yine çözen kişiye kalmaktadır (Altun, 2002). Genellikle bir konuyu bilen bir bireyin, o konu ile ilgili karşısına çıkan problemlerin çözümü için yeterli donanıma sahip olduğu düşünülür. Fakat çözüme kavuşamamasının sebebi çoğu zaman, bilgi eksikliğinden değil, bilgiyi etkin kullanamama, yani bireyin problem çözme becerisinin gelişmemiş olmasından kaynaklanır (Yıldırım, 2003).

OECD (2012)'e göre öğrencilerin problem çözümedeki performanslarına yönelik bazı önemli bulgular aşağıda verilmiştir:

- Birçok ülkede öğrencilerin %10'dan fazlası temel problemlerin çözümünü gerçekleştirememektedir.
- OECD ülkelerinde üst performans grubunda (5. seviye ve üzeri) yer alan öğrencilerin ortalaması yaklaşık %11 civarında kaldığı görülürken asgari performans düzeyine (2. seviye altı) ulaşamayan öğrencilerin ortalaması ise yaklaşık %21,4 olarak belirlenmiştir.
- OECD ülkelerinde ortalama öğrencilerin yarısı basit problemlerden biraz daha zor olan problemleri çözememektedir.

Bu araştırmaya katılan ülkeler çerçevesinde genel olarak sonuçlar, matematik programlarının en temel amaçlarından olan ve farklı disiplinler ile ilişkilendirilen problem çözme becerisinin henüz istenilen seviyede kazanılmadığını ortaya koymaktadır.

Okullarda matematik dilinin yaşamla bağlantısı iyi kurulamaz ise gerekli ilişki sağlanamaz ve matematik yaşamdan kopar (Umay, 1996). Ülkemizde de öğrencilerin okul öğrenmelerinin günlük yaşamda karşılaştıkları problemlerin çözümünde kullanmaları

açısından önemli eksiklikler olduğu görülmüştür (Gelbal ve Kelecioğlu, 2007). Bu yüzden öğrencilere sunulan problem durumlarının gerçek yaşam durumlarına dayalı (otantik) ya da yaşamda gerçekleşmemiş ama gerçekleşme olasılığı olan durumların kullanılması öğrencinin okulda öğrendiğini gerçek hayatında kullanabilmesi açısından önem taşımaktadır. Fakat öğrencilere sadece problem çözme hakkında fırsatlar sunmak da yeterli değildir (Bingham, 2004, s.38). Öğrencilerin farklı yöntemler kullanarak problem çözme becerilerini artırmaları için zengin problem durumlarının olduğu, öğrencilerin keşif ve araştırma yapabilecekleri sınıf ortamlarının sağlanması gerekmektedir (MEB, 2009; Toluk ve Olkun, 2002).

Birey tüm eğitim düzeylerinde ve yaşam boyu problem çözme becerisine ihtiyaç duyar. Fakat ilköğretim döneminin, çocukların zihinsel gelişimi açısından hızlı olduğu yıllar olması sebebi ile bu dönemde uygun yaklaşımlarla problem çözme becerisinde hızlı bir gelişme sağlanabilir (Baykul, 2006, s.61). Dolayısı ile eğitim etkinliklerinde, problem çözmenin de içinde bulunduğu üst düzey zihinsel becerilerin geliştirmesine ışık tutacak öğretim yöntem ve tekniklerin yanında buna bağlı olarak bu becerileri daha etkili olarak ölçen ölçme ve değerlendirme yaklaşımlarının kullanılması gerekmektedir (Kutlu vd. 2010, s.15-16). İlköğretimdeki değerlendirmelerin esas amacının daha çok öğrencilerin izlemelerine yönelik olduğu (Baykul, 2006, s.87) göz önünde bulundurulursa öğrenme sürecinden bağımsız olmayan öğrenciye zengin dönütler sağlayabilen performansa dayalı durum belirleme çalışmalarının öneminin büyük olduğu düşünülmektedir. Fakat okullarda öğretmenlerin bilgi eksikliği, isteksizliği, müfredatı yetiştirmek için zaman bulamaması, sınıfların kalabalık olması, kaynak ve araç gereç yetersizlikleri gibi sebeplerden dolayı matematik programında yer alan performansa dayalı ölçme-değerlendirme uygulamalarının sağlıklı gerçekleştirilemediği bilinmektedir. (Bal ve Doğanay, 2010; Gelbal ve Kelecioğlu, 2007).

Problem çözme sürecinin belirli aralıklarla gerçekçi bir şekilde değerlendirilerek çözümün ne kadar etkin olduğu ve yeni problemlerin ortaya çıkıp çıkmadığının belirlenmesi çok önemlidir. Bireyin “Bu çözüm yolunu nasıl bulduk?” sorusunu kendine sorması, uygulamanın aksayan yönlerini belirlemesine ve süreçte izlenen adımları çözümlemesine yardımcı olur. Bireyi işlemlerin sırasını ve organizasyonunu değerlendirmeye sevk ederek bireyin uygulama esnasında ortaya çıkan problemleri fark etmesi ve çözüme kavuşturması sağlanır. Sürecin bu şekilde ilerlemesi bireyin gelecekteki benzer problemleri daha kolay çözmesine ve problemlerdeki çözüm yollarının nitelik ve nicelik açısından iyileştirmesine

yardımcı olacaktır. (Bingham, 2004, s.38; Yıldırım, 2003, s.147-148). Ayrıca bireylerin problemleri farklı yollardan çözebileceği ve problem çözme ile ilgili fikirlerini akran ve öğretmenleri ile rahatlıkla paylaşabileceği öğrenme ortamları oluşturulmalıdır (MEB, 2009).

Öğrencilerin problem çözme gibi üst düzey becerilerini ölçmek için kullanılan açık uçlu etkinliklerin ve performansa dayalı öğrenci çalışmalarının değerlendirilmesi yapılırken olabildiğince nesnel bir ölçme yapılması çoğu zaman problem oluşturabilmektedir. Yeni değerlendirme yaklaşımları kapsamında, dereceli puanlama anahtarları (DPA) öğrenci çalışmalarının daha güvenilir ve yansız puanlanabilmesi amacıyla tercih edilmektedir (Çakıcı Eser ve Gelbal, 2013; Jonsson ve Svingby, 2007; Karakaya, Sarıtaş ve Salmaner 2015; Kutlu vd. 2010, s.52).

DPA'lar, öğrenciler arası iletişim sağlayarak akran geri bildirim için de kullanılabilir. Böylece öğrenciler arkadaşlarının çalışmasını tartışmaya yönelik puanlama anahtarının dilini kullandığı için kişisel tepkilerden çok ölçütler üzerine odaklanır. Öğrenci, akranlarına çalışmaları hakkında bilgiler verirken kendisinin de ölçütler hakkındaki kavrayışları gelişmektedir. Ayrıca puanlama anahtarları öğrencilerin kendi çalışmalarını izlemek için de uygun referans noktalarıdır (Brookhart, 2013, s.104-105). Öğrenciler ile öğretmenlerin aynı dereceli puanlama anahtarını kullanarak öğrenciye, kendi ve öğretmen puanlamalarını karşılaştırması açısından deneyim sağlayan çalışmalar bulunmaktadır (Karakaya, 2015; Karakaya vd. 2015). Bu sebeple öğrencilerin değerlendirme sürecinde kendi ve akranlarının çalışmalarını puanlaması, hem birey için önemli bir yaşantı olduğu hem de öz ve akran değerlendirme uygulamalarının daha başarılı uygulanmasına fayda sağlayacağı düşünülmektedir.

Noonan ve Randy (2005) ve Ross (2006) öz değerlendirmeyi, öğrencilerin öğrenme sürecinde yaptığı çalışmaları, öğretmen ya da öğrenciler tarafından belirlenmiş ölçütler doğrultusunda değerlendirmeleri ve öğrenmeleri hakkında karar vermek olarak tanımlamışlardır (Aktaran Kutlu vd. 2010, s.95). Problem çözme sürecinin bitiminde öğrenciler öz değerlendirme uygulaması ile öğrenciler durum belirleme sürecine katılarak problem çözme stratejilerini ne kadar bildiklerini ve ne kadarını uygulayabildiklerini görebilirler (MEB, 2009). Öğrencilerin durum belirleme sürecine katıldığı diğer bir değerlendirme çeşidi olan akran değerlendirme, Topping (1998)'e göre bireylerin benzer konumdaki akranlarının öğrenme ürünlerini veya kazanımlarını değer, başarı, düzey ve nitelik açısından eleştirel bakış açısıyla incelemesi için yapılan düzenleme olarak

tanımlamıştır. Problem çözme becerisinde akran değerlendirmenin kullanımı, öğrenciye farklı çözüm yollarının da olabileceğini sunmaktadır. Ayrıca süreç içerisinde öğrencinin farklı çözüm yollarını görüp değerlendirerek farklı fikirlere saygı duymayı öğrenebilmesi söz konusudur.

Akran değerlendirmede, öğrencilerin arkadaşlarının çalışmalarını incelerken ne kadar objektif olacağı tartışma konusudur. Fakat etkili bir akran değerlendirmenin gerçekleşmesi gerekli eğitim ve uygulama çalışmalarının yapıldığı takdirde öğrencilerin akranlarını objektif değerlendirdiği görülmüştür (Uysal, 2008, s.65).

Araştırma kapsamında problem çözme becerilerini değerlendirmeye yönelik performansa dayalı durum belirleme yaklaşımı ile öz, akran ve öğretmen puanlamaları kullanılarak öğrencileri de değerlendirme sürecine dâhil eden bir sürecin planlanması öğrenci için zengin dönütler sağlayabilecek önemli bir yaşantı olduğu düşünülmektedir. Fakat bu tür objektif olmayan testlerde puanlayıcı kanaatleri devreye girebilmektedir. Öğrencinin alacağı puan, puanlayıcıdan puanlayıcıya farklılık gösterebilmektedir (İlhan, 2015). Dolayısı ile puanlama güvenilirliği ön plana çıkmaktadır. Puanlama güvenilirliği, ölçme aracının farklı puanlayıcılar tarafından ya da aynı puanlayıcı tarafından farklı zamanlarda puanlanması ile elde edilen puanlar arasındaki tutarlılığına denir (Tekin, 1991, s.70). Çalışmada ele alınan öz, akran ve öğretmen puanlamalarının kullanılmasından dolayı puanlayıcılar arası güvenilirliğin kestirilmesi gereklidir. Puanlayıcılar-arası (inter-rater) güvenilirlik, farklı puanlayıcıların birbirinden bağımsız her bir öğrencinin kâğıdına aynı puanı verme tutarlılıklarının derecesidir (Baykul, 2010; Crocker ve Algina, 1986).

Literatür incelendiğinde puanlayıcı güvenilirliğini hesaplamaya yönelik klasik test kuramına dayalı güvenilirlik hesaplama yöntemleri, genellenebilirlik kuramı ve madde tepki kuramına dayalı çok yüzeyli Rasch ölçme modelinin kullanıldığı çalışmalar bulunmaktadır. Söz konusu üç kurama dayalı yapılan karşılaştırma çalışmalardan elde edilen sonuçlara göre güvenilirlik hesaplama yöntemleri iyi sonuçlar vermektedir (Güler, 2008; Macmillan, 2000). Fakat klasik test kuramında, farklı varyans kaynaklarının büyüklüğünün kestirilmesi çoklu analizlerin yapılmasını gerektirmektedir ve farklı varyans kaynakları arasındaki etkileşimden kaynaklanan hata hesaplanamamaktadır. Ayrıca klasik test kuramında kullanılan hesaplama yöntemleri gruptan bağımsız değer hesaplamaya imkân tanımaz (Fine'dan aktaran Güler, 2008). Genellenebilirlik kuramına dayalı yapılan güvenilirlik çalışmalarında; puanlayıcılardan, farklı değişkenlik kaynaklarından ve bunların etkileşimlerinden ortaya çıkan hata varyansları bir arada incelenebilir (Güler, 2008). Fakat

madde tepki kuramının bir modeli olan Rasch modeli test güvenilirliği genellenebilirlik kuramının verdiği güvenilirlikten daha yüksektir. Çünkü, Rasch hata varyansını madde ve puanlayıcı varyanslarına dahil etmez (Linacre, 1993). Çalışmada güvenilirlik kestirimlerini hesaplamak üzere belirlenecek yöntem, içinde bulunan değişkenlik kaynaklarının durumlarına göre seçilebilir. Çok yüzeyli Rasch ölçme modeli ile başarı puanlarını etkileyebileceği düşünülen tüm değişkenlik kaynaklarının göz önünde bulundurulması sağlanabilmektedir (Baird, Hayes, Johnson ve Lamprianou, 2013).

Bu araştırmada, problem çözme becerilerinin dereceli puanlama anahtarı ile öz, akran ve öğretmen tarafından puanlamalarının değerlendirilmesi ve zaman içerisinde öğrenci puanlamalarında gerçekleşen olası değişimlerin (katılık/cömertlik) izlenmesi amaçlanmıştır. Dolayısıyla bu çalışmada insan yargılarına dayalı değerlendirmeleri olabildiğince nesnelleştirme girişiminde olması (Hetherman 2004) avantajı ile çok yüzeyli Rasch ölçme modeli kullanılmıştır.

1.2.Kuramsal Çerçeve

Bu başlık altında problem çözme süreci, problem çözme becerisinin değerlendirilmesi ve puanlayıcı güvenilirliği ele alınmıştır.

1.2.1.Problem Çözme Süreci

Alan yazına bakıldığında problem ve problem çözmeye yönelik birçok tanım bulunmaktadır.

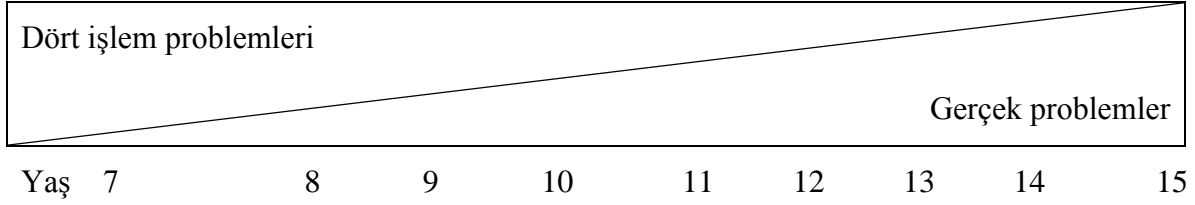
John Dewey problemi, insan zihnini karıştıran, insana meydan okuyan ve inancı belirsizleştiren her şey olarak tanımlamıştır. Problemin tanımı bu şekilde ele alındığında, problemin çözümü de zihni karıştıran ve inancı belirsizlikleri ortadan kaldırılması anlamına gelmektedir. Bir problemi çözmek için durumun analiz edilmesi, gerekli verilerin toplanması, bunlardan çözüme uygun olanların seçilmesi ve seçilen verilerin uygun şekilde kullanılması gerekir (Baykul, 2006, s.62). Bingham (2004, s.18-23) problemi, kişinin istenilen hedefe ulaşmak amacıyla mevcut güçlerinin karşısına çıkan engel olarak tanımlarken; problem çözmeyi, kişinin istenilen hedefe ulaşmak için karşılaştığı güçlükleri ortadan kaldırmasına yönelik çabalarının olduğu bir süreç olarak tanımlamıştır.

Morgan'a (2006, s.149) göre problem, bireyin bir amaca ulaşmada engellenme ile karşılaştığı bir çatışma durumu iken; problem çözme ise bireyin karşılaştığı engeli ortadan kaldırmanın en iyi yolunu bulmak olarak tanımlamıştır. Baykul 'a (1999, s. 67) göre problem çözme, istenilen amaca ulaşabilmek için etkili ve yararlı olan araç ve davranışları çeşitli imkânlar içinde seçme ve kullanmadır. Altun (2002, s.83-84) ise problemi, kişinin bir şeyler isteyip ama ne yapacağını hemen kestiremediği durumlar olarak ifade ederken; problem çözmeyi de bireyin ne yapacağını bilmediği durumlarda, yapılması gerekeni bilmesi olarak tanımlamıştır.

PISA 2012'ye göre problem çözme yeterliliği, çözümü açık olmayan bir problem durumunda, bireyin durumu anlaması ve bilişsel süreçler yardımı ile çözme kapasitesi olarak ifade edilmiştir. Alan yazında yapılan tanımlamalardan yola çıkarak problem çözmeyi, bireyin problem ile karşılaşma anından söz konusu probleme çözüm buluncaya kadar geçirdiği bir süreç olarak düşünülebilir.

Genel olarak; bir problem, önceden çözülmüş genel bir probleme özel veriler yerleştirilerek ya da hiçbir özgünlük katmadan iyice bilinen bir örneği adım adım izleyerek çözülebiliyorsa rutin bir problemdir (Polya, 1973, s.171). Matematik eğitiminde; öğrencinin bilgileri doğrudan kullanarak çözüme ulaştığı problemler “rutin problemler” olarak ele alınmaktadır. Öğretim programında vurgulanan “problem çözme becerileri” rutin olmayan problemler çerçevesinde düşünülmelidir (MEB, 2013). Rutin problemler öğrencinin daha çok ilkokula başladığı dört işlem becerilerinin kazanmaya başladığı yıllarda sık çözülür. Polya (1973, s.172), rutin problemlerin gerekli olduğunu fakat öğrencilerin rutin problemler dışındaki problemlerle karşılaşmasını sağlamamanın, öğrencileri hayal gücünden ve muhakeme becerisinden yoksun bıraktığı için rutin olmayan problemlerin önemine dikkat çekmiştir.

Altun (2002) şekil 1'de problem durumlarının öğrencilerin yaşları ile ilişkisini ortaya koymuştur.



Şekil 1. Problem durumları ile yaş ilişkisi

Problem çözme sürecinde ise beynimizde hangi işlemlerin olduğu, problem çözme işinin nasıl gerçekleştiği ve bu sürecin hangi parçalardan oluştuğu kesin olarak açıklanamamaktadır. Ancak ortaya atılan bazı kuramlar açısından ve yapılan araştırmalarla problem çözme sürecindeki bazı adımlar ayırt edilebilmektedir (Baykul, 2006, s.73) .

Dewey' in (1910) problem çözme sürecindeki temel aldığı aşamalar şunlardır (aktaran Topses, 2003, s. 251): Problemi fark etme, problem hakkında bilgi toplama, probleme çözümler üretme, çözüm önerilerini deneyerek test etme, uygulama hakkında bilgi toplama, o ana kadar uygulamaya konulan adımların özetini yapma ve yeni bir uygulama planı oluşturmaktır. Polya' ya (1973) göre problem çözme süreci; problemin anlaşılması, çözüm ile ilgili stratejinin seçilmesi, seçilen stratejinin uygulanması, çözümün değerlendirilmesi olarak dört aşamada incelenebilir.

Wallas (1926), büyük düşünürlerin problem çözme yaklaşımını inceleyerek ortak olarak aşağıdaki dört aşamayı vurgulamıştır (aktaran Morgan, 2006, s.150-151): Hazırlık aşamasında, problemi ortaya koyar ve probleme ilişkin veri toplar. Kuluçka aşamasında, çözülemeyen problem ileride daha başarılı olma umuduyla askıya alınır. Bilinçaltı süreçlerde iş başındadır. Kavrayış (aydınlanma) aşamasında, düşünürde kavrayış gözlenir ve düşünürün aklında aniden yeni bir fikir doğar. Son aşama olan değerlendirme ve düzeltme aşamasında, fikrin doğrulunu sınar. Fikir doğru ise ufak düzeltmeler yapabilir veya çözüm işe yaramazsa başlangıç noktasına döner.

PISA 2012 çalışmaları çerçevesinde hazırlanan raporlara göre problem çözme sürecinde izlenmesi gereken adımlar; keşfetme ile anlama, temsillendirme ile formüle etme, planlama ile uygulama, kontrol etme ile yansıtma olarak belirtilmiştir.

1.2.2.Problem Çözme Becerisinin Değerlendirilmesi

Ulusal Matematik Öğretmenleri Konseyi'nin (National Council of Teachers of Mathematics-NCTM) 2000 yılındaki raporuna göre, eğitimin bütün aşamalarında öğrencilerin problem çözme süreçlerini yansıtan yaklaşımların kullanılması gerekmektedir. Problem çözme sürecinin belirli aralıklarla gerçekçi bir şekilde değerlendirilerek çözümün ne kadar etkin olduğu ve yeni problemlerin ortaya çıkıp çıkmadığının belirlenmesi çok önemlidir. Bireyin "Bu çözüm yolunu nasıl bulduk?" sorusunu kendine sorması, uygulamanın aksayan yönlerini belirlemesine ve süreçte izlenen adımları çözümlemesine yardımcı olur. Bireyi işlemlerin sırasını ve organizasyonunu değerlendirmeye sevk ederek bireyin uygulama esnasında ortaya çıkan problemleri fark etmesi ve çözüme kavuşturması sağlanır. Sürecin bu şekilde ilerlemesi bireyin gelecekteki benzer problemleri daha kolay çözmeye ve problemlerdeki çözüm yollarının nitelik ve nicelik açısından iyileştirmesine yardımcı olacaktır. (Bingham, 2004, s.38; Yıldırım, 2003, s.147-148).

Öğretim sürecinin ayrılmaz bir parçası olan ölçme ve değerlendirme etkinliklerinde, geleneksel ölçme değerlendirme yaklaşımlarının öğretim sürecinden kopuk olduğu (Gelbal ve Kelecioğlu, 2007), öğrencilerin matematik yeteneğinin ve matematiksel bilgiyi oluşturma süreçlerinin çok az bir kısmına odaklandığı (Bal ve Doğanay, 2010), öğrencilerin problem çözme gibi üst düzey bilişsel süreçlerini yeterli bir şekilde ölçemediği bilinmektedir (Baştürk, 2008; Berberoğlu, 1988; White, 1999; Wiliam, 2003). MEB (2013)' e göre öğrencilerin araştırma ve sorgulama yapabilecekleri, eleştirel düşünebilecekleri, gerekçelendirme yapabilecekleri, fikirlerini rahatlıkla paylaşabilecekleri, farklı çözüm yöntemleri sunabilecekleri bir sınıf ortamının oluşturulabilmesi için öğrencilere açık uçlu sorulara, etkinliklere yer verilmesi ve performansa dayalı yöntemlerin kullanılması önerilmektedir.

Performansa dayalı durum belirleme daha çok problem çözme, eleştirel düşünme gibi üst düzey düşünme gerektiren karmaşık yapıdaki becerileri ölçmek için öğrenci başarısının gelişimini sağlamak amacıyla yapılan çalışmalar olarak tanımlanır (Kutlu vd. 2010, s.30). Performansa dayalı durum belirleme çalışmaları ile öğrencinin bilgiyi gerçek hayattaki ortamlarda kullanabilmesine yönelik öğretim uygulamalarının izlenmesi ve ölçme ve değerlendirmenin hem sonuç hem de süreç odaklı yapılması sağlanır (MEB, 2013).

Performansa dayalı durum belirlemenin bir başka avantajı ise, öğrencinin şans ile doğru cevabı bulma olasılığını ortadan kaldırmasıdır. Bu şekilde ölçme hataları şans faktöründen arındırılarak öğrencinin yetenek düzeyi için daha güvenilir sonuç elde edilmesi sağlanır

(Güler, 2014). Fakat açık uçlu etkinlikler ve performans görevlerinin objektif puanlanamama durumu da bir güvenilirlik sorunudur (Kutlu vd., 2010, s.49; Romagnano, 2001). Burada yaşanan güçlüklerden birisi öğrencilerin nasıl puanlanacağına karar vermek, diğeri ise ölçmenin güvenilirliğini sağlamaktır (Çakıcı Eser ve Gelbal, 2013). Alan yazında, iyi hazırlanmış puanlama anahtarı kullanmanın değerlendirmenin kalitesini artırdığı ve daha nesnel belirleme yapmaya katkı sağladığı birçok eğitimci ve araştırmacı arasında fikri kabul görmektedir (Jonsson ve Svingby, 2007; Karakaya vd. 2015; Kutlu vd. 2010; Parlak ve Doğan 2014). Dereceli puanlama anahtarındaki her ölçütün istenilen performans düzeyine göre tanımlanmış olması, birbirinden bağımsız puanlayıcıların aynı puanı verme olasılığını yükseltmektedir (Moskal, 2000). Dereceli puanlama anahtarı (DPA) öğrencilerin çalışmalarını veya ürünlerini analiz etmek için öğretmen ya da bir değerlendirici rehberliğinde geliştirilen ve tanımlanan bir puanlama tasarımı olarak tanımlanmıştır (Moskal 2000). Eğitimde DPA'nın kabul görmüş başka bir tanımı da “otantik ya da karmaşık öğrenci çalışmasının nitel değerlendirilmesi için performansın önemli bölümlerinin puanlanması için ölçütleri ve bu ölçütlere ulaşmanın standartlarını içeren bir puanlama aracı” olarak belirtilmiştir (Jonsson ve Svingby, 2007). Yapısal özellikleri açısından DPA, bütünsel (holistic) ve analitik (analytical) olmak üzere ikiye ayrılır. Bütünsel puanlama anahtarında puanlayıcı, performansın niteliği hakkında genel bir değerlendirme yapılarak tek bir puanlama yapılır. Analitik puanlama anahtarında puanlayıcı, performansın her bir boyutuna ilişkin ölçütlere göre bir puan belirler. Bütünsel puanlama genellikle kolay ve ekonomik olduğu için büyük ölçekli değerlendirmelerde kullanılır. Analitik puanlamanın sonuçları, öğrencilere ve öğretmenlere öğrencinin güçlü yönlerini ve öğrenme ihtiyaçlarını belirlemesine yardım ettiği için analitik puanlama anahtarının sınıf içi kullanımı daha yararlıdır. Her iki türde verilen göreve göre puanlama anahtarı değişiyorsa göreve özel (task specific), tüm görevler aynı puanlama anahtarı ile puanlanıyorsa genel (generic) olarak adlandırılmaktadır (Jonsson ve Svingby, 2007). DPA'lar etkili geri bildirimler yoluyla öğrencilerin sonraki görevlerinde kavram ve becerileri daha iyi anlamasını kolaylaştırması açısından bir potansiyele sahiptir (Kuehl, Sofronas ve Lau, 2015). Değerlendirmede hem öğretmene hem öğrenciye ne arandığını ve neyin önemli olduğunu anlatır. Bu hem kritik kararların verildiği hem de öğrenme amaçlı değerlendirmeler için geçerlidir (Jonsson ve Svingby, 2007).

Açık uçlu etkinlik uygulamalarında ve performans değerlendirmelerinde objektif puanlanamama durumuna yönelik performans değerlendirmede yapılan ölçme işleminin

güvenirliğini artırmak adına puanlayıcı sayısı artırılabilir (Çakıcı Eser ve Gelbal, 2013; Ebel'den aktaran İlhan 2015).

1.2.3.Puanlayıcı Güvenirliği

Ölçme ve değerlendirme de üç temel kuram olan klasik test kuramı (KTK), genellenebilirlik kuramı (GK) ve madde tepki kuramında (MTK) puanlayıcı güvenirliğini sağlamaya ilişkin güvenirlilik hesaplama yöntemleri bulunmaktadır.

1.2.3.1.KTK'ya Dayalı Puanlayıcılar Arası Güvenirlilik Hesaplama Yöntemleri

Performans değerlendirmede puanlayıcılar arası güvenirliği hesaplamak için KTK 'ya dayalı birçok yöntem kullanılmaktadır. Farklı ölçme durumları için geliştirilen söz konusu hesaplama yöntemlerinden Cohen' in Kappa istatistiği, ağırlıklandırılmış Kappa (Weighted Kappa) istatistiği, Fleiss Kappa istatistiği, Kendall' ın uyum katsayısı, Krippendorff'un Alpha katsayısı, Kendall τ , Goodman ve Kruskal' ın Υ , sınıf içi korelasyon (ICC) katsayısı, Pearson Moment Çarpım korelasyon katsayısı, Spearman sıra korelasyon katsayısı gibi yöntemler genellikle kullanılmaktadır. KTK' da puanlayıcı güvenirliğinde kullanılan verinin ölçek türü ve puanlayıcı sayısına göre özelliklerini Kılıç (2009) tablo 1'de belirtmiştir.

Tablo 1. Klasik Test Kuramına Ait Bazı Güvenirlilik Hesaplama Yöntemleri

	Ölçek Türleri					
	Sınıflama		Sıralama		Eşit Aralıklı-Eşit Oranlı	
	Puanlayıcı sayısı		Puanlayıcı sayısı		Puanlayıcı sayısı	
Uyumluluk Ölçütü	İki	Çoklu	İki	Çoklu	İki	Çoklu
Cohen' in Kappa istatistiği	+	-	-	-	-	-
Fleiss Kappa istatistiği	-	+	-	+	-	-
Ağırlıklandırılmış Kappa	-	-	+	-	-	-
Kendall'in uyum katsayısı	-	-	-	+	-	-
Krippendorff Alfa	+	+	+	+	+	+
Kendall τ	-	-	+	-	-	-
Goodman ve Kruskal'in γ	-	-	+	-	-	-
Sınıf içi korelasyon ICC	-	-	+	+	+	+
Pearson Korelasyon	-	-	-	-	+	-
Spearman Korelasyon	-	-	-	-	+	-

1.2.3.2. Genellenabilirlik Kuramı

Genellenebilirlik kuramı, davranış ölçmede güvenilirliğin değerlendirilmesini, güvenilir gözlemlerin tasarlanmasını, araştırılmasını ve kavramlaştırılmasını sağlayan, temelinde varyans analizine (ANOVA) dayalı olan istatistiksel bir kuramdır (Shavelson ve Webb 1991).

Genellenebilirlik kuramının amacı, ölçme sonuçlarını farklı varyans kaynaklarına ayırarak, yorumlayarak ve tanımlayarak, ölçme konusu olan bireyler ya da objelerin gözlenen puanlarının evren puanlarına (gerçek puan) genellenmesini sağlamaktadır (Atılğan, 2004). Shavelson ve Webb (1991), genellenebilirlik kuramının klasik test kuramının genişletilmiş bir uzantısı olduğunu dört farklı açıdan belirtmiştir: İlk olarak, genellenebilirlik kuramı çoklu varyans kaynaklarını tek bir analiz ile inceler. İkinci olarak her bir varyans kaynağının büyüklüğünün belirlenmesini sağlamaktadır. Üçüncü olarak, bireylerin

performanslarına ilişkin hem bağıl kararlar hem de mutlak kararlar alınabilmesine yönelik iki farklı güvenilirlik katsayısı (G ve phi katsayıları) hesaplamaktadır. Son olarak da belirli bir amaca yönelik ölçme hatasının en aza indirgenebileceği ölçmelerin düzenlenmesine yani karar çalışmalarına imkân tanımaktadır.

1.2.3.3. Çok Yüzeyle Rasch Ölçme Modeli

MTK'nın bir parametrelili logistik modeli olarak da bilinen Rasch ölçme modelinin genişletilmesiyle çok yüzeyle Rasch ölçme modeli (ÇYRÖM) oluşturulmuştur (Linacre, 1993). Model, aynı analizde her birey için ölçümü, diğer bireylerin dağılım özelliklerinden bağımsız hesaplamaya çalışır (Linacre, 1993).

ÇYRÖM'ün doğrusal eşitliği parametrelerin ayrı olarak kestirilmesine imkân sağlamaktadır (Atılğan, 2004). Böylece model, her bir değişkenlik kaynağına ilişkin örneklemeden bağımsız parametre hesaplar. Her bir değişkenlik kaynağına ait aynı anda istatistiksel olarak birbirinden bağımsız parametre kestirimi yapar. Yani, bireylerin yetenekleri, belirli maddelerin dağılım özellikleri ve puanlayıcı katılık/cömertliklerinden bağımsız olarak kestirilmesine imkân tanımaktadır (Güler, 2008).

ÇYRÖM'de her türlü sistematik hata kaynağı değişkenlik kaynağı olarak ele alınabildiği için genellenebilirlik kuramında ele alınamayan durumları değişkenlik kaynağı olarak incelemektedir (Alharby, 2006).

1.3. Araştırmanın Önemi

İlköğretim döneminde zihinsel gelişimin hızlı olduğu göz önüne alındığında (Baykul, 2006, s.61); yaşam boyu ihtiyaç duyulan, sınırları okul ile çizili olmayan, matematik programının temel amaçlarından olan problem çözme becerisinin (NCTM, 2000; MEB 2013) önemini vurgulamak gerekir. MEB (2013) göre, problem çözme becerisi gibi üst düzey zihinsel süreçlerin gelişim düzeyini belirlemek ve değerlendirmek için ölçme araçlarının çeşitlendirilmesi bütüncül ve çok yönlü bir değerlendirme açısından önem taşımaktadır. Öyle ki süreci içine alan, öğrencinin neyi başarıp neyi başaramadığını fark edeceği performansa dayalı durum belirleme araçlarının öğrenciyi öz ve akran olarak değerlendirmede aktif kılarak yer alması öğrencinin öğrenme sürecindeki farkındalığını artıracak düşünülmemektedir. Dolayısıyla ile bu yıllardaki öğrencilerin hem kendi hem de

akranlarının problem çözme süreçlerini puanlamaları, öğrencinin gelecekte yaşayacağı problem çözme süreçlerine katkı sağlayacaktır. Fakat bununla beraber performansa dayalı durum belirlemenin en önemli sınırlılığı olan güvenilirlik (Kutlu vd. 2010, s.49) sorunu zihinlere gelmektedir. Performansa dayalı öğrenci çalışmalarının değerlendirilmesinde yapılırken olabildiğince nesnel bir ölçme yapılması güvenilirlik açısından önem taşımaktadır. Nitekim yurt dışında özellikle üst düzey zihinsel becerilerin değerlendirilmesinde farklı değişkenlik kaynaklarına göre güvenilirlik kestirimleri veren çok yüzeyli Rasch ölçme modelinin kullanıldığı çalışmalar vardır (Alharby, 2006; Brown, O’Gorman ve Du, 1996; Farrokhi, Esfandiari ve Dalili, 2011; Farrokhi, Esfandiari ve Schaefer, 2012; Hetherman, 2004; Macmillan, 2000; Matsuno, 2006; Nakamura, 2002; Smith ve Kulikowich, 2004; Weigle, 1999; Wolfe ve Chiu 1997). Türkiye’de ise araştırmaların ağırlıklı olarak son yıllarda yapıldığı görülmektedir (Akın ve Baştürk, 2012; Atılğan, 2004; Baştürk, 2008; Baştürk, 2010; Güler, 2008; Güler ve Gelbal 2010; Güler, 2014; İlhan, 2015; Karakaya, 2015; Karakaya vd., 2015; Yüzüak vd., 2015). Fakat hem yurt dışı hem yurt içi alan yazında problem çözme becerilerinin değerlendirilmesinde öz, akran ve öğretmen puanlamalarının güvenilirliği açısından zaman içerisinde gelişimini gösteren bir çalışmaya rastlanmamıştır. Bu yüzden ilköğretimde problem çözme becerisi gibi üst düzey zihinsel beceriler gerektiren ve değerlendirmesi performansa dayalı olabilecek öğrenmelerin değerlendirilmesinde öz, akran ve öğretmen puanlamalarının kullanılarak çok yüzeyli Rasch ölçme modeli ile süreç içerisinde incelenmesinin alan yazına katkı verebileceği düşünülmektedir.

Okul uygulamalarında da performansa dayalı durum belirleme (Kutlu, vd. 2010) ve öğrencinin durum belirme sürecine katılmasına (Uysal, 2008) yönelik ölçme araçlarının kullanımında sıkıntı yaşandığı bilinmektedir. Dolayısıyla bu çalışma uygulamadaki bu tür sorunlara yönelik örnek niteliği taşımaktadır. Ayrıca bu çalışma ile öğrencinin kendi çözüm süreci ve arkadaşlarının çözüm sürecine olabildiğince objektif yaklaşmasına ve eleştirel bir bakış açısı ile incelemesine fırsat sunulmuştur. Ek olarak bu durumun öğrencinin sonraki problem çözme süreçleri için olumlu etki bırakacağı düşünülmektedir.

1.4.Araştırmanın Amacı

Bu araştırmada ilköğretim altıncı sınıf öğrencilerinin problem çözme becerilerini değerlendirmeleri üzere öğrenci performansları; öz, akran ve öğretmen tarafından dereceli puanlama anahtarı kullanılarak puanlanmış ve süreç içerisindeki öz ile akran

puanlamalarının öğretmen puanlamalarına göre uyumunun izlenmesi amaçlanmıştır. Bu amaca yönelik mevcut program kapsamında, öğrencilerin problem çözme becerilerini ölçmeye yönelik performans görevleri verilmiş; bu görevler her hafta öz, akran ve öğretmenler tarafından puanlanmıştır. Süreçte puanlayıcılar arasındaki karşılaştırmalar çok yüzeyle Rasch ölçme modeliyle incelenmiştir. Araştırmanın bu temel amacı doğrultusunda aşağıdaki sorulara yanıtlar aranmıştır.

1.5.Araştırma Soruları

Çalışma kapsamında “Problem çözme becerilerini ölçmek üzere hazırlanan performans görevi uygulamalarının öz, akran ve öğretmen puanları ile süreç içinde çok yüzeyle Rasch ölçme modeli ile değerlendirilmesi nedir?” sorusuna cevap aranmıştır.

1. Birinci performans görevi uygulamasında problem çözme becerisinin öz, akran ve öğretmen puanlarının çok yüzeyle Rasch ölçme modeliyle analizi sonucu kalibrasyon haritasının durumu nedir?
 - a. Birinci performans görevi uygulamasında öğrencilerin başarı durumu ölçüm raporuna ait istatistikler nedir?
 - b. Birinci performans görevi uygulamasında aşamalara ait ölçüm raporundaki istatistikler nedir?
 - c. Birinci performans görevi uygulamasında puanlayıcı türüne ait ölçüm raporundaki istatistikler nedir?
2. İkinci performans görevi uygulamasında problem çözme becerisinin öz, akran ve öğretmen puanlarının çok yüzeyle Rasch ölçme modeliyle analizi sonucu kalibrasyon haritasının durumu nedir?
 - a. İkinci performans görevi uygulamasında öğrencilerin başarı durumuna ait ölçüm raporundaki istatistikler nedir?
 - b. İkinci performans görevi uygulamasında aşamalara ait ölçüm raporundaki istatistikler nedir?
 - c. İkinci performans görevi uygulamasında puanlayıcı türüne ait ölçüm raporundaki istatistikler nedir?
3. Üçüncü performans görevi uygulamasında problem çözme becerisinin öz, akran ve öğretmen puanlarının çok yüzeyle Rasch ölçme modeliyle analizi sonucu kalibrasyon haritasının durumu nedir?

- a. Üçüncü performans görevi uygulamasında öğrencilerin başarı durumuna ait ölçüm raporundaki istatistikler nedir?
 - b. Üçüncü performans görevi uygulamasında aşamalara ait ölçüm raporundaki istatistikler nedir?
 - c. Üçüncü performans görevi uygulamasında puanlayıcı türüne ait ölçüm raporundaki istatistikler nedir?
4. Birinci, ikinci ve üçüncü performans görevi uygulamalarının tümünü içeren süreçte problem çözme becerisinin öz, akran ve öğretmen puanlarının çok yüzeyli Rasch ölçme modeliyle analizi sonucu kalibrasyon haritasının durumu nedir?
- a. Tüm görevlerin bulunduğu analizde öğrencilerin başarı durumuna ait ölçüm raporundaki istatistikler nedir?
 - b. Tüm görevlerin bulunduğu analizde aşamalara ait ölçüm raporundaki istatistikler nedir?
 - c. Tüm görevlerin bulunduğu analizde görevlere ait ölçüm raporundaki istatistikler nedir?
 - d. Tüm görevlerin bulunduğu analizde puanlayıcı türüne ait ölçüm raporundaki istatistikler nedir?

1.6.Sınırlılıklar

Araştırma kapsamında uygulama programı ilk olarak iki şube toplamında 75 öğrenciye uygulanmıştır. Fakat araştırma sürecinde devamlılık esas alındığı için 57 öğrenciye ait veriler kullanılmıştır.

Uygulama öncesinde iki akran puanlaması planlanmışken yapılan deneme uygulamasında süre sınırlaması nedeni ile bir akran puanlanmıştır.

Araştırmada performans görevleri öz, akran ve üç öğretmen tarafından puanlanmıştır. Fakat üçüncü uygulanan performans görevine ilişkin verilerin varyans düşüklüğü sebebiyle analiz sonucunu vermemesi üzerine analizden bir öğretmen puanlayıcı çıkarılarak analizde öz, akran ve iki öğretmen puanlamaları kullanılmıştır.

1.7.Varsayımlar

Araştırmada öğretmen ve öğrenci puanlayıcılar ciddiyetle ve birbirlerinden bağımsız olarak puanlama yapmışlardır.

Araştırmaya katılan öğrenciler problem durumlarında gerçek bilgi ve becerilerini yansıtacak şekilde performans göstermiştir. Bunun için süreçte olası sorulara yönelik açıklamalar yapılarak anlaşılabilirlik ve açıklık sağlanmaya çalışılmıştır. Ayrıca öğrencilere motivasyon sağlamak adına öğretmen puanlarına en yakın puanlama yapan öğrencilere çeşitli pekiştiriciler verilmiştir.

1.8.İlgili Çalışmalar

Bu bölümde, problem çözme becerisi, öz ve akran değerlendirme hakkında ve çok yüzeyli Rasch ölçme modeli ile ilgili çalışmalara yer verilmiştir.

1.8.1.Problem Çözme Becerisi Hakkında Yapılan Çalışmalar

Karataş ve Güven (2003), çalışmalarında problem çözme becerilerinin değerlendirildiği yöntemleri incelemiştir. Çalışmanın ilk kısmında, problem çözenin ve problem çözme becerilerinin öğrencilere kazandırılmasının önemine değinmiş; çalışmanın ikinci kısmında da problem çözme becerilerinin değerlendirilmesinde sık kullanılan yöntemlerden standart testler, performans değerlendirme, yazılı cevap gerektiren sorular ve klinik mülakat yöntemleri ele alınmıştır. İncelenen yöntemlerin problem çözme becerilerini değerlendirmedeki potansiyeli tartışılmıştır.

Gündüz Sefer (2006) çalışmasında, matematik dersinde problem çözme becerilerinin dereceli puanlama anahtarı ile değerlendirilerek geri bildirim verilmesinin söz konusu becerinin gelişimine etkisi incelemiştir. Uygulamalı araştırma niteliğinde olan çalışmanın araştırma grubunu İstanbul ili Üsküdar ilçesinde bir devlet okulunun biri kontrol (23 öğrenci), diğeri deney grubu (19 öğrenci) olmak üzere iki tane 5. sınıf şubesi oluşturmaktadır. Çalışma kapsamında öğrencilere açık uçlu sorulardan oluşan ön test uygulanarak durum belirlemesi yapılmıştır. Daha sonra deney ve kontrol gruplarına bir aylık zaman diliminde toplam beş problem çözme etkinliği uygulanmıştır. Araştırma sonuçlarına göre, etkinliklerin dereceli puanlama anahtarı ile değerlendirildiği deney

grubunun problem çözüme becerileri kontrol grubuna göre daha çok artmış ancak iki grubun ortalamaları arasında manidar bir fark bulunmamıştır.

Yayan (2010) çalışmasında, altıncı sınıf öğrencilerinin problem çözüme becerilerini problemi anlama, plan geliştirme, planı uygulama ve çözümü kontrol etme ve değerlendirme olarak dört ana süreci dikkate alarak incelemiştir. Bulgulara göre, altıncı sınıf öğrencilerinde genel olarak problem çözüme becerisinde düşük performans gösterdiği görülmüştür. Öğrencilerin problemi anlama sürecinde en yüksek performansı gösterdiği görülürken, çözümü kontrol etme ve değerlendirme sürecinde en düşük performansı gösterdiği görülmüştür.

Büyükkıdık (2012) araştırmasında, 2011-2012 bahar döneminde 132 ilköğretim 6, 7 ve 8. sınıf öğrencisinin rutin olmayan problem çözüme becerilerinin değerlendirilmesine yönelik iki adet performans görevi kullanmıştır. Dört puanlayıcı hem analitik hem de bütünsel dereceli puanlama anahtarı ile on gün ara ile puanlama yapmıştır. Elde edilen puanlamalar hem klasik test kuramı ve hem de genellenebilirlik kuramı ile puanlayıcılar arası güvenilirlik açısından karşılaştırma amacıyla incelenmiştir. Araştırmanın sonuçlarına göre her iki kuramdan da elde edilen bulguların güvenilirliğinin yüksek olduğu fakat genellenebilirlik kuramından elde edilen katsayıların klasik test kuramından elde edilen katsayılara nazaran daha yüksek olduğu ve klasik test kuramına göre daha detaylı bilgi sağladığı görülmüştür. Ayrıca, dereceli puanlama anahtarlarının ölçüm güvenilirliğini etkileyebileceği ve analitik dereceli puanlama anahtarından elde edilen verilerin bütünsel dereceli puanlama anahtarına göre daha yüksek güvenilirliğe sahip olduğu söylenmiştir.

Özmen-Hızarcıoğlu (2013) çalışmasında, problem çözüme becerisinin değerlendirilmesinde dereceli puanlama anahtarı kullanarak ve kullanmadan elde edilen puanlar arasındaki uyumu incelemeyi amaçlamıştır. Amaç doğrultusunda 2011-2012 eğitim öğretim yılında Bartın ilindeki bir özel okulda 6. sınıf düzeyindeki 27 öğrenciye uygulamak üzere 4 adet açık uçlu problem ve problemleri puanlamak için dereceli puanlama anahtarı hazırlanmıştır. Araştırma bulgularına göre, öğretmenlerin öğrencilere dereceli puanlama anahtarı kullanmadan verdikleri puanlar arasındaki uyumun düşük olduğu gözlenmiştir. Öğretmenlerin dereceli puanlama anahtarı kullanarak verdikleri puanlar arasındaki uyumun ise yüksek olduğu belirlenmiştir.

Kuehl, Sofronas ve Lau (2015) “Öğretmen Adaylarının ve Çalışan Öğretmenlerin Matematiksel Problem Çözmeyi Dereceli Puanlama Anahtarı ile Değerlendirmesi” isimli çalışmasında 33 öğretmen adayı ve 43 çalışan öğretmenin 4. Sınıf öğrencilerinin problem

çözme becerilerinin puanlama anahtarı kullanımının karşılaştırılması amaçlanmıştır. 5 boyut ve 4 düzeyden oluşan bütünsel puanlama anahtarı kullanılmıştır. Araştırma sonuçlarına göre beklentilerin aksine çalışan öğretmenlerin puanladığı ölçümlerden kaynaklanan hata varyansı daha yüksek olduğu gözlenmiş ve öğretmen adaylarının puanlamalarında daha az farklılaşma görülmüştür.

1.8.2.Öz ve Akran Değerlendirme Hakkında Yapılan Çalışmalar

Uysal (2008), araştırmasında yükseköğretim hazırlık sınıflarında öğrenim gören öğrencilerin akran ve öz değerlendirme çalışmalarını kullanarak etkin bir biçimde değerlendirme sürecine dâhil edilip edilemeyeceği sorusunu yanıtlamayı amaçlamıştır. Araştırmanın sonuçlarına göre, öğrencilerin akran ve öz değerlendirme için olumlu görüş bildirdikleri, akranlarını ve kendilerini değerlendirmeye istekli oldukları saptanmıştır. Ayrıca, öğrencilerin akranlarının puanları ile öğretmen puanları arasında yüksek korelasyon bulunmuştur.

Uzun (2010), ilköğretim 4 ve 7. sınıf öğrencilerinin alternatif değerlendirme yaklaşımlarından olan öz değerlendirme uygulamalarına yönelik tutumlarını incelemeyi amaçlamıştır. İstanbul'da 1500 ilköğretim öğrencisinden alınan verilerin analizleri sonucunda öğrencilerin öz değerlendirme hakkındaki tutumları; cinsiyete, okulun bulunduğu yere, öğretmenin cinsiyetine, sınıf düzeyine, öz değerlendirme uygulamalarını yapma sıklığına, öğretmenin tecrübesine ve sınıftaki öğrenci sayısı değişkenlerine göre anlamlı düzeyde farklılaştığı görülmüştür.

Olğun (2011), ilköğretim dördüncü sınıf fen ve teknoloji dersinde öz ve akran değerlendirmelerinin kullanıldığı işbirlikli öğrenme yönteminin öğrencilerin akademik başarılarına, tutumlarına ve biliş üstü becerilerine etkisini incelemeyi amaçlamıştır. Araştırmanın bulgularına göre, deney grubu öğrencilerinin başarı, tutum ve biliş üstü farkındalık düzeylerinin kontrol grubu ile arasında istatistiksel olarak anlamlı derecede daha yüksek olduğu saptanmıştır. Ayrıca, deney grubunda oturumların bitiminde uygulanan öz ve akran değerlendirme puanlarıyla, öğrencilerin başarı ve biliş üstü farkındalık ölçeği puanları arasındaki korelasyon son oturumlara doğru ilerledikçe anlamlı olarak arttığı tespit edilmiştir.

Bozkurt ve Demir (2012) çalışmalarında, ilköğretim 5. sınıf öğrencilerine yönelik akran değerlendirme etkinliği uygulamıştır. Etkinlikten yola çıkarak öğrencilerin akran

değerlendirmesine yönelik görüşlerinin belirlenmesi amaçlanmıştır. Araştırmanın bulgularında değerlendirmeden yüksek alan öğrencilerin görüşleri olumlu iken düşük alan öğrencilerin görüşleri olumsuz olmuştur. Fakat değerlendirmeye olumsuz yaklaşan öğrenciler de akran değerlendirmenin diğer ünitelerde de uygulanabilir yönünde görüş bildirmişlerdir.

1.8.3.Çok Yüzeyle Rasch Modeli ile İlgili Yapılan Çalışmalar

Brown, O’Gorman ve Du (1996) “Matematik Performansı Değerlendirmenin Güvenirlik ve Geçerliliği” adlı çalışmasında, Minneapolis Devlet Okuları’nda kullanılan matematik problemi çözme içerikli performans değerlendirmenin psikometrik özelliklerini puanlayıcılar arası güvenilirlik, puanlama güvenilirliği ve geçerlilik açısından incelemiştir. Çalışmada yaklaşık 1300 5. sınıf öğrencisine iki karmaşık problem sorulmuş ve yazılı iletişim değerlendirmenin önemli bir parçası olmuştur. Problemler strateji, çözümün niteliği, matematik dili, matematiksel düşünme, cümle yapısı ve imla ve teknik açısından puanlanmıştır. Araştırmada hem çok yüzeyle Rasch analizi hem klasik test teorisinden güvenilirlik hesaplama yöntemleri kullanılmıştır. Araştırmanın sonucunda, çoklu psikometrik yöntemlerle puanlayıcılar arası güvenilirlik, puanlama güvenilirliği ve değerlendirmenin geçerliliğini incelemek ve değerlendirme sistemimin gelişmesi ve izlenmesi yönelik kanıt elde etmenin mümkün olduğu görülmüştür.

Wolfe ve Chiu (1997) çalışmasında, büyük ölçekli bir performans değerlendirme ortamında değerlendirici hatalarının ortak desenleri nasıl belirlenebileceği tartışılmıştır. Ortak değerlendirici etkilerinin her birini sergilemek için simüle veri setleri oluşturulur. En yaygın değerlendirici etkileri: (1) doğruluk/rastgelelik; (2) katılık / yumuşaklık ve (3) merkezîyetçilik/ aşırılık. Simulasyon verideki değerlendirici etkilerini görmek için çok yüzeyle Rasch ölçme analizi yapılmıştır. Değerlendirici etkilerini normatif bir çerçevede çok yüzeyle Rasch analizi ile belirlenmiştir.

Nakamura (2002) çalışmasında, Japon sınıflarındaki İngilizce sözel sunum becerilerinin öğretmen ve akran değerlendirmelerinin güvenilirliğini çok yüzeyle Rasch ölçme modeli ile incelemiştir. Çok yüzeyle Rasch ölçme modelinde; ölçme birey yüzeyi için on iki üniversite öğrencisi, puanlayıcı yüzeyi için bir tanesi öğretmen olmak üzere random olarak seçilen dört tane akran ile beraber beş puanlayıcı, madde yüzeyi olarak da sınıf ortamında yapılan İngilizce sunum görevleri kullanılmıştır. Araştırma bulgularında, akran

değerlendirmelerin öğrenci motivasyonu için sunumlarında yararlı olabileceği ve öğrencilerin ekranlarını güvenilir puanlayıcı olarak kullanabilecekleri belirtilmiştir. Ayrıca sunum değerlendirmesi, akademik yeterlilik ölçümü, madde güçlüğü, puanlayıcı katılığı ve öğrencilerin sosyal uyum yeteneklerinin ölçümü konularında Rasch modeli kullanılmasının yararlı olabileceği belirtilmiştir.

Atılgan (2004) çalışmasında, 2002 ve 2003 yıllarında yapılan müzik öğretmenliği özel yetenek seçme sınavı verileri kullanılarak genellenebilirlik kuramının tek ve çok değişkenli modelleri ile çok değişkenlik kaynaklı Rasch modeli ile kestirilen istatistikler karşılaştırmıştır. Çalışmadan elde edilen bulgulara göre, genellenebilirlik kuramının tek değişkenli ve çok değişkenli modellerinin aynı ölçme durumu için alt boyutlardan oluşan testlerde farklı sonuçlar verdiği görülmüştür. Hem genellenebilirlik kuramı hem de çok değişkenlik kaynaklı Rasch modeli yaklaşımlarının ikisinin de kullanışlı sonuçlar ürettiği ve değişkenlik kaynakları için kestirilen istatistiklerin de kısmen tutarlı sonuçlar olduğu tespit edilmiştir.

Hetherman (2004) çalışmasında, New York şehri eğitim birimi insan kaynakları bölümünde personel alımı için kullanılan İngilizce kompozisyon yazma testlerinin etkililiğini çok yüzeysel Rasch ölçme modeli ile izlemeyi ve incelemeyi amaçlamıştır. Hem 0-1 puanlama (dichotomous) hem çoklu puanlama (polytomous) Rasch modeli kullanılan çalışmada değerlendirme sınav sistemi ile ilişkili tüm yüzeysel ve etkileşimleri incelenmiştir. Araştırmada 230 kişiden elde edilen verilerin analizi sonucunda: a) Puanlayıcıdan kaynaklı hatalar dikkate alınmadan, uygulamaya katılan bireylerin dil yeteneklerinin, (0,1) yetenek sürekliliğinde iki farklı düzeye ayrılmıştır. b) Dil yeteneği bileşenlerinin güçlük seviyeleri -1.68 den 1.33' e kadar anlamlı bir farklılık gösterdiği görülmüştür. c) Tümüyle puanlayıcılar arası güvenilirliğin yüksek olduğu ve standartlaştırılmış z puanlarının -1.48 ile 3.5 değerleri arasında olduğu bulunmuştur. d) Kompozisyon yazma testlerinin konu başlıklarının güçlük düzeylerinin farklı olduğu bulunmuştur. e) İncelenen değerlendirme sisteminin altındaki varsayımların uyumsuzlukları teori ve pratikteki bulgularının açıklamalarına yer verilmiştir.

Smith ve Kulikowich (2004) araştırmalarında, 44 tane 4. sınıf öğrencilerinin okuldaki kickball (futbolla beysbol karışımı bir spor türü) faaliyetlerini içeren karmaşık problem çözme becerileri değerlendirme verilerinin psikometrik özelliklerini genellenebilirlik kuramına göre ve çok yüzeysel Rasch ölçme modeline göre incelemiştir. Araştırma bulgularına göre, genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch ölçme

yaklaşımlarının değişkenlik kaynaklarının göreceli büyüklüklerinin değişimlerini benzer, her iki ölçüm tekniğinin değişkenlik kaynaklarının farklı olduğu gösterilmiştir.

Baştürk (2008) araştırmasında, yükseköğretimde öğrencilerin powerpoint sunumlarına ilişkin performansların kalitesini değerlendirmede çok yüzeyle Rasch modelinin kullanılabilirliğini incelemiştir. Bu çalışmada, sunum becerileri madde/görev gücü ve puanlayıcı katılık/cömertliklerine göre değerlendirilmiştir. Öğrenciler altı gruba ayrılarak her alt gruba sunum hazırlayacakları dersin içeriği kapsamında bir konu başlığı verilmiştir. Dersi veren öğretim elemanlarından oluşan yedi puanlayıcı, her bir alt grubun sunum performanslarını A+ powerpoint dereceli puanlama anahtarı ile puanlamıştır. Araştırma sonuçlarına göre yükseköğretimde çoklu kategorili verilerin performans ve akran değerlendirme çok yüzeyle Rasch modelinin kullanımının etkili olduğu sonucuna varılmıştır.

Güler (2008) araştırmasında, TIMSS (1999) uygulamasından aldığı 24 açık uçlu matematik sorusu, 2007 bahar öğretim yılında, 203 öğrenciye uygulanmıştır. Öğrencilerin verdikleri cevaplar dört puanlayıcı tarafından bütünsel dereceli puanlama anahtarı (rubrik) ile puanlanmıştır. Daha sonra elde edilen puanların güvenilirliği klasik test kuramı, genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch ölçme modeline göre incelenmiştir. Klasik test kuramına göre, Cronbach alfa güvenilirlik katsayısı 0,92, Kendall'ın uyum istatistiği 0,52 bulunurken puanlayıcılar arası korelasyon katsayıları 0.90 ile 0.97 arasında değişen yüksek ilişki gösteren değerler elde edilmiştir. Genellenebilirlik kuramına göre, puanların genellenebilirlik katsayısı 0,92 ve güvenilirlik katsayısı 0,90 bulunmuştur. Ayrıca puanlayıcı değişkenlik kaynağının toplam varyansı açıklama yüzdesi 2,1 olarak oldukça düşük bulunmuştur. Çok değişkenlik kaynaklı Rasch modeline göre, öğrenci boyutunun güvenilirliği 0,95 bulunurken puanlayıcılar arası güvenilirlik ise 0,99 olarak bulunmuştur. Araştırma bulgularına göre matematik başarısını ölçmek için kullanılan ölçme aracının güvenilir sonuçlar verdiği, puanlayıcıların birbirleri ile uyumlu puanlamalar yaptığı görülmüştür. Araştırma kapsamındaki üç kuramdan hangisinin puanlama güvenilirliğinde kullanılması gerektiğine ise puanların hangi amaçla kullanılacağına dikkate alınarak karar verileceği önerilmiştir.

Baştürk (2010) çalışmasında, öğrencilerin bilimsel araştırma becerilerini ölçmeye yönelik öğrenci performanslarına dayalı oluşturulan bilimsel araştırma ödevlerini çok yüzeyle Rasch ölçme yöntemi ile değerlendirmeyi amaçlamıştır. Araştırmada öğrencilerin bilimsel araştırma becerileri puanlayıcıların katılık/cömertliklerine ve davranışı gerçekleştirme gücüne göre değerlendirilmiştir. Araştırmanın çalışma grubu Pamukkale Üniversitesi,

Eđitim Fakóltesi, Eđitim Bilimleri Bölümü, Rehberlik ve Psikolojik anabilim dalında “Bilimsel Arařtırma Yöntemleri dersini alan 20 öđrenci olarak belirlenmiřtir. Verilerin analizine göre; hazırlanan bilimsel arařtırma ödevleri arasında “uygunluk ii” ve “uygunluk dıřı” istatistiklere sahip ödevlerin olmadığı, ödevlerin istatistiksel olarak anlamlı bir řekilde 3 ödev kalite tabakasına ayrıldıđı görölmüřtür. Puanlayıcıların yapmış oldukları puanlamalar arasında katılık ve cömertlik açısından istatistiksel olarak anlamlı farklılıklar gözlemlenmediđi, sadece bir puanlayıcı için “uygunluk dıřı” kareler toplamının beklenen deđerden farklı olduđu görölmüřtür.

Farrokhi, Esfandiari ve Dalili (2011), arařtırmalarında ok yüzeyle Rasch modelini kullanarak öz, akran ve öđretmen deđerlendirmelerindeki merkezi eđilimi belirlemeyi amaçlamıřlardır. alıřmada İnan Üniversitesinde öđrenim gören ve asıl branřı İngilizce olan İnanlı öđrenciler tarafından yazılan 188 makaleyi, 188’ i öz ve akran ve 6’sı öđretmen olmak üzere toplamda 194 puanlayıcı altı dereceli analitik puanlama anahtarı ile deđerlendirmiřtir. Arařtırma sonucuna göre; grup düzeyinde ve bireysel düzeyde üç eřit deđerlendirmede merkeziyetiliđin olmadığı görölmüřtür.

Semerci (2011a) alıřmasında mikro öđretim uygulamalarını ok yüzeyle Rasch ölme modeli ile incelemeyi amaçlamıřtır. alıřma grubunu 2008-2009 öđretim yılında Fırat Üniversitesi Eđitim Fakóltesi Bilgisayar ve Öđretim Teknolojileri Eđitimi Bölümü’nde mikro öđretim uygulaması yapılan 32 öđretmen adayı oluřtırmaktadır. Arařtırma sonuçlarından en katı puanlayıcı J1 ve en cömert puanlayıcı J2 olarak belirlenmiřtir. Rasch modelinin uygulandıđı alıřmalarda yanlılık analizinin sebeplerini aıđa ıkarmak için anket, görüřme gibi ek ölme araç ve yöntemlerin iřleme konulması önerilmektedir.

Semerci (2011b) arařtırmasında, doktora yeterlilikler erevesinde öđretim üyesi, akran ve öz deđerlendirmelerin Rasch ölme modeli ile incelemeyi amaçlamıřtır. Arařtırmacı madde yüzeyi olarak 15 yeterlik (YÖK tarafından kabul edilen ölütler), puanlayıcı yüzeyi olarak 5 öđretim üyesi ve 6 doktora öđrencisinden oluřan 11 jüri, birey yüzeyi olarak da 6 doktora öđrencisi belirlenmiřtir. Arařtırma sonuçlarına göre; P1 ve P2 kodlu doktora öđrencileri daha yeterli bulunurken P5 ve P6 kodlu doktora öđrencileri daha az yeterli bulunmuřtur. Puanlayıcılar arasından J5 kodlu jürinin en cömert J4 kodlu jürinin de en katı deđerlendirme yaptıđı tespit edilmiřtir. B3 yani; “Özgün ve disiplinler arası alıřmalarda liderlik yapar.” görevi en kolay görev, İ3 yani; ”Bir yabancı dili en az Avrupa Dil Portföyü C1 Genel Düzeyinde kullanarak ileri düzeyde yazılı, sözlü ve görsel iletiřim kurar.” görevi

en zor görev olarak belirlenmiştir. Çalışmada doktora eğitimi alan öğrencilerin yeterlilikleri arasında farklılıklar bulunmuştur.

Akın ve Baştürk (2012) çalışmalarında, Güzel Sanatlar ve Spor Lisesi öğrencilerinin keman eğitiminde kazandıkları temel becerileri çok yüzeyli Rasch ölçme modeli ile değerlendirmişlerdir. Çalışmada müzik öğretmenliği alanında eğitim gören öğretmen adaylarının keman eğitimi ile ilgili temel becerileri, puanlayıcıların katılık /cömertliklerine ve davranışı sergileme güçlüğüne göre değerlendirilmiştir. Örneklem olarak 3 farklı Anadolu Güzel sanatlar Lisesi'nde "Keman eğitimi" dersine devam eden 27 öğrenci belirlenmiştir. Veriler "Keman Çalma Becerilerini Değerlendirme Ölçeği" aracılığıyla toplanmıştır. Araştırmada müzik eğitiminde özellikle performansa dayalı öğrenmelerin değerlendirilmesinde çok yüzeyli Rasch ölçme modelinin etkin kullanılabileceği sonucuna varılmıştır.

Farrokhi, Esfandiari ve Schaefer (2012) çalışmalarında öz değerlendirici, akran değerlendirici ve öğretmen değerlendirici olmak üzere üç çeşit değerlendiricinin çok yüzeyli Rasch ölçme modeliyle farklılaşan puanlayıcı etkilerini veya puanlayıcıların katılıklarını/ cömertliklerini araştırmayı amaçlamışlardır. Araştırmanın uygulaması İran'da iki devlet üniversitesinde esas branşı İngilizce olan 188 İranlı öğrenci üzerinde yapılmıştır. Söz konusu 188 öğrencinin yazmış olduğu makaleler öz ve akran değerlendiriciler olarak öğrenciler ve öğretmenler tarafından altı dereceli analitik puanlama anahtarı ile değerlendirilmiştir. Araştırma sonuçlarına göre üç değerlendirme tipinde puanlama katılık ve cömertlik açısından farklılaşma olduğu görülmüştür.

Güler (2014), çalışmasında, açık uçlu maddelerden oluşan istatistik sınavı puanlarının güvenilirliğini; öğrenci, madde ve puanlayıcı yüzeyinin beraber ele alındığı madde tepki kuramında yer alan Rasch modelinin bir uzantısı olan çok yüzeyli Rasch modeli kullanarak incelemiştir. Araştırmada veriler 55 öğrencinin 10 maddeye verdiği cevapların 3 puanlayıcı tarafından puanlanarak elde edilmiştir. Araştırma sonuçlarına göre çok yüzeyli Rasch modelinin ölçme araçlarının geliştirilmesinde ve uygun ölçme koşullarının düzenlenmesinde önemli rolü olduğu belirtilmiştir. Özellikle birden fazla puanlayıcının olduğu öğrencilerin geleceğine yön veren sınavlarda kullanılması için uygun olacağı önerilmiştir.

İlhan (2015) çalışmasında açık uçlu matematik sorularının standart ve SOLO taksonomisine dayalı puanlama anahtarı ile puanlanmasında puanlayıcı etkilerinin çok yüzeyli Rasch ölçme modeli ile incelenmesini amaçlamıştır. Bu amaç kapsamında

araştırmanın katılımcılarını 8. sınıf öğrencisi 104 kişi, puanlayıcı olarak da yedi matematik öğretmeni oluşturmaktadır. Araştırma sonuçlarına göre, hem standart hem de SOLO taksonomisine dayalı olan puanlama anahtarlarının kullanıldığı puanlamalarda yetenek düzeyi farklı olan öğrencilerin yüksek güvenilirlikte birbirinden ayırt edilebildiği ve maddelerin güçlük düzeylerinin anlamlı olarak farklılaştığı belirlenmiştir. Standart puanlama anahtarının kullanıldığı puanlamalarda; halo etkisi, tutarsızlık, puanlayıcı x birey ve puanlayıcı x birey x madde yanlılıklarının olmadığı görülmüştür. Puanlamalarda merkeze yönelme etkisi ile puanlayıcı x madde yanlılığının olduğu, puanlayıcılar arasında katılık/cömertlik yönüyle anlamlı farklılık olduğu ve puanlayıcı güvenliğinin düşük olduğu belirlenmiştir. SOLO taksonomisine dayalı puanlama anahtarı ile yapılan puanlamalarda puanlayıcı katılık/cömertliği, merkeze yönelme etkisi, halo etkisi, tutarsızlık ve yanlılık puanlayıcı etkilerinden birine rastlanmamıştır. Ayrıca puanlayıcı güvenliğinin yüksek olduğu görülmüştür.

Karakaya (2015) çalışmasında portfolyoların değerlendirilmesinde öz, akran ve öğretmen değerlendirmelerini çok yüzeyli Rasch ölçme modeliyle karşılaştırmayı amaçlamıştır. Çalışma grubu 74 stajyer öğretmen olan araştırmada analiz için öğrenci başarısı, öğretmenlerin cinsiyeti, öz, akran ve öğretmen katılık/cömertliği ve portfolyo değerlendirme ölçütleri olmak üzere dört yüzey tanımlanmıştır. Araştırma sonucunda en cömert puanlayıcıların öz, en katı puanlayıcıların akran olduğu belirlenerek aralarında istatistiksel olarak manidar fark bulunmuştur. Çalışma kapsamında üniversitede yürütülmüş olan uygulamanın ayrıntılı bilgiler sağlaması ve pratik olması açısından kullanımı önerilmiştir.

Karakaya, Sarıtış ve Salmaner (2015) çalışmalarında istatistik dersi kapsamında öğrencilerin grup olarak yaptıkları görevlerden oluşan ürünlerini her öğrencinin bireysel olarak puanlamaları hem de üç öğretim elemanı tarafından puanlanmaları çok yüzeyli Rasch ölçme modeli ile incelemiştir. Araştırmanın çalışma grubunu 2014-2015 eğitim ve öğretim yılı, bahar dönemi Gazi Üniversitesi, Eğitim Fakültesi, Rehberlik ve Psikolojik Danışmanlık bölümü 2. Sınıf öğrencilerinden istatistik dersini alan 34 öğrenci oluşturmaktadır. Araştırmaya ait görevleri öncelikle öğrenciler dereceli puanlama anahtarı ile kendileri puanlamış sonrasında yine dereceli puanlama anahtarı ile üç tane öğretim elemanı tarafından değerlendirilmiştir. Araştırma sonucunda genel olarak, öğrencilerin kendilerini öğretim elemanlarına göre cömert puanladıkları görülmüştür. Çalışmada ürün ile süreci değerlendirmeye yönelik kullanılan dereceli puanlama anahtarının da amacına

hizmet ettiđi belirlenmiř ve puanlayıcıların ölçütlere göre yanlılıđı da incelenmiřtir. Söz konusu yanlılıđın olası sebepleri arařtırılmak üzere nitel çalıřmalar yapılması önerilmiřtir.

Yüzüak, Yüzüak ve Kaptan (2015) tarama modeli kullandıđı çalıřmalarında, performans görevinin akran grupları ve öğretmen yaklařımları dođrultusunda çok yüzeyle Rasch ölçme modeli ile deđerlendirmeyi amaçlamıřtır. Arařtırmanın çalıřma grubunu 2013-2014 eđitim yılı güz döneminde Bartın ilinde çalıřan bir öğretmen ve 7. sınıfta öğrenim gören 50 öğrenci oluřturmaktadır. Deđerlendirmede beřer kiřilik öğrenci grupları oluřturulmuř ve her grup bir ürün oluřturmuřtur. Analiz sonuçlarına göre, Grup 74'ün en cömert puanlayıcı grup olduđu, öğretmenin ise en katı puanlayıcı olduđu; P4'ün ölçütleri en iyi sađlayan görev olduđu, P6'nın ise ölçütleri daha iyi sađlamayan görev olduđu; Bilgi-dođruluk kodlu ölçütün en kolay karřılandıđı, Kaynakça kodlu ölçütün ise en zor karřılandıđı görülmüřtür. Çok-yüzeyle Rasch ölçme modelinin fen eđitiminde performans ve akran grup deđerlendirmeleri için etkili bir řekilde kullanılabileceđi sonucuna ulařılmıřtır.

Yurt içinde ve yurt dıřında yapılmıř olan çalıřmaları incelediđimizde; çok yüzeyle Rasch ölçme modelinin performans deđerlendirmeye yönelik bilgi ve becerilerin deđerlendirilmesinde etkili olduđu görülmektedir. Ayrıca modelde puanlayıcılar arasında öz, akran ve öğretmen deđerlendirmelerin beraber bulunduđu az sayıda çalıřma olduđu görülmüřtür. İlköđretim düzeyindeki öğrencilerin problem çözmeye becerilerini deđerlendirmeye yönelik puanlayıcılar arasında öz, akran ve öğretmen deđerlendirmelerinin çok yüzeyle Rasch modeli ile birlikte analiz yapıldıđı çalıřmaya rastlanmamıřtır.

BÖLÜM II

YÖNTEM

Bu bölümde araştırmanın modeli, çalışma grubu, veri toplama araçları hakkında bilgiler verilmiş; araştırma için kullanılan verilerin toplanması ve verilerin hangi analiz ile yapıldığı belirtilmiştir.

2.1.Araştırmanın Modeli

Araştırmada ilköğretim altıncı sınıf öğrencilerinin problem çözme becerilerini değerlendirmeleri üzere; öğrenci performansları öz, akran ve öğretmen tarafından dereceli puanlama anahtarı kullanılarak puanlanmış ve süreç içerisinde öz ile akran puanlayıcıların öğretmen puanlamalarına göre izlenmesi amaçlanmıştır. Araştırma kapsamında üç haftanın uygulama verisi ayrı olarak ve bu üç uygulamanın bulunduğu veri ÇYRÖM ile analiz edilmiştir.

Bu çalışmada mevcut konu hakkında detaylı bilgi toplama ve konuyu açıklama amaçlandığı için betimsel bir araştırmadır. Betimsel araştırmalarda ilgi duyulan bir konu hakkında detaylı bilgi toplama ve konuyu tanımlama amaçlanır. Olayların, objelerin, varlıkların, kurumların, grupların ve çeşitli alanların “ne” olduğunu betimlemeye, açıklamaya çalışan araştırmalardır (Punch 1998).

2.2.Çalışma Grubu

Araştırmada çalışma grubu olarak, 2014-2015 eğitim-öğretim yılı bahar döneminde, Ankara ili, Keçiören ilçesinde yer alan Hacı Sabancı Orta Okulu isimli devlet okulunda öğrenim gören 6. sınıf düzeyindeki iki şubedeki öğrenciler ile üç öğretmenden

oluşmaktadır. Çalışmanın altı haftalık uygulama aşamasında, öğrenci puanlarının öğretmen puanlarına göre gelişimi amaçlar arasında bulunmasından dolayı sürecin bütünlüğü önem taşımaktadır. Bu sebeple performans görevi uygulamalarına herhangi kısmına katılmayan öğrenciler analize dâhil edilmemiştir. Nitekim başlangıçta iki şube toplamı öğrenci sayısı 75 iken yapılan analizlerde üç performans görevi uygulamasına katılmış olan 57 öğrenciye ait veriler kullanılmıştır. Tablo 2’de çalışma grubunu oluşturan öğrencilerin cinsiyetlerine göre dağılımları verilmiştir.

Tablo 2. Çalışma Grubunun Cinsiyetlerine Göre Frekans ve Yüzdeleri

	Çalışma Grubu	
	f	%
Cinsiyet		
Kız	29	50,9
Erkek	28	49,1
Toplam	57	100

Çalışmada, öğrencilerin problem çözme becerilerinin değerlendirilmesi için performans görevlerini puanlamak üzere üç öğretmen puanlayıcı belirlenirken gönüllülük esas alınmıştır. Farklı bölgelerde görev yapmış öğretmen puanlayıcı grubunun özellikleri; R1 ve R3 puanlayıcıları matematik öğretmeni olup matematik eğitimi alanında uzman seviyesindedir. R2 puanlayıcısı da yine matematik öğretmeni olup ölçme ve değerlendirme alanında yüksek lisans yapmaktadır. Çalışma tecrübesi açısından üç puanlayıcının da deneyimi 3-5 yıl arasında değişmektedir.

İlköğretim dönemi zihinsel beceri gelişiminin diğer dönemlere göre daha hızlı olan yıllar olduğu (Baykul, 2006, s.61) ve sınıf düzeyi yükseldikçe öğrencilerin kendilerini nesnel değerlendirmekte zorlandığı (Bourke ve Positt’den aktaran Kutlu vd. 2010, s.96) göz önünde bulundurulduğu için araştırmada ilköğretim düzeyi öğrenciler tercih edilmiştir. Öğrencilerle akran ve öz değerlendirme çalışması yapılmadan önce konu ile ilgili eğitim verilmesinin çok önemli olduğu için (Uysal, 2008) öğrenci ve öğretmenlere konu ile ilgili eğitim verilmiştir. Öğrenciler süreçte hem kendilerini hem de akranlarını değerlendirdiği için puanlayıcı türü olarak hem “öz” hem de “akran” olarak yer almışlardır. Öğretmen puanlayıcılar ise matematik alanında probleme dayalı öğretimi bilen, problem çözme becerisine yönelik sınıf içi uygulamalar gerçekleştiren ve dereceli puanlama anahtarı

hakkında yeterli bilgi düzeyine sahip üç ilköğretim matematik öğretmeninden oluşmaktadır.

2.3. Veri Toplama Araçları

Araştırmanın amaçları doğrultusunda, öğrencilerin problem çözme becerilerini ölçmek üzere rutin olmayan problem durumu içeren dört tane performans görevi araştırmacı tarafından geliştirilmiştir. Söz konusu görevlerin öz, akran ve öğretmenler tarafından değerlendirilmeleri için yine araştırmacı tarafından geliştirilen analitik dereceli puanlama anahtarları (rubrik) kullanılmıştır.

2.3.1. Performans Görevleri

Performans görevleri hazırlanırken öncelikle öğrencilerin problem çözme becerisini günlük yaşamda kullanmaya ihtiyaç duyabileceği öğrenci seviyesine uygun problem durumları oluşturulması amaçlanmıştır. Alan yazında yer alan rutin ve rutin olmayan problemler de incelenerek araştırmacı tarafından 6. sınıf öğrencilerine yönelik rutin olmayan problem durumunu içeren dört adet performans görevi hazırlanmıştır.

Her bir performans görevinin uygunluğu birbirinden bağımsız sekiz kişiden oluşan uzman grubundan alınan görüşler çerçevesinde belirlenmiştir. Uzman grubunu; bir dilbilimci ve aynı zamanda ölçme ve değerlendirme alanında Dr., ölçme ve değerlendirme alanında bir Yard. Doç. Dr ve bir Doç. Dr olmak üzere iki öğretim üyesi, öğretmenlik deneyimi olan hem matematik eğitimi hem de ölçme ve değerlendirme alanlarında üç uzman ve halen MEB' te görev yapan iki ilköğretim matematik öğretmeni oluşturmaktadır. Ek 2 'de verilen performans görevlerinin geliştirme sürecinde görevlerin etkililiğine ilişkin kullanılan uzman görüş formu Ek 1.a'da verilmiştir.

Performans görevlerini değerlendirmek üzere analitik dereceli puanlama anahtarı hazırlanırken öncelikle beklenen performansa göre amaç belirlenmiştir. Daha sonra problem çözme sürecinin aşamaları incelenerek beklenen performansa göre aşamalar yazılmıştır. Tanımlanan problem çözme aşamalarının performans düzeylerine karşılık gelen niteliklileri puanlayıcıdan puanlayıcıya değişmemesi için olabildiğince somut olarak belirlenmeye çalışılmıştır.

Son olarak analitik dereceli puanlama anahtarı, yukarıda açıklanan sekiz kişiden oluşan aynı uzman grubuna Ek 1.b'de bulunan Dereceli Puanlama Anahtarı Uzman Değerlendirme Formu'ndaki ölçütler dikkate alınarak gönderilmiştir. Uzmanlardan alınan görüş ve öneriler ile Ek 3.a da öğretmene yönelik, Ek3.bde öğrenciye yönelik olan dereceli puanlama anahtarlarına son hali verilmiştir. Öğrenciye yönelik hazırlanan DPA'da öğretmene yönelik DPA'dan farklı olarak, performansa ait seviyeler rakamlar yerine harfler ile simgelenerek sözel olarak da motive edici ifadeler yer verilmiştir.

Performans görevlerinin ve dereceli puanlama anahtarının psikometrik özelliklerine yönelik kanıt toplamak için uzman görüşlerinden faydalanılarak ölçme araçlarının kapsam geçerliliği belirlenmeye çalışılmıştır.

Ölçek geliştirme çalışmalarında deneysel uygulamaların yapılamadığı durumlarda uzman görüşlerine dayalı nitel çalışmaların nicel çalışmalara dönüştürülmesi amacıyla kullanılan kapsam geçerlik oranlarına başvurulmaktadır (Yurdugül, 2005). Lawshe (1975) tarafından geliştirilen kapsam geçerlik oranları 3'lü likert türünde ölçekler için kullanılmaktadır. Bu üçlü derecelendirmede her bir madde için maddenin yapıyı ölçmesi ile ilgili olarak "gerekli", "yararlı ama gerekli değil" ve "gereksiz" uzman görüşleri alınır. Kapsam geçerlik oranları (KGO) herhangi bir maddeye ilişkin "gerekli" görüşünü belirten uzman sayısının, maddeye ilişkin görüş belirten tüm uzman sayısının yarısına oranının 1 eksiği olarak hesaplanmaktadır. Buna göre aşağıda KGO için formül belirtilmiştir:

$$KGO = \frac{NG}{N/2} - 1$$

NG = Gerekli görüşünü belirten uzman sayısı

N = Tüm uzmanların sayısı

Formüle göre hesaplanan KGO'ların ortalaması ile testin tümüne ait kapsam geçerlik indeksi hesaplanabilmektedir. Lawshe tekniğine göre, kapsam geçerlik için en az 5 en fazla 40 uzman görüşüne ihtiyaç vardır. Tablo 3'te uzman sayısına göre minimum kapsam geçerlik oranları verilmiştir (Lawshe, 1975).

Tablo 3. $\alpha = 0,05$ Düzeyinde Kapsam Geçerlik Oranları İçin Minimum Değerler

Uzman Sayısı	Minimum Değer	Uzman Sayısı	Minimum Değer
5	0,99	13	0,54
6	0,99	14	0,51
7	0,99	15	0,49
8	0,78	20	0,42
9	0,75	25	0,37
10	0,62	30	0,33
11	0,59	35	0,31
12	0,56	40+	0,29

Çalışma kapsamında oluşturulan performans görevleri ve puanlama anahtarının uygulamada yer alıp almamasına karar verilirken kapsam geçerliği için Lawshe tekniğinin sonuçları dikkate alınmıştır. Bu noktadan hareketle her bir performans görevi ve dereceli puanlama anahtarı için kapsam geçerlik indeksi (KGİ) hesaplanmıştır. Çalışmada ölçme araçlarının ölçütlere uygunluk açısından “uygun”, “düzeltilmeli” ve “uygun değil” düzeyleri ile 9 uzman görüşünden yararlanılarak KGO’lar hesaplanmıştır. Tablo 3’e göre 9 uzman görüşü için kapsam geçerlik ölçütü (KGÖ) 0,75 değeri olduğu gözükmektedir. Buna göre her bir ölçme aracı için elde edilen KGİ’ler 0,75 olan KGÖ değeri ile karşılaştırılmıştır. Dört performans görevi için hesaplanan KGİ’ler tablo 4’te verilmiştir.

Tablo 4. Performans Görevleri Kapsam Geçerliği

	Görev 1	Görev 2	Görev 3	Görev 4
1. Ölçüt	1	0,78	0,78	0,78
2. Ölçüt	1	1	0,78	1
3. Ölçüt	0,78	0,78	0,78	0,78
4. Ölçüt	1	0,78	1	0,78
5. Ölçüt	0,56	0,78	0,78	0,56
Kapsam Geçerlik İndeksi	0,87*	0,82*	0,82*	0,78*
Kapsam Geçerlik Ölçütü				0,75

Tablo 4’e göre 9 uzman tarafından değerlendirilen dört performans görevine ait kapsam geçerlik indekslerinin kapsam geçerlik ölçütünden yüksek değerde olduğu gözlenmiştir.

Çalışmada kullanılan puanlama anahtarına ilişkin kapsam geçerlik değerleri aşağıda bulunan Tablo 5’te gösterilmiştir.

Tablo 5. Dereceli Puanlama Anahtarı Kapsam Geçerliği

Analitik Dereceli Puanlama Anahtarı	
1. Ölçüt	1
2. Ölçüt	1
3. Ölçüt	0,78
4. Ölçüt	1
5. Ölçüt	0,56
6. Ölçüt	1
Kapsam Geçerlik İndeksi	0,89*
Kapsam Geçerlik Ölçütü	0,75

Tablo 5’e göre 9 uzman tarafından değerlendirilen analitik dereceli puanlama anahtarı için hesaplanan kapsam geçerlik indeksinin kapsam geçerlik ölçütünden yüksek olduğu görülmüştür.

2.4.Verilerin Toplanması

Araştırmacı tarafından toplanan verilerin elde edilmesine ilişkin süreç aşağıda belirtilmiştir.

2.4.1.Program Tanıtım ve Deneme Uygulaması

Altı haftalık uygulama programının ilk haftası tanıtım, bir hafta da deneme uygulaması olmak üzere iki haftası program tanıtım ve deneme uygulama sürecine ayrılmıştır. İlk hafta öğretmen ve öğrenciler ile tanışılmış, niçin beraber bir çalışma yürütüleceğinden, sürecin içerik ve öneminden bahsedilmiştir. İkinci hafta, süreçte uygulanacak performans görevlerinin aşamaları ve dereceli puanlama anahtarı hakkında bilgilendirme yapılmıştır. Eğitim esnasında öğrencilerden gelen sorular yanıtlanmaya çalışılmış ve öğrenci motivasyonunun uygulama süreci boyunca devamlılığını sağlamak için bazı uygulamalar hakkında ön bilgi verilmiştir. Deneme olarak ilk performans görevi uygulanmış devamında

öğrenciler önce kendi çalışmalarını sonra rasgele olarak dağıtılan kâğıtlardan herhangi bir arkadaşının çalışmalarını puanlamışlardır. Uygulamada bir ders saati içerisinde performans görevi uygulaması, öğrencinin kendini değerlendirmesi ve iki akranını puanlaması planlanmaktaydı fakat deneme uygulamasından sonra süre sınırlı olduğu için bir akran puanlamasına karar verilmiştir.

2.4.2. Performansa Dayalı Durum Belirleme Süreci

Bu aşamada, üç hafta boyunca öğrencilere performans görevlerinin uygulaması yapılmıştır. Her hafta bir önceki haftaki öğrencilerin puanlamalarının objektifliği hakkında genel bir bilgilendirme yapılarak sonraki uygulama başlatılmıştır. Performans görevi uygulamaları için öğrencilere yaklaşık 15-20, öz değerlendirme için 10, akran değerlendirme için de 10 dakika süre tanınarak uygulamalar bir ders saati içerisinde tamamlanmıştır. Ayrıca öğrenci motivasyonunun uygulama süreci boyunca devamlılığını sağlamak için bazı uygulamalar yapılmıştır. Altıncı hafta, öğrenci görüşlerini almak için yarı yapılandırılmış sorular yöneltilmiş, yazılı olarak toplanmıştır.

Son olarak öğretmenlere araştırmacı tarafından ölçme araçları tanıtılmış ve öğrenci performanslarının dereceli puanlama anahtarı ile puanlanması için çalışmaya gönüllü olarak katılan üç öğretmen puanlayıcıya gönderilmiştir.

2.5. Verilerin Analizi

Araştırmanın verileri, 57 öğrencinin problem çözme becerilerini ölçmeye yönelik (deneme haftası dışında kalan) üç performans görevi uygulamasının öz, akran ve üç öğretmen tarafından DPA ile puanlanması ile elde edilmiştir. Veriler araştırma sorularına göre düzenlenmiştir. Araştırmanın ilk üç sorusunu yanıtlamak adına birinci, ikinci ve üçüncü uygulamaya ait verilerin analizi için yüzeyle; öğrenci, aşama, puanlayıcı türü ve puanlayıcı olmak üzere belirlenmiştir. Devamında araştırmanın dördüncü sorusunu yanıtlamak adına üç uygulamanın verilerini içeren genel analiz için yüzeyle; öğrenci, aşama, puanlayıcı türü, puanlayıcı ve görev olarak belirlenmiştir. Genel analizde; öz, akran ve öğretmen puanlayıcıların katılık/esnekliğini her görev için hafta hafta izlenebilmesi adına ayrı kodlanmıştır. Düzenlenen verilerin analizini gerçekleştirmek için çok yüzeyle Rasch ölçme modeli yaklaşımı kapsamında FACETS 3.71.4 (Linacre, 1987-2014) paket programı kullanılmıştır.

Araştırmada ÇYRÖM 'yi uygulamadan önce model veri uyumunun incelenmesi gereklidir. Model veri uyumunun sağlanması için standartlaştırılmış artık değerlerin yaklaşık en fazla %5'i -2 ile +2 dışında ve yaklaşık en fazla %1'i -3 ile +3 dışında olması gerekmektedir (Linacre, 2014). Buna göre yapılan dört analiz için sırasıyla model veri uyum değerleri aşağıda açıklanmıştır:

- Problem çözme becerilerinin ölçülmesine yönelik yapılan birinci performans görevi uygulamasından alınan 1100 verinin standartlaştırılmış artık değerlerinin 44'ü yani %4'ü -2 ile +2 ve 9'u yani %0,8'i -3 ile +3 dışında yer almıştır. Dolayısıyla birinci performans görevi uygulamasından alınan verilerin FACET analizi için model uyumunun sağladığı görülmektedir.
- Problem çözme becerilerinin ölçülmesine yönelik yapılan ikinci performans görevi uygulamasından alınan 1092 verinin standartlaştırılmış artık değerlerinin 44'ü yani yaklaşık %4'ü -2 ile +2 ve 12'si yani yaklaşık %1'i -3 ile +3 dışında yer almıştır. Dolayısıyla, ikinci performans görevi uygulamasından alınan verilerin FACET analizi için model uyumunun sağladığı söylenebilir.
- Problem çözme becerilerinin ölçülmesine yönelik yapılan üçüncü performans görevi uygulamasından alınan 852 verinin standartlaştırılmış artık değerlerinin 31'i yani yaklaşık %3,6'sı -2 ile +2 ve 8'i yani yaklaşık %0,9'u -3 ile +3 dışında yer aldığı için üçüncü uygulamadan alınan verilerin FACET analizi için model uyumu sağlamıştır.
- Problem çözme becerilerinin ölçülmesine yönelik yapılan üç uygulamanın verilerinin bulunduğu genel analizdeki 2928 verinin standartlaştırılmış artık değerlerinin 100'ü yani yaklaşık % 3,4'ü -2 ile +2 ve 21'i yani yaklaşık % 0,7'si -3 ile +3 dışında yer almaktadır. Sonuç olarak üç uygulama verilerinin de içinde bulunduğu genel analiz verilerinin FACET analizi için model uyumu gösterdiği görülmüştür.

2.5.1.Çok Yüzeyle Rasch Ölçme Modeli

Danimarkalı bir matematikçi George Rasch, bir parametrelili lojistik model ya da Rasch modeli olarak bilinen madde tepki modelinde bireylerin yetenek düzeyini ve maddelerin güçlük düzeyini aynı zamanda tanımlamaya çalışmıştır (Linacre'den aktaran Atılğan, 2004). Linacre (1989) Rasch'ın bir parametrelili modeline, puanlayıcı katılık/cömertliği,

soru güçlüğü gibi yüzeyleri eklenerek çok parametrelili bir modele dönüştürmüştür. Çok yüzeyli Rasch ölçme modelinde odak noktası her birey için puanlayıcı, madde ve görev etkileri ortadan kaldırıldığında değerlendirilen her birey için ölçümüdür (aktaran Linacre, 1993). Rasch yöntemlerinin avantajı insan yargılarına dayalı değerlendirmeleri olabildiğince nesnelleştirme girişiminde olmasıdır (Hetherman 2004).

ÇYRÖM her bir yüzey için yüzeyler arası olması gereken doğrusal bir bağlantının kurallarını belirlemede ve bu bağlantıları oluşturmaktadır (Baştürk, 2010). Böylece aynı logit cetvelde yüzeylere ait değerler yerleştirilir.

Çok yüzeyli Rasch ölçme modeli genel olarak bireylerin yetenek seviyelerini, görev veya maddelerin güçlük seviyelerini ve puanlayıcıların katılıklarını kestirmek üzere üç değişkenlik kaynağı içerir (Güler, 2008). Bu değişkenlik kaynakları aşağıda verilen eşitlikteki gibi belirtilmiştir (Linacre, 1993).

$$\text{Log}(P_{nij}/P_{nij-1}) = B_n - D_i - C_j - F_k$$

P_{nij} : (n) bir bireyin (i) bir maddeye gösterdiği performansa (j) bir puanlayıcının k kategorisinde verdiği puanın olasılığı

P_{nij-1} : (n) bir bireyin (i) bir maddeye gösterdiği performansa (j) bir puanlayıcının k kategorisinde verdiği puanın olasılığı

B_n : (n) bir bireyin yetenek düzeyi

D_i : (i) bir maddenin güçlük düzeyi

C_j : (j) bir puanlayıcının katılık düzeyi

F_k : k-1 kategorisinden k kategorisine geçişin güçlük düzeyi

Regresyon analizi alan yazınına göre; bağımlı değişken, başarılı kategoriden olma olasılığı oranının logaritmik dönüşümü, bağımsız değişkenler de değişkenlik kaynaklarıdır (Hetherman, 2004). Gözlenen puanların logaritmik olarak logit ölçeğine dönüşümüne dayalı doğrusal bir model olan çok yüzeyli Rasch ölçme modelinde de (Hetherman, 2004) bireylerin yetenekleri, belirli madde dağılımlarının özelliklerinden, belirli puanlayıcıların performansa verdikleri puanlardan bağımsız bir şekilde kestirilir (Smith ve Kulikowich, 2004). Her bir değişkenlik kaynağının her bir elemanına ilişkin beklenmedik durumlarda performansların durumu hakkında bilgi verir (Alharby, 2006).

Çok yüzeyli Rasch ölçme modelinde her değişkenlik kaynağı için ayrı ayrı parametreler hesaplanmakta ve ayrı ayrı güvenilirlik katsayıları kestirilmektedir. Bireylerin yetenekleri tüm puanlayıcıların tüm maddeler için verilen puanlardan kestirilmektedir (Güler, 2008). FACETS programının analizi sonucunda tekrarlama (iteration) raporu, yüzeylerin haritası,

her deęişkenlik kaynađına iliřkin logit ölçme kestirimleri standart hata ve model uyum istatistikleri gibi farklı istatistik sonuçları elde edilir (Linacre, 1993). Veriden iyi kestirimde bulunabilmek için gerekli iterasyon sayısı verinin Rasch modeline ne kadar uyumlu olduđuna bađlıdır. Normal yaklaşım algoritmi ve koşulsuz maksimum olasılık algoritmi kullanılarak mümkün olduđunca hızlı ve kesin kestirime ulařılır (Linacre'den aktaran Güler, 2008).

Çok yüzeyli Rasch ölçme modelinde veri matrisindeki uyumsuzlukların derecesini gösteren uyum istatistikleri her madde için artıkların yönü ve büyüklüğü hakkında bilgi verir. Veriler modele yaklařtıkça standartlařtırılmıř artığın kestirim ortalaması 0 ve varyansı 1'e yaklařır (Hetherman, 2004). Her bir deęişkenlik kaynađının her bir elemanı için hesaplanan standartlařtırılmıř artıkların kareler ortalaması deęişkenlik kaynađının modelle olan uyumunu gösteren dıř uyum (outfit) deđeridir. İç uyum (infit) istatistiđinin, dıř uyum istatistiđinden farkı standartlařtırılmıř kareler ortalamasının ađırlıklandırılmıř durumu olmasıdır. Puanlayıcı deęişkenlik kaynađı için iç ve dıř uyum istatistikleri modelle kestirilen puanlayıcı katılık parametrelerinin tutarlılıđı için kanıt ortaya koymaktadır (Güler, 2008).

ÇYRÖM' nin hesapladıđı diđer bir istatistik olan ayırma indeksi deęişkenlik kaynađındaki elemanların birbirinden ayrılma derecesini gösterir. Puanlayıcı deęişkenlik kaynađı için belirtilen deđerin 0'a yakın, madde ve bireyler için büyük olması beklenir. Puanlayıcı yüzeyine ait ayırma indeksinin büyük olması puanlayıcıların ne kadar farklı puanlama yaptıklarını göstergesidir. Ayırma indeksi (G) $(4G + 1)/3$ formülü ile her bir deęişkenlik kaynađı için, deęişkenlik kaynađındaki elemanların kaç düzeye ayrılabilceđini göstermektedir (Hetherman 2004). Ayırma indeksinin güvenilirliđi; klasik test kuramındaki KR-20 Cronbach alfa veya genellenebilirlik kuramındaki genellenebilirlik katsayısına eř deđer olduđu için iç tutarlılık olarak yorumlanabilir (Nakamura, 2000). Dolayısıyla ayırma indeksinin bire yaklařması, puanlayıcıların katılıđı açısından birbirinden ne kadar güvenilir şekilde ayrıldıđını gösterir (Güler, 2008).

FACETS analizi sonucu hesaplanan diđer bir istatistik de her bir deęişkenlik kaynađındaki elemanların birbirinden farklılařmasının istatistiksel olarak anlamlı olup olmadıđını test eden X^2 (ki-kare) deđeridir. X^2 testinin anlamlı olması yüzeye ait elemanlar arasındaki farklılıđın anlamlı olduđuna iřarettir.

BÖLÜM III

BULGULAR

Bu bölümde, araştırma sürecinde toplanan verilerin analizleri sonucu elde edilen bulgulara yer verilmiştir. Alt problemler sıra ile ele alınmıştır.

3.1. Birinci Alt Probleme İlişkin Bulgular

Alt problem 1: Birinci performans görevi uygulamasında problem çözme becerisinin öz, akran ve öğretmen puanlarının çok yüzeyli Rasch ölçme modeliyle analizi sonucu kalibrasyon haritasının durumu nedir?

Öğrencilerin problem çözme becerilerini ölçmeye yönelik birinci performans görevi uygulamasının öz, akran ve öğretmen puanlarının çok yüzeyli Rasch ölçme modeli ile analizinde öğrenci, aşama, puanlayıcı türü ve puanlayıcı olarak kullanılan dört yüzeyin kalibrasyon haritası Şekil 2’de yer almaktadır. Şekil 2’de + ve – arasında her bir yüzey için ölçümlerin bulunduğu logit cetveli bulunmaktadır. Problem çözme becerilerine ait performansı gösteren logit cetvelde yukarıdan aşağıya doğru öğrenci yetenekleri azalmaktadır. İkinci yüzey olarak aşama, öğrencilere en kolay gelen aşamadan en zor gelen aşamaya doğru sıralanmaktadır. Üçüncü yüzey olan puanlayıcı türünde puanlayıcıların katılığı logit cetvelde en cömertten en katıya doğru sıralanmıştır. Dördüncü yüzey olarak belirlenen puanlayıcı yüzeyi yapay (dummy) yüzey olarak tanımlanarak içindeki elemanlar sıfıra sabitlemiştir.

Measr	+Öğrenci	+Aşama	+Tür	RATIN
4	+	+	+	(4)
	7			
	24 52			
	18			
3	+	+	+	
	39 48 50			
	20 9			
	29 32 36 44			
	45			
	49			
	12 30 55			---
2	+	+	+	
	16 17 37 51			
	13 15 42 5 8			
	26 27 34 4 40			
	11			
	31			3
	19			
1	+	+	+	
	23 53		Öz	---
	33 35 41			
	47			
	10	Uygulama		
	22 28 3 46 56			2
	1 54 6			
*	0 *	* Anlama	* Akran	*
	14	Çözüm yolu		
	38	Kontrol etme		---
			Öğretmen	
-1	+	+	+	1
-2	+	+	+	
	57			

-3	+	+	+	
	43			
-4	+	+	+	(0)
	2 21 25			
Measr	+Öğrenci	+Aşama	+Tür	RATIN

Şekil 2. Birinci analizdeki uygulamaya ait kalibrasyon haritası

Şekil 2' deki kalibrasyon haritasına bakıldığında öğrencilerden en yüksek performansı gösteren 7, en düşük performansı gösterenlerin 2, 21 ve 25 numaralı öğrenciler olduğu görülmektedir. Öğrencilerin genel olarak logit cetveldeki yeri pozitif alanda olduğu için genel olarak grubun başarılı olduğu söylenebilir. Dereceli puanlama anahtarındaki problem çözme aşamalarının zorluk/kolaylık düzeyini gösteren logit cetvelde en kolay cevaplanan

“Uygulama” (seçilen çözüm stratejisini uygulama) aşaması olurken en zor cevaplanan aşama, “Kontrol etme” (çözümü kontrol etme ve değerlendirme) aşaması olmuştur. “Anlama” (problemi anlama) ile “Çözüm yolu” (çözüm ile ilgili stratejinin seçimi) aşamalarının ise yaklaşık orta düzey zorlukta olduğu görülmüştür. Puanlayıcı türü olarak öz, akran ve öğretmen puanlarının bulunduğu logit cetvelde; en cömert “Öz”, en katı puanlayıcı “Öğretmen” puanları olduğu görülürken “Akran” puanlarının orta düzey katılımda olduğu gözlenmiştir.

Alt problem 1: a) Birinci performans görevi uygulamasında öğrencilerin başarı durumu ölçüm raporuna ait istatistikler nedir?

Birinci uygulamada öğrencilerin problem çözme becerilerinin başarı durumu ölçüm raporuna ait değerler şekil 3’te belirtilmiştir.

Gözlenen Puan	Gözlem Sayısı	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçü Uyg.		Dışü Hes.		Nu Öğrenci	
				Ölçüm	Hata	K.O.ZStd	K.O.ZStd	Ayrırt.	Nu		
76	20	3.80	3.90	3.70	.50	.81	-.1	.77	-.1	1.00	7 7
75	20	3.75	3.88	3.56	.45	.86	-.1	.69	-.4	1.13	24 24
60	16	3.75	3.88	3.56	.50	1.70	1.1	1.53	.9	.94	52 52
70	20	3.50	3.82	3.15	.35	1.13	.4	.96	.0	.94	18 18
58	16	3.63	3.80	3.04	.41	.82	-.2	1.17	.4	.83	50 50
74	20	3.70	3.80	3.03	.41	1.28	.6	1.14	.4	.97	48 48
71	20	3.55	3.78	2.95	.35	.88	-.1	.78	-.3	1.06	39 39
73	20	3.65	3.76	2.85	.39	.80	-.3	.66	-.3	1.06	20 20
72	20	3.60	3.76	2.84	.36	.77	-.4	.88	.0	1.02	9 9
81	24	3.38	3.70	2.64	.29	.86	-.3	1.75	1.5	.69	44 44
71	20	3.55	3.70	2.64	.35	1.22	.6	1.10	.3	.89	29 29
71	20	3.55	3.70	2.62	.35	1.38	.9	1.12	.4	1.02	36 36
71	20	3.55	3.69	2.60	.35	1.10	.3	1.06	.2	.95	32 32
71	20	3.55	3.67	2.54	.35	.36	-1.9	.32	-1.6	1.41	45 45
69	20	3.45	3.62	2.40	.32	.72	-.7	.67	-.7	1.25	49 49
62	20	3.10	3.55	2.21	.26	.82	-.5	.94	.0	.91	55 55
52	16	3.25	3.54	2.20	.31	1.06	.2	1.36	.8	.63	30 30
51	16	3.19	3.54	2.19	.30	.90	-.1	.83	-.3	1.16	12 12
66	20	3.30	3.49	2.08	.29	1.16	.5	1.14	.4	1.00	37 37
64	20	3.20	3.48	2.06	.28	.55	-1.6	.83	-.2	1.28	16 16
62	20	3.10	3.47	2.04	.27	.95	.0	.90	-.2	.79	17 17
65	20	3.25	3.45	2.00	.29	.54	-1.6	.50	-1.5	1.40	51 51
61	20	3.05	3.39	1.90	.26	.46	-2.2	.41	-1.7	1.60	8 8
48	16	3.00	3.38	1.88	.29	.77	-.6	.88	-.1	1.03	13 13
61	20	3.05	3.37	1.85	.26	.59	-1.5	.60	-1.4	1.71	5 5
62	20	3.10	3.34	1.82	.27	.62	-1.3	1.41	.8	1.00	15 15
59	20	2.95	3.34	1.82	.25	.80	-.6	.81	-.5	1.14	42 42
62	20	3.10	3.29	1.74	.28	1.13	.5	1.08	.3	.93	40 40
58	20	2.90	3.27	1.70	.25	.90	-.2	1.16	.5	.97	4 4
63	20	3.15	3.25	1.67	.30	.42	-2.2	.44	-2.1	1.96	34 34
45	16	2.81	3.25	1.67	.27	1.15	.5	1.11	.4	.94	26 26
60	20	3.00	3.21	1.62	.27	.78	-.7	.74	-.6	1.29	27 27
43	16	2.69	3.14	1.52	.27	1.12	.4	1.42	1.2	.76	11 11
41	16	2.56	2.95	1.29	.28	1.91	2.4	1.78	1.9	.26	31 31
55	24	2.29	2.90	1.23	.21	.88	-.4	.92	-.2	.72	19 19
41	20	2.05	2.73	1.05	.23	1.27	1.0	1.24	.9	.93	53 53
64	24	2.67	2.67	.99	.24	1.18	.7	1.24	.8	.83	23 23
36	16	2.25	2.57	.89	.29	1.05	.2	.96	.0	1.10	35 35
66	24	2.75	2.50	.82	.28	1.07	.3	1.12	.4	1.02	41 41
45	20	2.25	2.49	.81	.25	2.03	2.8	1.98	2.6	-.74	33 33
41	20	2.05	2.32	.63	.25	1.29	1.0	1.31	1.0	.89	47 47
29	16	1.81	2.24	.55	.28	.62	-1.2	.56	-1.3	1.45	10 10
31	16	1.94	2.07	.37	.31	.84	-.3	.80	-.4	1.30	3 3
57	32	1.78	2.06	.37	.21	1.22	.8	1.21	.8	.89	56 56
48	20	2.40	2.04	.35	.33	1.21	.6	1.05	.2	1.21	46 46
22	16	1.38	2.01	.31	.32	2.65	3.0	2.78	3.1	-.20	28 28
39	20	1.95	1.96	.27	.28	.77	-.6	.73	-.7	1.21	22 22
43	20	2.15	1.91	.21	.29	.66	-.9	.70	-.7	1.36	1 1
36	20	1.80	1.88	.17	.29	.52	-1.4	.53	-1.3	1.30	54 54
33	20	1.65	1.80	.08	.28	.81	-.4	.73	-.7	1.38	6 6
29	16	1.81	1.75	.02	.34	.44	-1.5	.45	-1.4	1.39	14 14
29	20	1.45	1.59	-.18	.29	.35	-2.4	.36	-2.2	1.54	38 38
10	8	1.25	.63	-2.19	.47	2.51	2.0	2.67	2.1	-.37	57 57
6	20	.30	.31	-3.28	.49	1.49	1.3	1.37	.9	.58	43 43
0	20	.00	.04	(-5.50	1.85)	Minimum					2 2
0	20	.00	.04	(-5.55	1.85)	Minimum					25 25
0	16	.00	.03	(-5.93	1.85)	Minimum					21 21
51.0	19.3	2.63	2.82	1.15	.39	1.00	-.1	1.03	.0		Ort. (N:57)
20.5	3.1	.98	1.05	2.06	.35	.47	1.2	.49	1.1		S.D

RMSE (Model)= 0,32 Ayırma İndeksi= 3,99 Güvenirlilik=0,94
Tamamı aynı ki-kare değeri= 715,0 s.d.= 56 p= 0.00
Rasgele normal ki-kare değeri= 41,1 s.d.= 55 p= 0,92

Şekil 3. Birinci uygulamada öğrencilere ait ölçüm raporu

Şekil 3'te görüldüğü üzere problem çözme becerilerinin ölçüldüğü toplam 57 öğrencinin logit değerleri 3,70 ile -5,93 arasında değişim göstermektedir. Öğrencilerin uygulanan

performans görevine ait başarı ölçümündeki standart hatayı gösteren RMSE değeri 0,32 bulunmuştur. RMSE, aşırı uçlarda yer alan değerler hariç bütün veriler için ölçme hatasını göstermektedir (Baştürk, 2010). Değişkenlik kaynaklarına ilişkin elde edilen ölçmelerin sürekliliği boyunca birbirlerinden ne derece ayrıldığını gösteren istatistik olan (Turner'dan aktaran Güler, 2008) ayırma indeksi 3,99 olarak bulunmuştur. Bireyler için yüksek olması beklenen bu değer, bireylerin problem çözme becerilerinin birbirinden ne kadar farklı olduğunu göstermektedir. Ayırma indeksinin (G) kullanıldığı $(4G + 1)/3$ bir yüzeydeki elemanların kaç farklı düzeye ayrıldığını gösteren formüle göre (Hetherman, 2004), bireylerin bu göreve ait problem çözme performanslarının yaklaşık altı düzeye ayrılacağı görülmüştür. Dolayısı ile bu değer öğrenci performanslarının puanlamasında merkeze yönelme etkisi olmadığına işaret etmektedir. Öğrencilerin problem çözme becerilerinde gösterdiği performans sıralamasının güvenilirliğini gösteren istatistik 0,94 olarak yeterince yüksek bir değer olarak bulunmuştur. Diğer bir deyişle öğrenciler güvenilir bir şekilde farklı beceri düzeylerine ayrılmıştır. Sabit etkiye ait “Öğrencilerin problem çözme becerilerini ölçmeye yönelik birinci performansları arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2=715,0$, $sd=56$, $p=0.00$) yokluk hipotezi reddedilmiştir. Yani, öğrencilerin problem çözme becerilerini ölçmeye yönelik birinci performansları arasında istatistiksel açıdan anlamlı farklılıklar bulunmaktadır.

Rasch analizinde iç uyum (infit) ve dış uyum (outfit) değerlerinin 0,5 ile 1,5 arasında olması beklenir (Linacre, 2014). Myford ve Wolfe (2003) 1,5 ile 2 arasındaki değerler verilerin ölçüm için yararlı olmadığını ancak 2'ye kadar olan uyum istatistiklerinin kabul edilebilir olarak nitelendirmiştir (Aktaran İlhan, 2015). 57 öğrenci arasından 55 öğrencinin yani, öğrencilerin yaklaşık % 96'sı uyum içi ve uyum dışı kareler ortalamalarının kabul edilebilir aralıkta olduğu gözlenmiştir.

Alt problem 1: b) Birinci performans görevi uygulamasında problem çözme aşamalarına ait ölçüm raporundaki istatistikler nedir?

Birinci performans görevi uygulamasında öğrencilerin problem çözmelerinin değerlendirilmesinde kullanılan puanlama anahtarındaki problem çözme aşamalarının zorluk/kolaylıklarına ilişkin istatistikler Tablo 6'da verilmiştir.

Tablo 6. Birinci Uygulamada Problem Çözme Aşamalarına Ait Ölçüm Raporu

No Aşamalar	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçi		Uygunluk Dışı		
			Ölçüm	Hata	Kareler Ort.	Z Std.	Kareler Ort.	Z Std.	
3 Uygulama	2,89	3,44	0,44	0,08	1,02	0,2	0,93	-0,4	
1 Anlama	2,68	3,19	0,06	0,08	1,06	0,6	1,22	1,6	
2 Çözüm yolu	2,54	3	-0,19	0,08	0,92	-0,9	0,99	-0,1	
4 Kontrol	2,47	2,89	-0,31	0,08	0,95	-0,5	0,93	-0,6	
Ortalama	2,64	3,13	0	0,08	0,99	-0,2	1,02	0,2	
Std. sapma	0,16	0,21	0,29	0	0,06	0,6	0,12	1	
RMSE (Model)= 0,08		Std sapma= 0,28		Ayırma İndeksi= 3,44				Güvenirlilik=0,92	
Tamamı aynı ki-kare değeri= 49,4		s.d.= 3	p= 0,00						
Rasgele normal ki- kare değeri= 2,8		s.d.= 2	p= 0,24						

Tablo 6’da görüldüğü üzere dereceli puanlama anahtarındaki “Anlama”, “Çözüm yolu”, “Uygulama” ve “Kontrol” olmak üzere dört problem çözme aşamasının logit değerleri 0,44 ile -0,31 arasında yer almaktadır. Problem çözme aşamalarının zorluk/kolaylıklarını gösteren logit değerlere ait standart hata (RMSE) değerinin 0,08 olduğu görülmektedir. Bu değer standart hatanın düşük olduğunu gösterir. Ayırma indeksi 3,44 ve problem çözme aşamalarının zorluk/kolaylık sıralamasının güvenirliliğini gösteren istatistik 0,92 olarak yüksek bulunmuştur. Aşamalar yüzeyi için de ayırma indeksinin yüksek olması aşamaların güçlüğünün birbirinden farklı olması anlamına gelir, bu da istenilen bir durumdur. Sabit etkiye ait “Problem çözme aşamalarının arasında zorluk/kolaylık açısından anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2 = 49,4$, $sd=3$, $p=0,00$) yokluk hipotezi reddedilmiştir. Yani, birinci performans görevi uygulamasında problem çözme aşamalarının zorluk/kolaylık düzeyleri arasında istatistiksel açıdan anlamlı farklılıklar bulunmaktadır. Rasch analizinde etkili bir ölçüm yapımı için iç uyum ve dış uyum değerlerinin 0,5 ile 1,5 arasında olması beklenir (Linacre, 2014). Tablo 6’da iç uyum ve dış uyum değerlerinin 0,92 ile 1,22 arasında yer alması verilerin modele çok iyi uyum sağladığını göstermektedir.

Alt problem 1: c) Birinci performans görevi uygulamasında puanlayıcı türüne ait ölçüm raporundaki istatistikler nedir?

Birinci performans görevi uygulamasında öz, akran ve öğretmen puanlayıcıların bulunduğu puanlayıcı türünün puanlama katılık/cömertliğine ilişkin istatistikler Tablo 7’de verilmiştir.

Tablo 7. Birinci Uygulamada Puanlayıcı Türüne Ait Ölçüm Raporu

N P. Türü	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçi		Uygunluk Dışı	
			Ölçüm	Hata	Kareler Ort.	Z Std.	Kareler Ort.	Z Std.
1 Öz	3,25	3,59	0,78	0,11	1,29	2	1,44	2,2
2 Akran	2,95	3,14	-0,02	0,11	1,09	0,7	1,04	0,2
3 Öğretmen	2,36	2,46	-0,76	0,05	0,91	-1,7	0,90	-1,7
Ortalama	2,85	3,06	0	0,09	1,1	0,4	1,13	0,3
Std. sapma	0,37	0,47	0,63	0,03	0,16	1,6	0,23	1,6
RMSE (Model)= 0,09		Std sapma= 0,62			Ayrırma İndeksi= 6,62			
Güvenirlik= 0,98								
Tamamı aynı ki-kare değeri= 181		s.d.= 2	p= 0,00					
Rasgele normal ki- kare değeri= 2		s.d.= 1	p= 0,16					

Puanlayıcı türü yüzeyini oluşturan “Öz”, “Akran” ve “Öğretmen” olmak üzere üç tür puanlayıcının katılık/cömertliğini gösteren logit değerleri sıra ile 0,78, -0,02 ve -0,76 olarak bulunmuştur. Puanlayıcı türünün puanlama katılık/cömertliklerini gösteren logit değerlere ait standart hata (RMSE) değeri 0,09 olmakla beraber bu değer aşırı uçlarda yer alan değerler hariç tüm veriler için ölçme hatasının küçük olduğuna işaret etmektedir. Bu Ayrırma indeksi 6,62 ve puanlayıcı türlerinin puanlama katılık/cömertlik sıralamasının güvenilirliğini gösteren istatistik 0,98 olarak oldukça yüksek bir değer bulunmuştur.

Sabit etkiye ait “Puanlayıcı türlerinin puanlama katılık/cömertlikleri arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2 = 181$, $sd=2$, $p=0.00$) yokluk hipotezi reddedilmiştir. Yani, birinci performans görevi uygulamasında puanlayıcı türünün puanlama katılık/cömertlikleri arasında istatistiksel açıdan anlamlı farklılıklar bulunmaktadır. Ek olarak, tablo 7’ye göre iç uyum ve dış uyum değerlerinin 0,90 ile 1,44 arasında yer alması verilerin modele çok iyi uyum sağladığını göstermektedir. Bu da puanlayıcıların tutarlı puanlama yapmış olduğu anlamına gelmektedir.

3.2.İkinci Alt Probleme İlişkin Bulgular

Alt problem 2: İkinci performans görevi uygulamasında problem çözme becerisinin öz, akran ve öğretmen puanlarının çok yüzeyli Rasch ölçme modeliyle analizi sonucu kalibrasyon haritasının durumu nedir?

Şekil 4’te öğrencilerin problem çözme becerilerini ölçmeye yönelik ikinci performans görevi uygulamasının öz, akran ve öğretmen puanlarının çok yüzeyli Rasch ölçme modeli ile analizinde öğrenci, aşama, puanlayıcı türü ve puanlayıcı olarak kullanılan dört yüzeyin kalibrasyon haritası yer almaktadır. Şekil 4’te yine her bir yüzey için + ve – arasında ölçümlerin bulunduğu logit cetveli bulunmaktadır. Problem çözme becerilerine ait performansı gösteren logit cetvelde yukarıdan aşağıya doğru öğrenci yetenekleri azalmaktadır. İkinci yüzey olarak problem çözme aşamalarının bulunduğu logit cetvelde, en kolay aşamadan en zor aşamaya doğru sıralanmaktadır. Üçüncü yüzey olan puanlayıcı türünde puanlayıcıların katılığı logit cetvelde en cömertten en katıya doğru sıralanmıştır. Dördüncü yüzey olarak belirlenen puanlayıcı yüzeyi yapay (dummy) yüzey olarak tanımlanarak yüzeydeki elemanlar sıfıra sabitlemiştir.

Kalibrasyon haritasına bakıldığında şekil 4’te öğrencilerden en yüksek performansı gösteren 41, en düşük performansı gösteren ise 28 numaralı öğrenciler olduğu gözlenmiştir. Öğrencilerin genel olarak logit cetveldeki yeri pozitif alanda olduğu için grubun genel olarak başarılı olduğu söylenebilir. Dereceli puanlama anahtarındaki aşamaların zorluk/kolaylık düzeylerini gösteren logit cetvelde öğrenciler tarafından en kolay yapılan “Anlama” (problemi anlama) aşaması olurken en zor yapılan “Kontrol etme” (çözümü kontrol etme ve değerlendirme) aşaması olmuştur. “Uygulama” (seçilen çözüm stratejisini uygulama) ile “Çözüm yolu” (çözüm ile ilgili stratejinin seçimi) aşamalarının ise yaklaşık orta düzey zorlukta olduğu görülmüştür. Puanlayıcı türü yüzeyinin bulunduğu logit cetvelde en cömert puanlayıcının “Öz” en katı puanlayıcının “Öğretmen” puanları olduğu görülürken “Akran” puanlarının orta düzey katılıkla olduğu gözlenmiştir.

Measr	Öğrenci	Aşama	Tür	RATIN
3	+	+	+	(4)
	41			
	18 51			
	55			
	50			
2	+	+	+	---
	5 7			
	13 43			
	1 3			
	20 49			
	35			3
1	+	+	+	+
	12 24			
	29 37			
	2 32 36 56	Anlama		
	31 45 48 54		Öz	---
	16 25 34 4 40 52 9			
	15 21 26 8			
	14 17 44			2
*	0 *	+	+	+
	23 27 57	* Uygulama	* Akran	* *
	11 19 30 47 53	Çözüm yolu		
	42 46			---
	10			
	38 6	Kontrol etme	Öğretmen	
-1	+	+	+	1
	39			
	22			
-2	+	+	+	---
	33			
-3	+	+	+	+
	28			
-4	+	+	+	(0)
Measr	Öğrenci	Aşama	Tür	RATIN

Şekil 4. Analizdeki uygulamaya ait kalibrasyon haritası

Alt problem 2: a) İkinci performans görevi uygulamasında öğrencilerin başarı durumuna ait ölçüm raporundaki istatistikler nedir?

İkinci uygulamada öğrencilerin problem çözme becerilerinin başarı durumu ölçüm raporuna ait değerler şekil 5'te yer almaktadır.

Şekil 5'te görüldüğü üzere problem çözme becerilerinin ölçüldüğü ikinci uygulamada toplam 57 öğrencinin logit değerleri 2,72 ile -3,45 arasında bulunmaktadır. Öğrencilerin uygulanan ikinci performans görevine ait başarı ölçümünün standart hata (RMSE) değeri 0,28 olarak bulunmuştur. Ayırma indeksi 3,75 olarak bulunmuştur bu da öğrencilerin ikinci uygulamada da problem çözme becerilerinin birbirinden farklı olduğuna işaret etmektedir. Güvenirliliğe ilişkin kat sayı 0,93 olarak yeterince yüksek bulunmuştur. Öğrenciler yüksek bir güvenilirlikle farklı beceri düzeylerine ayrılmıştır. Sabit etkiye ait “Öğrencilerin problem çözme becerilerini ölçmeye yönelik ikinci performansları arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2=588,9$, $sd=56$, $p=0.00$) yokluk hipotezi reddedilmiştir. Yani, öğrencilerin problem çözme becerilerini ölçmeye yönelik ikinci performansları arasında istatistiksel açıdan anlamlı farklılıklar bulunmaktadır. Ayırma indeksinin kullanıldığı $(4G + 1)/3$ formülüne göre, bireylerin bu göreve ait problem çözme performanslarının yaklaşık beş düzeye ayrılacağı görülmüştür. Bu değeri ki kare değerinin de desteklediği üzere, öğrenci performanslarının puanlanmasında merkeze yönelme etkisi olmadığı görülmüştür. Uyum indekslerine bakıldığında, 57 öğrenci arasından 53 öğrencinin yani, öğrencilerin yaklaşık % 93'ü uyum içi ve uyum dışı kareler ortalamalarının kabul edilebilir aralıkta olduğu gözlenmiştir.

Gözlenen Puan	Gözlem Sayısı	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçerik Uyg. Dışı				Hesap.	
				Ölçüm	Hata	K.O.	ZStd	K.O.	ZStd	Ayırt.	Nu Öğrenci
58	16	3.63	3.82	2.72	.43	.48	-1.0	.43	-1.1	1.27	41 41
72	20	3.60	3.81	2.61	.37	1.44	1.0	1.21	.5	.87	51 51
58	16	3.63	3.80	2.57	.43	1.27	.6	1.92	1.5	.51	18 18
71	20	3.55	3.79	2.53	.35	1.20	.5	1.08	.3	.75	55 55
69	20	3.45	3.74	2.28	.32	.97	.0	1.14	.4	.71	50 50
66	20	3.30	3.58	1.79	.29	1.56	1.4	2.08	2.1	.35	7 7
65	20	3.25	3.57	1.74	.29	.41	-2.0	.50	-1.5	1.35	5 5
60	20	3.00	3.49	1.56	.26	1.57	1.7	2.22	2.5	-1.26	13 13
49	16	3.06	3.48	1.55	.30	1.16	.5	1.81	1.6	.39	43 43
49	16	3.06	3.45	1.48	.30	.88	-.2	.87	-.2	.64	3 3
58	20	2.90	3.44	1.47	.25	1.85	2.4	1.85	2.2	.00	1 1
55	20	2.75	3.36	1.31	.24	.93	-.1	.90	-.2	1.31	49 49
55	20	2.75	3.34	1.29	.24	1.29	1.0	1.35	1.1	.50	20 20
67	24	2.79	3.31	1.23	.22	.80	-.7	.85	-.4	1.14	35 35
41	16	2.56	3.20	1.07	.26	.57	-1.5	.57	-1.5	1.26	12 12
54	20	2.70	3.11	.94	.25	.56	-1.7	.54	-1.5	1.49	24 24
47	20	2.35	3.04	.84	.23	.47	-2.3	.50	-2.1	1.76	37 37
48	20	2.40	3.00	.80	.23	.87	-.4	.84	-.4	1.51	29 29
34	16	2.13	2.97	.76	.25	.44	-2.3	.44	-2.3	2.05	32 32
46	20	2.30	2.92	.70	.24	1.90	2.6	1.84	2.3	.11	56 56
49	20	2.45	2.90	.68	.24	.61	-1.4	.58	-1.5	1.86	2 2
47	20	2.35	2.89	.67	.23	.63	-1.4	.62	-1.4	1.43	36 36
44	20	2.20	2.85	.63	.24	1.78	2.2	1.69	1.8	.36	54 54
60	24	2.50	2.83	.61	.22	1.30	1.1	1.28	1.0	.54	31 31
45	20	2.25	2.83	.61	.23	1.42	1.4	1.35	1.2	.39	48 48
48	20	2.40	2.80	.58	.24	.76	-.7	.75	-.7	1.59	45 45
38	20	1.90	2.73	.50	.23	1.65	2.0	1.65	1.9	.32	34 34
44	20	2.20	2.70	.47	.24	.82	-.5	.77	-.7	1.68	40 40
48	20	2.40	2.69	.46	.25	.74	-.8	.82	-.4	1.16	9 9
47	20	2.35	2.68	.45	.25	.63	-1.2	.57	-1.4	1.78	25 25
45	20	2.25	2.66	.43	.24	.77	-.7	.85	-.3	1.44	52 52
30	16	1.88	2.63	.40	.26	1.07	.3	1.09	.3	.50	16 16
36	20	1.80	2.51	.28	.23	.84	-.5	.86	-.4	1.05	4 4
29	16	1.81	2.50	.28	.26	1.18	.6	1.16	.5	.84	8 8
27	16	1.69	2.49	.26	.26	.70	-.9	.61	-1.2	1.80	21 21
45	20	2.25	2.47	.25	.26	.39	-2.4	.41	-2.0	1.54	26 26
54	24	2.25	2.45	.23	.25	.41	-2.3	.42	-1.9	1.32	15 15
35	20	1.75	2.38	.16	.24	.57	-1.6	.50	-1.9	1.56	14 14
38	20	1.90	2.36	.14	.24	.74	-.8	.79	-.6	.92	17 17
31	16	1.94	2.29	.08	.29	.39	-2.0	.33	-2.1	1.76	44 44
29	16	1.81	2.24	.03	.28	.18	-3.5	.17	-3.4	1.95	23 23
33	20	1.65	2.18	-.02	.24	1.34	1.1	1.26	.8	.69	27 27
33	20	1.65	2.14	-.06	.26	1.64	1.7	1.58	1.5	.58	57 57
29	16	1.81	2.06	-.13	.31	.41	-1.8	.97	.0	1.38	19 19
22	16	1.38	2.06	-.13	.28	.35	-2.4	.28	-2.7	1.74	47 47
40	20	2.00	2.03	-.16	.27	.25	-3.2	.25	-3.0	1.63	30 30
35	20	1.75	2.00	-.18	.28	.80	-.4	1.63	1.4	.70	53 53
35	20	1.75	1.99	-.20	.25	1.62	1.8	1.43	1.2	.19	11 11
38	20	1.90	1.94	-.24	.27	.60	-1.3	.67	-.9	1.31	46 46
26	16	1.63	1.92	-.26	.29	.34	-2.3	.32	-2.3	1.79	42 42
36	20	1.80	1.76	-.41	.28	.29	-2.8	.40	-1.9	1.56	10 10
22	16	1.38	1.66	-.51	.30	1.22	.6	1.32	.8	.58	38 38
34	20	1.70	1.56	-.62	.27	1.41	1.2	1.19	.6	.68	6 6
14	16	.88	1.27	-.97	.34	1.38	.9	1.48	1.1	.44	39 39
16	20	.80	1.07	-1.25	.32	2.82	3.5	2.37	2.9	-.09	22 22
12	20	.60	.61	-2.15	.36	2.65	3.1	2.63	3.2	-.23	33 33
8	24	.33	.23	-3.45	.42	1.68	1.4	.65	-.6	1.10	28 28
42.5	19.2	2.22	2.62	.50	.27	1.00	-.2	1.03	-.1		Ort. (N=57)
14.9	2.2	.73	.78	1.08	.05	.57	1.7	.58	1.6		S.D.

RMSE (Model)= 0,28 Ayırma İndeksi= 3,75 Güvenirlik= 0,93
Tamamı aynı ki-kare değeri= 588,9 s.d. = 56 p=0.00
Rasgele normal ki-kare değeri= 50,8 s.d. = 55 p=0.64

Şekil 5. İkinci uygulamada öğrencilere ait ölçüm raporu

Alt problem 2: b) İkinci performans görevi uygulamasında aşamalara ait ölçüm raporundaki istatistikler nedir?

İkinci performans görevi uygulamasında öğrencilerin problem çözme becerilerinin değerlendirilmesinde kullanılan dereceli puanlama anahtarındaki problem çözme aşamalarının zorluk/kolaylıklarına ilişkin istatistikler Tablo 8’de verilmiştir.

Tablo 8. İkinci Uygulamada Aşamalara Ait Ölçüm Raporu

No Aşamalar	Gözlenen Ort.	Yansız Ort.	Model		<u>Uygunluk İçi</u>		<u>Uygunluk Dışı</u>	
			Ölçüm	Hata	Kareler Ort.	Z Std.	Kareler Ort.	Z Std.
1 Anlama	2,74	3,29	0,69	0,07	1,09	1,1	1,37	3,2
3 Uygulama	2,19	2,69	-0,04	0,07	0,70	-3,9	0,71	-3,4
2 Çözüm yolu	2,13	2,61	-0,12	0,07	1,03	0,3	1,07	0,7
4 Kontrol	1,83	2,18	-0,52	0,07	1,10	1,1	0,99	-0,1
Ortalama	2,22	2,69	0	0,07	0,98	-0,3	1,03	0,1
Std. sapma	0,33	0,39	0,44	0	0,17	2,1	0,24	2,4
RMSE (Model)= 0,07		Std sapma= 0,43			Ayırma İndeksi= 6,13			
Güvenirlilik=0,97								
Tamamı aynı ki-kare değeri= 150,1		s.d.= 3	p= 0,00					
Rasgele normal ki- kare değeri= 2,9		s.d.= 2	p= 0,23					

Tablo 8’de görüldüğü üzere dereceli puanlama anahtarındaki “Anlama”, “Çözüm yolu”, “Uygulama” ve “Kontrol” olmak üzere dört problem çözme aşamasının logit değerleri 0,69 ile -0,52 arasında değişmektedir. Aşamaların zorluk/kolaylık düzeylerini gösteren logit değerlere ait standart hata (RMSE) değeri 0,07 olduğu bulunmuştur. Ayırma indeksi 6,13 olarak hesaplanmıştır. Ayırma indeksinin için yüksek olması istenilen bir durum olmakla beraber, problem çözme becerisinin değerlendirilmesi üzere belirtilen aşamaların birbiri ile aynı kapsamı yoklamadığına işaret ettiği söylenebilir. Güvenirlilik kat sayısının 0,97 olarak bulunması, problem çözme aşamalarının zorluk/kolaylık sıralamasının yüksek bir güvenirlikle yapıldığı anlamına gelir.

Sabit etkiye ait “İkinci uygulamada problem çözme aşamalarının zorluk/kolaylıkları arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2 = 150,1$, $sd=3$, $p=0.00$) yokluk hipotezi reddedilmiştir. Dolayısıyla, ikinci performans görevi uygulamasında problem çözme aşamalarının zorluk/kolaylıkları arasında istatistiksel

açından anlamlı farklılıklar bulunmaktadır. Tablo 8’de aşamalara ait iç uyum ve dış uyum değerleri 0,70 ile 1,37 arasında yer almıştır ve Rasch analizinde etkili bir ölçüm yapımı için verilerin modele çok iyi uyum sağladığı söylenebilir.

Alt problem 2: c) İkinci performans görevi uygulamasında puanlayıcı türüne ait ölçüm raporundaki istatistikler nedir?

İkinci performans görevi uygulamasında öz, akran ve öğretmen puanlayıcıların bulunduğu puanlayıcı türünün puanlama katılığı/cömertliğine ilişkin istatistikler Tablo 9’da verilmiştir.

Tablo 9. İkinci Uygulamada Puanlayıcı Türüne Ait Ölçüm Raporu

N P. Türü	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçi		Uygunluk Dışı		
			Ölçüm	Hata	Kareler Ort.	Z Std.	Kareler Ort.	Z Std.	
1 Öz	3,02	3,23	0,60	0,09	1,39	3,4	1,37	2,6	
2 Akran	2,54	2,72	-0,01	0,09	1,20	1,8	1,27	2,2	
3 Öğretmen	1,87	2,10	-0,59	0,04	0,82	-3,6	0,87	-2,4	
Ortalama	2,48	2,69	0	0,07	1,14	0,5	1,17	0,8	
Std. Sapma	0,47	0,46	0,49	0,02	0,24	3	0,22	2,3	
RMSE (Model)= 0,07		Std sapma= 0,48			Ayırma İndeksi= 6,45				
Güvenirlilik= 0,98									
Tamamı aynı ki-kare değeri= 168,6			s.d.= 2	p= 0.00					
Rasgele normal ki- kare değeri= 2			s.d.= 1	p= 0,16					

Puanlayıcı türü yüzeyini oluşturan “Öz”, “Akran” ve “Öğretmen” olmak üzere üç tür puanlayıcının katılık/cömertliğini gösteren logit değerleri sıra ile 0,60, -0,01 ve -0,59 olarak bulunmuştur. Puanlayıcı türünün puanlama katılık/cömertliklerini gösteren logit değerlere ait standart hata (RMSE) değeri 0,07 olarak bulunmuştur. Ayırma indeksi 6,45 ve puanlayıcı türlerinin puanlama katılık/cömertlik sıralamasının güvenilirliğini gösteren güvenirlilik katsayısı 0,98 olarak yeterince yüksek olduğu söylenebilir.

Sabit etkiye ait “İkinci uygulamada puanlayıcı türlerinin puanlama katılık/cömertlikleri arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2 = 168,6$, $sd=2$, $p=0.00$) yokluk hipotezi reddedilmiştir. Yani, ikinci performans görevi uygulamasında

puanlayıcı türünün puanlama katılık/cömertlikleri arasında manidar farklılık bulunmaktadır. Tablo 9’da iç uyum ve dış uyum değerlerinin 0,82 ile 1,39 arasında yer alması verilerin modele çok iyi uyum sağladığını göstermektedir. Uyum istatistiklerinin iyi olması, puanlayıcıların tutarlı bir puanlama yapmış olduğuna işaret etmektedir.

3.3.Üçüncü Alt Probleme İlişkin Bulgular

Alt problem 3: Üçüncü performans görevi uygulamasında problem çözme becerisinin öz, akran ve öğretmen puanlarının çok yüzeyli Rasch ölçme modeliyle analizi sonucu kalibrasyon haritasının durumu nedir?

Öğrencilerin problem çözme becerilerini ölçmeye yönelik üçüncü performans görevi uygulamasının öz, akran ve öğretmen puanlarının çok yüzeyli Rasch ölçme modeli ile analizinde öğrenci, aşama, puanlayıcı türü ve puanlayıcı olmak üzere kullanılan dört yüzeyin kalibrasyon haritası Şekil 6’da yer almaktadır. Şekil 6’da her bir yüzey için + ve – arasında ölçümlerin bulunduğu logit cetvel bulunmaktadır. Problem çözme becerilerine ait performansı gösteren logit cetvelde yukarıdan aşağıya doğru öğrenci yetenekleri azalmaktadır. İkinci yüzey olarak problem çözme aşamalarının bulunduğu logit cetvelde, en kolay aşamadan en katı aşamaya doğru sıralanmaktadır. Üçüncü yüzey olan puanlayıcı türünde, puanlayıcıların katılığı logit cetvelde en cömertten en katıya doğru sıralanmıştır. Dördüncü yüzey olarak belirlenen puanlayıcı yüzeyi, yapay (dummy) yüzey olarak tanımlanarak yüzeydeki elemanlar sifıra sabitlenmiştir.

Measr	+Öğreci	+Aşama	+Tür	RATIN
3	24			(4)
	45			
	13 25			
2	48 8			
	20 29 51			---
	49			
	34 43 52			
	11 12 19 21 9			
	10 14 23 44			
	39 41 6 7			
1	1 3			3
	22 47			
	28 53		Öz	
	18 50 55	Uygulama	Anlama	
	17 32 5			
	26 33 35 36			---
	15 27 31 4 42 46			
	30			
* 0 *				* 2 *
	2 40 56			
	57	Çözüm yolu	Kontrol etme	---
	54			
	16			
-1	37 38			1

-2				
-3				(0)
Measr	+Öğrenci	+Aşama	+Tür	RATIN

Şekil 6.Üçüncü analizdeki uygulamaya ait kalibrasyon haritası

Şekil 6'daki kalibrasyon haritasına bakıldığında öğrencilerden en yüksek performansı gösteren 24, en düşük performansı gösteren ise 37 ile 38 numaralı öğrenciler olduğu gözlenmiştir. Öğrencilerin genel olarak logit cetveldeki yeri pozitif alanda olduğu için grubun genel olarak başarılı olduğu söylenebilir. Dereceli puanlama anahtarındaki problem çözme aşamalarının zorluk/kolaylık düzeylerini gösteren logit cetvelde "Uygulama" (seçilen çözüm stratejisini uygulama) ile "Anlama" (problemi anlama) öğrenciler tarafından en kolay yapılan aşamalar olurken; "Çözüm yolu" (çözüm ile ilgili stratejinin

seçimi) ile “Kontrol etme” (çözümü kontrol etme ve değerlendirme) aşamaları öğrenciler tarafından en zor yapılan aşamalar olarak bulunmuştur. Puanlayıcı türü yüzeyinin bulunduğu logit cetvelde “Öz” en cömert puanlayıcı olurken “Öğretmen” ile “Akran” puanlayıcıların en katı puanlayıcılar oldukları gözlenmiştir.

Alt problem 3: a) Üçüncü performans görevi uygulamasında öğrencilerin başarı durumuna ait ölçüm raporundaki istatistikler nedir?

Üçüncü uygulamada öğrencilerin problem çözme becerilerinin başarı durumu ölçüm raporuna ait değerler şekil 7’de yer almaktadır.

Şekil 7’de görüldüğü üzere problem çözme becerilerinin ölçüldüğü son uygulamada toplam 57 öğrencinin logit değerleri 2,90 ile -0,99 arasında değer aldığı görülmüştür. Öğrencilerin uygulanan üçüncü performans görevine ait başarı ölçümünün standart hata (RMSE) değeri 0,33 olarak bulunmuştur. Ayırma indeksi 2,33 olarak bulunmuş öğrencilerin üçüncü uygulamada da problem çözme becerilerinin birbirinden farklı olduğuna işaret etmektedir. Güvenirliğe ilişkin kat sayı 0,84 olarak bulunmuştur.

Sabit etkiye ait “Öğrencilerin problem çözme becerilerini ölçmeye yönelik üçüncü performansları arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2=357,4$, $sd=56$, $p=0.00$) yokluk hipotezi reddedilmiştir. Yani, öğrencilerin problem çözme becerilerini ölçmeye yönelik üçüncü performansları arasında istatistiksel açıdan anlamlı farklılıklar bulunmaktadır. Öğrencilerin problem çözme performanslarının ayırma indeksinin kullanıldığı formüle göre yaklaşık üç düzeye ayrılabilceği görülmüştür. Ki kare değerinin de bu sonucu desteklediği üzere yapılan puanlama da merkeze yönelme etkisinin olmadığı görülmüştür. Uyum indekslerine bakıldığında ise 57 öğrenci arasından 56 öğrencinin yani, öğrencilerin yaklaşık % 98’i uyum içi ve uyum dışı kareler ortalamalarının kabul edilebilir aralıkta olduğu gözlenmiştir.

Gözlenen Puan	Gözlem Sayısı	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçi Uyg. Dışı				Hes. Ayırt. Nu Öğrenci	
				Ölçüm	Hata	K.O.	ZStd	K.O.	ZStd		
61	16	3.81	3.86	2.90	.60	.67	-.3	.50	-.4	1.13	24 24
58	16	3.63	3.75	2.32	.44	.60	-.7	.41	-.8	1.32	45 45
43	12	3.58	3.74	2.28	.48	1.39	.8	1.81	1.3	.58	13 13
42	12	3.50	3.72	2.21	.44	.79	-.2	.76	-.3	1.12	25 25
42	12	3.50	3.65	1.94	.45	.51	-.9	.62	-.4	1.27	8 8
57	16	3.56	3.63	1.89	.41	.88	.0	.55	-.7	1.30	48 48
41	12	3.42	3.60	1.80	.42	1.07	.2	.78	-.2	1.13	51 51
55	16	3.44	3.59	1.79	.37	.60	-.9	.90	.0	1.19	20 20
55	16	3.44	3.58	1.77	.37	1.38	.9	1.34	.7	.85	29 29
52	16	3.25	3.54	1.66	.32	1.71	1.6	1.74	1.5	.53	49 49
39	12	3.25	3.49	1.53	.38	.42	-1.4	.61	-.7	1.26	43 43
38	12	3.17	3.47	1.49	.36	1.38	.9	1.27	.7	.63	52 52
52	16	3.25	3.45	1.46	.33	.90	-.1	1.17	.4	.83	34 34
53	16	3.31	3.44	1.42	.34	1.16	.5	1.05	.2	.87	19 19
52	16	3.25	3.42	1.40	.33	1.17	.5	1.55	1.1	.73	9 9
39	12	3.25	3.41	1.38	.38	1.66	1.3	1.70	1.1	.36	21 21
38	12	3.17	3.40	1.36	.36	.31	-2.0	.32	-1.7	1.56	11 11
51	16	3.19	3.39	1.32	.32	.82	-.3	1.26	.7	1.01	12 12
48	16	3.00	3.37	1.30	.29	.94	.0	.89	-.2	.86	44 44
51	16	3.19	3.35	1.26	.32	.76	-.5	.81	-.2	1.17	10 10
52	16	3.25	3.34	1.25	.33	.86	-.2	.87	-.1	1.01	23 23
51	16	3.19	3.33	1.22	.32	.64	-.9	.67	-.5	1.23	14 14
38	12	3.17	3.29	1.16	.38	1.56	1.1	1.51	1.0	.12	39 39
38	12	3.17	3.29	1.16	.38	1.03	.2	1.23	.6	.89	41 41
39	16	2.44	3.25	1.08	.26	.99	.0	.97	.0	1.12	6 6
51	16	3.19	3.24	1.07	.33	.77	-.4	1.16	.4	1.01	7 7
47	16	2.94	3.21	1.03	.30	.80	-.4	.78	-.5	1.35	1 1
52	16	3.25	3.20	1.02	.35	.62	-.8	.62	-.6	1.26	3 3
47	16	2.94	3.15	.94	.29	.49	-1.6	.57	-1.0	1.32	22 22
39	16	2.44	3.09	.85	.26	.81	-.5	.81	-.5	1.07	47 47
40	16	2.50	3.05	.80	.27	.51	-1.6	.47	-1.7	1.57	28 28
48	16	3.00	3.04	.79	.30	.99	.0	.89	-.1	.98	53 53
34	12	2.83	2.92	.63	.35	1.90	1.7	1.82	1.6	-.12	18 18
45	16	2.81	2.88	.58	.29	.80	-.4	.87	-.2	1.11	55 55
47	16	2.94	2.87	.58	.31	.98	.0	.90	-.1	1.12	50 50
37	16	2.31	2.83	.53	.25	1.56	1.6	1.50	1.4	.49	5 5
29	12	2.42	2.82	.52	.30	.90	-.1	.85	-.2	.96	32 32
40	16	2.50	2.81	.50	.27	1.13	.4	.99	.1	1.22	17 17
30	12	2.50	2.75	.44	.31	.96	.0	.91	.0	1.18	33 33
41	16	2.56	2.74	.43	.27	.97	.0	.96	.0	.86	36 36
28	12	2.33	2.70	.39	.30	1.54	1.3	1.51	1.3	.60	26 26
37	16	2.31	2.70	.39	.25	.65	-1.1	.63	-1.2	1.51	35 35
41	16	2.56	2.62	.30	.27	.81	-.4	.89	-.1	1.17	15 15
30	16	1.88	2.62	.30	.29	.62	-1.1	.59	-1.0	1.53	4 4
38	16	2.38	2.61	.29	.26	1.38	1.1	1.44	1.2	.68	27 27
34	16	2.13	2.57	.25	.26	.57	-1.5	.59	-1.3	1.51	42 42
33	16	2.06	2.52	.20	.25	.59	-1.4	.58	-1.4	1.42	46 46
31	16	1.94	2.51	.19	.25	.84	-.4	.84	-.4	1.12	31 31
30	12	2.50	2.47	.15	.35	.92	.0	.92	.0	.80	30 30
34	16	2.13	2.07	-.22	.29	.84	-.3	.91	-.1	.80	40 40
22	16	1.38	2.05	-.24	.27	1.42	1.2	1.47	1.2	.89	2 2
32	16	2.00	2.00	-.29	.26	.92	-.1	.93	.0	1.16	56 56
23	16	1.44	1.77	-.49	.29	1.12	.4	.96	.0	1.02	57 57
29	16	1.81	1.63	-.61	.29	.93	.0	.86	-.2	.66	54 54
21	16	1.31	1.49	-.74	.27	2.61	3.5	2.72	3.5	-.95	16 16
22	16	1.38	1.27	-.95	.27	1.56	1.5	1.40	1.1	.78	37 37
17	16	1.06	1.23	-.99	.28	1.32	.9	1.15	.5	1.00	38 38
40.6	14.9	2.74	2.96	.86	.32	1.00	.0	1.01	.1		Ort. (N=57)
10.3	1.8	.67	.64	.84	.07	.42	1.0	.43	1.0		S.D.

RMSE (Model)= 0,33

Std Sapma= 0,77

Ayırtma İndeksi= 2,33

Güvenirlilik= 0,84

Tamamı aynı ki-kare değeri= 357,4 s.d.= 56

p= 0.00

Rasgele aynı ki-kare değeri= 47,9 s.d.= 55

p= 0,74

Şekil 7. Üçüncü uygulamada öğrencilere ait ölçüm raporu

Alt problem 3: b) Üçüncü performans görevi uygulamasında aşamalara ait ölçüm raporundaki istatistikler nedir?

Üçüncü performans görevi uygulamasında öğrencilerin problem çözmelerinin değerlendirilmesinde kullanılan puanlama anahtarındaki problem çözme aşamalarının zorluk/kolaylık düzeylerine ilişkin istatistikler tablo 10’da verilmiştir.

Tablo 10. Üçüncü Uygulamada Aşamalara Ait Ölçüm Raporu

No Aşamalar	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçi		Uygunluk Dışı	
			Ölçüm	Hata	Kareler Ort.	Z Std.	Kareler Ort.	Z Std.
1 Anlama	3,06	3,39	0,48	0,09	1,06	0,6	1,02	0,1
3 Uygulama	3,04	3,38	0,46	0,09	1,22	1,9	1,22	1,6
2 Çözüm yolu	2,38	2,71	-0,46	0,08	0,75	-2,8	0,76	-2,3
4 Kontrol	2,38	2,70	-0,47	0,08	1,07	0,7	1,01	0,1
Ortalama	2,72	3,05	0	0,08	1,02	0,1	1	-0,1
Std. sapma	0,33	0,34	0,47	0,01	0,17	1,8	0,16	1,4
RMSE (Model)= 0,08		Std. sapma= 0,46			Ayrırma İndeksi= 5,59			
Güvenirlilik=0,97								
Tamamı aynı ki-kare değeri= 128,9			s.d.= 3	p= 0.00				
Rasgele normal ki- kare değeri= 2,9			s.d.= 2	p= 0,23				

Tablo 10’da görüldüğü üzere dereceli puanlama anahtarındaki “Anlama”, “Çözüm yolu”, “Uygulama” ve “Kontrol etme” olmak üzere dört problem çözme aşamasının logit değerleri 0,48 ile -0,47 arasında değişmektedir. Aşamaların zorluk/kolaylık düzeylerini gösteren logit değerlere ait standart hata (RMSE) değerinin 0,08 olduğu bulunmuştur. Ayrırma indeksi 5,59 ve aşamaların puanlanma zorluk/kolaylık düzeylerinin sıralamasının güvenirlilik kat sayısı 0,97 olarak bulunmuştur. Elde edilen güvenirlilik kat sayısı değerinin yeterince yüksek olduğu söylenebilir.

Sabit etkiye ait “Üçüncü uygulamada aşamaların puanlanma katılık/cömertlikleri arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2 = 128,1$, $sd=3$, $p=0.00$) yokluk hipotezi reddedilmiştir. Dolayısıyla, üçüncü uygulamada problem çözme aşamalarının zorluk/kolaylık düzeyleri arasında istatistiksel açıdan anlamlı farklılıklar bulunmaktadır. Rasch analizinde verimli bir ölçüm yapımı için iç uyum ve dış uyum

değerlerinin tablo 10'dan bakılarak iç uyum ve dış uyum değerleri 0,75 ile 1,22 arasında yer aldığı ve verilerin model uyum gösterdiği söylenebilir.

Alt problem 3: c) Üçüncü performans görevi uygulamasında puanlayıcı türüne ait ölçüm raporundaki istatistikler nedir?

Son performans görevi uygulamasında öz, akran ve öğretmen puanlayıcıların bulunduğu puanlayıcı türünün puanlama katılık/cömertliğine ilişkin istatistikler tablo 11'de verilmiştir.

Tablo 11. Üçüncü Uygulamada Puanlayıcı Türüne Ait Ölçüm Raporu

N P. Türü	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçi		Uygunluk Dışı	
			Ölçüm	Hata	Kareler Ort.	Z Std.	Kareler Ort.	Z Std.
1 Öz	3,35	3,52	0,74	0,10	1,18	1,4	0,94	-0,3
2 Akran	2,68	2,84	-0,32	0,09	1,12	1	1,14	1
3 Öğretmen	2,41	2,75	-0,42	0,05	0,94	-0,9	0,99	-0,1
Ortalama	2,82	3,03	0	0,08	1,08	0,5	1,02	0,2
Std. Sapma	0,40	0,34	0,52	0,02	0,10	1,1	0,08	0,6
RMSE (Model)= 0,09		Std sapma= 0,52			Ayrırma İndeksi= 6,08			
Güvenirlilik= 0,97								
Tamamı aynı ki-kare değeri= 105,7			s.d.= 2	p= 0.00				
Rasgele normal ki- kare değeri= 2			s.d.= 1	p= 0,16				

Puanlayıcı türü yüzeyini oluşturan “Öz”, “Akran” ve “Öğretmen” olmak üzere üç tür puanlayıcının katılık/cömertliğini gösteren logit değerleri sırayla 0,74, -0,32 ve -0,42 olarak bulunmuştur. Puanlayıcı türünün puanlama katılık/cömertliklerini gösteren logit değerlere ait standart hata (RMSE) değeri 0,09 olarak bulunmuştur. Ayrırma indeksi 6,08 ve puanlayıcı türlerinin puanlama katılık/cömertlik sıralamasının güvenilirliğini gösteren güvenirlilik katsayısı 0,97 olarak yeterince yüksek olduğu söylenebilir.

Sabit etkiye ait “Üçüncü uygulamada puanlayıcı türlerinin puanlama katılık/cömertlikleri arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2 = 105,7$, $sd=2$, $p=0.00$) yokluk hipotezi reddedilmiştir. Dolayısıyla son performans görevi uygulamasında puanlayıcı türünün puanlama katılık/cömertlikleri arasında manidar farklılık

bulunmaktadır. Rasch analizinde verimli bir ölçüm yapımı için iç uyum ve dış uyum değerlerinin 0,5 ile 1,5 arasında olması beklenir (Linacre, 2014). Tablo 11’de görüldüğü üzere iç uyum ve dış uyum değerlerinin 0,94 ile 1,18 arasında yer alması verilerin modele çok iyi uyum sağladığını göstermektedir.

3.4.Dördüncü Alt Probleme İlişkin Bulgular

Alt problem 4: Birinci, ikinci ve üçüncü performans görevi uygulamalarının tümünü içeren süreçte problem çözme becerisinin öz, akran ve öğretmen puanlarının çok yüzeyli Rasch ölçme modeliyle analizi sonucu kalibrasyon haritasının durumu nedir?

Öğrencilerin problem çözme becerilerini ölçmeye yönelik birinci, ikinci ve üçüncü performans görevi uygulamasının öz, akran ve öğretmen puanlarının çok yüzeyli Rasch ölçme modeli ile analizinde oluşan kalibrasyon haritası Şekil 8’de yer almaktadır. Kalibrasyon haritasında öğrenci, aşama, puanlayıcı türü, görev ve puanlayıcı olmak üzere beş yüzeye ait her bir yüzey için + ve – arasında ölçümlerin bulunduğu logit cetvel bulunmaktadır. Bireyin problem çözme becerilerine ait performansını gösteren logit cetvelde yukarıdan aşağıya doğru öğrenci yetenekleri azalmaktadır. Problem çözme aşamalarını gösteren ikinci yüzeyde aşamalar yukarıdan aşağıya en kolay yapılan aşamadan en zor yapılan aşamaya doğru sıralanmaktadır. Üçüncü yüzey olan puanlayıcı türünde, puanlayıcıların katılığı logit cetvelde yukarıdan aşağı en cömertten en katıya doğru sıralanmıştır. Puanlayıcı türüne ait yüzeyde öz, akran ve öğretmen puanlayıcıların değerlendirme sürecindeki farklılaşmalarını her uygulamada daha net görebilmek için puanlayıcılar zaman sırasına göre kodlanmıştır. Birinci, ikinci ve üçüncü uygulamadaki öğrencilerin kendini puanlaması analizde sıra ile “Öz 1”, “Öz 2” ve “Öz 3” olarak kodlanmıştır. Akran ve öğretmen puanlamaları da yine haftalara göre “Akran 1”, “Akran 2”, “Akran 3” ve “Öğretmen 1”, “Öğretmen 2”, “Öğretmen 3” olarak numaralandırılmıştır. Dördüncü yüzey olan görev, yukarıdan aşağıya en kolay görevden en zor göreve doğru sıralanmaktadır. Performans görevleri de puanlayıcı türünde olduğu gibi uygulandığı haftaya göre “Görev 1”, “Görev 2” ve “Görev 3” şeklinde numaralandırılmıştır. Son olarak, beşinci yüzey olarak belirlenen puanlayıcı yüzeyi de yapay (dummy) yüzey olarak tanımlanmış içindeki elemanlar sıfıra eşitlenmiştir.

Measr	+Öğrenci	+Aşama	+Tür	+Görev	RATIN
2	+	+	+	+	(4)
7					
24					
50	51				---
18					
13	20 55				
45	48 49				
1	+29 41 5 52 9	+	+	+	+
12					
32	44				3
34	36				
19	39 8				
1	15 17 3 31 35				
11	23 26 27 30 40		Öz 2 Öz 3		---
16	37 4 42 53 56	Anlama	Öz 1		
10	14	Uygulama			
43	47 54		Akran 1	Görev1 Görev3	2
* 0	*46	*	*Akran 2	*	*
25	6	Çözüm yolu			
21	22 33 57				
2		Kontrol etme	Öğret.1 Öğret.2 Öğret.3 Akran3	Görev 2	---
28	38				
					1
-1	+	+	+	+	(0)
Measr	+Öğrenci	+Aşama	+Tür	+Görev	RATIN

Şekil 8. Üç uygulamanın verilerini içeren kalibrasyon haritası

Şekil 8’deki kalibrasyon haritasına bakıldığında öğrencilerden en yüksek performansı gösteren 7 ile 24, en düşük performansı gösteren ise 28 ile 38 numaralı öğrencilerin olduğu görülmektedir. Üç uygulamanın verilerini içeren analizde öğrencilerin genel olarak logit cetveldeki yeri pozitif alanda olduğu için grubun genel olarak başarılı olduğu gözükmemektedir. Dereceli puanlama anahtarındaki aşamaların zorluk/kolaylık düzeylerini gösteren logit cetvelde “Anlama” (problemi anlama) ile “Uygulama” (seçilen çözüm stratejisini uygulama) kolay yapılan aşamalar olarak belirlenirken “Çözüm yolu” (çözüm ile ilgili stratejinin seçimi) ile “Kontrol etme” (çözümü kontrol etme ve değerlendirme) aşamalarının ise zor yapılan aşamalar olarak gözlenmiştir. Puanlayıcı türü yüzeyinin bulunduğu logit cetvel haftalara göre incelendiğinde; “Öz 1”, “Öz 2” ve “Öz 3” cetvelde pozitif alanda olduğu için öğrencilerin kendi çalışmalarını puanlamaları genel olarak cömert oldukları belirlenmiştir. Akran puanlamaları incelendiğinde ise haftalara göre bir farklılaşma olduğu görülmektedir. Birinci uygulamadaki akran puanlamalarını temsil eden “Akran 1” logit cetvelde sıfırın üzerinde kalarak cömert puanlayıcı olurken, ikinci haftaki akran puanlarını temsil eden “Akran 2” orta düzeyde katılık/cömertlik göstermiştir. “Akran

3” ise logit cetvelde negatif alanda yer alarak son uygulamadaki akran puanlamalarının birinci ve ikinci haftaya göre puanlamada daha katı olduğu belirlenmiştir. Şekil 8’deki kalibrasyon haritasına bakıldığında genel olarak öğretmen puanlamalarının da üç uygulamada logit cetvelde negatif alanda yer aldıkları gözlenerek en katı puanlayıcılar oldukları belirlenmiştir.

Alt problem 4: a) Üç görevin bulunduğu analizde öğrencilerin başarı durumuna ait ölçüm raporundaki istatistikler nedir?

Öğrencilerin problem çözme becerilerini ölçmeye yönelik uygulanan üç performans görevlerine ait puanlamaları kapsayan genel analizde başarı durumuna ait değerler şekil 9’da yer almaktadır.

Şekil 9’da görüldüğü üzere problem çözme becerilerinin ölçüldüğü üç uygulamanın verilerinin yer aldığı genel analizde, toplam 57 öğrencinin logit değerleri 1,56 ile -0,50 arasında değer almaktadır. Öğrencilerin üç haftalık uygulanan performans görevine ait başarı ölçümünün standart hata (RMSE) değeri 0,14 olarak bulunmuştur. Ayırma indeksi 3,58 ve güvenilirliğe ilişkin kat sayısı 0,93 olarak yeterince yüksek bulunmuştur. Sabit etkiye ait “Öğrencilerin problem çözme becerilerini ölçmeye yönelik üç haftalık süreçte öğrencilerin performansları arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2 = 723,5$, $sd=56$, $p=0.00$) yokluk hipotezi reddedilmiştir. Yani, öğrencilerin problem çözme becerilerini ölçmeye yönelik üç haftalık süreçte öğrencilerin performansları arasında istatistiksel olarak anlamlı farklılıklar bulunmaktadır. Bireylerin ayırma indeksinin (G) kullanıldığı $(4G + 1)/3$ yüzeydeki elemanların kaç farklı düzeye ayrıldığını gösteren formül sonucu yaklaşık beş düzeye ayrılabilceği görülmüştür. Ki kare değerinin de bu durumu destekleyen sonuç vermesi ile yapılan puanlamada merkeze yönelme etkisi bulunmadığı belirlenmiştir. Uyum değerleri ele alınacak olursa; 57 öğrenci arasından 55 öğrencinin yani öğrencilerin yaklaşık % 96’sı uyum içi ve uyum dışı kareler ortalamalarının kabul edilebilir aralıkta olduğu gözlenmiştir. Analiz sonucu uygunluk içi ve uygunluk dışı değer aralıklarının dışında kalan iki birey için ise uygulamalarda gösterdiği performansın beklenen seviyede olmadığı söylenebilir.

Gözlenen Puan	Gözlem Sayısı	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçerik Uyg. Dışı Hesap.					
				Ölçüm	Hata	K.O.	ZStd	K.O.	ZStd	Ayrırt.	Nu Öğrenci
193	56	3.45	3.69	1.56	.17	1.00	.0	1.00	.0	.95	7 7
192	56	3.43	3.68	1.52	.17	.80	-.7	.63	-1.4	1.21	24 24
178	52	3.42	3.65	1.43	.17	1.03	.2	.96	.0	.95	51 51
174	52	3.35	3.64	1.41	.16	1.13	.5	1.10	.4	.91	50 50
162	48	3.38	3.58	1.28	.18	1.57	2.0	1.54	1.7	.65	18 18
151	48	3.15	3.55	1.22	.16	1.11	.5	1.19	.7	.59	13 13
183	56	3.27	3.55	1.21	.15	.71	-1.4	.67	-1.4	1.15	20 20
116	36	3.22	3.53	1.18	.18	1.43	1.5	1.44	1.2	.57	55 55
176	56	3.14	3.48	1.09	.14	1.09	.5	1.17	.7	1.07	48 48
177	56	3.16	3.48	1.09	.14	.71	-1.5	.59	-1.9	1.39	45 45
172	56	3.07	3.47	1.07	.14	.79	-1.1	.84	-.7	1.18	49 49
143	48	2.98	3.45	1.04	.14	1.05	.3	1.03	.2	1.12	52 52
150	48	3.13	3.44	1.03	.15	1.20	.9	1.02	.1	1.04	41 41
174	56	3.11	3.43	1.01	.14	.99	.0	.96	-.1	1.10	29 29
172	56	3.07	3.40	.97	.14	.75	-1.3	.72	-1.3	1.22	9 9
163	56	2.91	3.40	.97	.13	.99	.0	.95	-.1	.87	5 5
143	48	2.98	3.39	.94	.14	.60	-2.3	.66	-1.6	1.09	12 12
150	52	2.88	3.28	.81	.13	.69	-1.8	.65	-1.8	1.35	44 44
134	48	2.79	3.28	.81	.13	.72	-1.6	.72	-1.5	1.30	32 32
157	56	2.80	3.18	.69	.13	.73	-1.6	.75	-1.2	1.23	36 36
153	56	2.73	3.15	.66	.13	1.04	.3	1.07	.4	.98	34 34
132	48	2.75	3.14	.65	.14	.90	-.4	.98	.0	1.07	8 8
123	48	2.56	3.05	.56	.13	1.34	1.8	1.30	1.5	.83	39 39
136	52	2.62	3.04	.55	.13	.55	-3.1	.57	-2.6	1.55	19 19
131	48	2.73	3.03	.54	.14	.91	-.4	1.03	.1	.86	3 3
140	56	2.50	2.99	.51	.12	.90	-.5	.85	-.7	1.14	17 17
148	56	2.64	2.99	.50	.13	1.27	1.5	1.26	1.2	.69	1 1
140	56	2.50	2.96	.48	.12	.87	-.7	.87	-.7	.82	35 35
132	56	2.36	2.96	.47	.12	1.00	.0	.99	.0	1.09	31 31
156	60	2.60	2.94	.45	.12	.60	-2.8	.58	-2.4	1.37	15 15
131	56	2.34	2.93	.45	.12	1.02	.1	1.04	.2	1.11	27 27
131	52	2.52	2.92	.44	.13	.69	-2.0	.67	-1.9	1.45	23 23
122	48	2.54	2.91	.43	.14	.70	-1.8	.74	-1.0	1.18	30 30
116	48	2.42	2.90	.42	.13	.89	-.5	1.16	.8	.97	11 11
118	48	2.46	2.88	.41	.13	.84	-.8	.90	-.4	1.22	26 26
140	56	2.50	2.84	.37	.12	.82	-1.0	.76	-1.2	1.47	40 40
83	36	2.31	2.80	.34	.15	.49	-3.0	.49	-2.6	1.60	53 53
135	56	2.41	2.77	.32	.12	1.27	1.6	1.25	1.3	.66	37 37
119	52	2.29	2.76	.31	.12	.60	-2.8	.58	-2.7	1.61	42 42
124	56	2.21	2.75	.30	.12	.73	-1.8	.76	-1.5	1.26	4 4
78	36	2.17	2.73	.29	.15	1.17	.8	1.19	.9	.63	56 56
115	52	2.21	2.70	.26	.12	1.52	2.8	1.59	3.0	.03	16 16
115	52	2.21	2.68	.25	.12	.52	-3.5	.50	-3.4	1.82	14 14
116	52	2.23	2.55	.15	.13	.54	-3.1	.55	-2.8	1.53	10 10
94	48	1.96	2.48	.10	.13	2.18	5.2	2.32	5.5	-1.16	43 43
102	52	1.96	2.47	.09	.12	.59	-2.9	.56	-3.0	1.62	47 47
73	36	2.03	2.42	.06	.16	1.37	1.5	1.34	1.2	.67	54 54
119	56	2.13	2.35	.02	.13	.56	-2.9	.57	-2.5	1.48	46 46
89	52	1.71	2.22	-.07	.13	1.99	4.5	1.90	4.0	-.12	25 25
106	56	1.89	2.16	-.12	.12	1.07	.4	1.06	.3	.93	6 6
102	56	1.82	2.08	-.17	.12	1.16	.9	1.12	.7	.80	22 22
56	36	1.56	2.05	-.19	.16	1.16	.7	1.05	.2	.94	57 57
87	52	1.67	2.01	-.22	.13	1.37	1.9	1.32	1.6	.70	33 33
66	44	1.50	1.98	-.24	.14	2.06	4.3	1.98	3.8	-.23	21 21
71	56	1.27	1.89	-.30	.13	1.91	4.0	1.98	4.0	.11	2 2
70	56	1.25	1.67	-.45	.13	1.39	1.9	1.34	1.6	.79	28 28
68	52	1.31	1.61	-.50	.14	.82	-.9	.89	-.5	1.02	38 38
129.8	51.4	2.52	2.92	.53	.14	1.02	-.1	1.01	-.1		Ort. (No=57)
34.6	5.9	.58	.54	.51	.02	.39	2.0	.40	1.9		S.D.

RMSE (Model)= 0,14 Std. Sapma= 0,50 Ayrırma İndeksi= 3,58 Güvenirlilik= 0,93
Tamamı aynı ki-kare= 723,5 s.d.= 56 p= 0,00
Rasgele normal ki-kare: 51,9 s.d.= 55 p= 0,60

Şekil 9. Üç uygulamanın verilerini içeren analizde öğrencilerin ölçüm raporu

Alt problem 4: b) Üç görevin bulunduğu analizde aşamalara ait ölçüm raporundaki istatistikler nedir?

Üç haftalık performans görevi uygulamalarında öğrencilerin problem çözmelerinin değerlendirilmesinde kullanılan puanlama anahtarındaki problem çözme aşamalarının zorluk/kolaylık düzeylerine ilişkin istatistikler tablo 12’de verilmiştir.

Tablo 12. Üç Uygulamanın Verilerini İçeren Analizde Aşamalara Ait Ölçüm Raporu

No Aşamalar	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçi		Uygunluk Dışı	
			Ölçüm	Hata	Kareler Ort.	Z Std.	Kareler Ort.	Z Std.
1 Anlama	2,81	3,28	0,27	0,04	0,85	-3,2	0,84	-2,6
3 Uygulama	2,71	3,19	0,16	0,04	1,11	2,1	1,09	1,5
2 Çözüm yolu	2,37	2,85	-0,15	0,03	0,96	-0,9	0,97	-0,4
4 Kontrol	2,22	2,68	-0,28	0,03	1,11	2,2	1,09	1,7
Ortalama	2,72	3	0	0,04	1,01	0,1	1	0
Std. sapma	0,33	0,24	0,22	0	0,11	2,3	0,10	1,8
RMSE (Model)= 0,04		Std sapma= 0,22			Ayrırma İndeksi= 6,16			
Güvenirlilik=0,97								
Tamamı aynı ki-kare değeri= 155,9			s.d.= 3	p= 0.00				
Rasgele normal ki- kare değeri= 2,9			s.d.= 2	p= 0,23				

Tablo 12’de görüldüğü üzere dereceli puanlama anahtarındaki “Anlama”, “Çözüm yolu”, “Uygulama” ve “Kontrol etme” olmak üzere dört aşamanın logit değerleri 0,27 ile -0,28 arasında değişmektedir. Problem çözme aşamalarının zorluk/kolaylık düzeylerini gösteren logit değerlere ait standart hata (RMSE) değeri 0,04 olarak belirlenmiştir. Ayrırma indeksi 6,16 ve problem çözme aşamalarının zorluk/kolaylıklarının sıralamasının güvenirlilik kat sayısı 0,97 olarak bulunmuştur. Güvenirliliğe ilişkin kat sayı değerinin yeterince yüksek olduğu söylenebilir. Sabit etkiye ait “Öğrencilerin problem çözme becerilerini ölçmeye yönelik üç haftalık uygulamada aşamaların zorluk/kolaylıkları arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2 = 155,9$, $sd=3$, $p=0.00$) yokluk hipotezi reddedilmiştir. Yani, üç haftalık uygulamada problem çözme aşamalarının zorluk/kolaylık düzeylerini arasında istatistiksel açıdan anlamlı farklılıklar bulunmaktadır. Uyum indeksleri için de tablo 12’de iç uyum ve dış uyum değerleri 0,84 ile 1,11 arasında yer almıştır ve verilerin model uyumunun sağladığı söylenebilir.

Alt problem 4: c) Üç görevin bulunduğu analizde puanlayıcı türüne ait ölçüm raporundaki istatistikler nedir?

Üç haftalık performans görevi uygulamasının verilerini içeren analizdeki; öz, akran ve öğretmen puanlayıcılar bulunduğu puanlayıcı türünün puanlama katılık/cömertliğine ilişkin istatistikler haftalara göre tablo 13’te verilmiştir.

Tablo 13. Üç Uygulamanın Verilerini İçeren Analizde Puanlayıcı Türüne Ait Ölçüm Raporu

N	P. Türü	Gözlenen Ort.	Yansız Ort.	Model		Uygunluk İçi		Uygunluk Dışı	
				Ölçüm	Hata	Kareler Ort.	Z Std.	Kareler Ort.	Z Std.
4	Öz	3,02	3,37	0,39	0,07	1,39	3,6	1,33	2,7
7	Öz	3,35	3,36	0,38	0,08	1,32	2,4	1,05	0,4
1	Öz	3,27	3,30	0,29	0,08	1,31	2,4	1,13	0,9
2	Akran	3,11	3,15	0,12	0,09	0,94	-0,4	0,83	-1
5	Akran	2,54	3,05	0	0,07	1,07	0,7	1,11	1
8	Akran	2,68	2,71	-0,26	0,08	1,04	0,4	1	0
9	Öğretmen	2,41	2,66	-0,30	0,04	1,07	1,2	1,14	2,1
6	Öğretmen	1,87	2,65	-0,31	0,03	0,86	-3,2	0,87	-2,7
3	Öğretmen	2,41	2,65	-0,31	0,04	0,89	-2,3	0,87	-2,5
Ortalama		2,74	2,99	0	0,07	1,10	0,6	1,04	0,1
Std. Sapma		0,46	0,30	0,29	0,02	0,19	2,1	0,15	1,8
RMSE (Model)= 0,07		Std sapma= 0,28		Ayrırma İndeksi= 4,07					
Güvenirlilik= 0,94									
Tamamı aynı ki-kare değeri= 196,8		s.d.= 8		p= 0,00					
Rasgele normal ki- kare değeri= 7,6		s.d.= 7		p= 0,36					

Puanlayıcı türü yüzeyini oluşturan “Öz”, “Akran” ve “Öğretmen” olmak üzere üç tür puanlayıcının katılık/cömertliğini gösteren istatistikler ve uyum indeksleri Tablo 13’te yer almaktadır. Öğrencilerin kendi performanslarını puanladığı öz puanlar incelendiğinde; logit değerleri haftalara göre sırayla 0,29, 0,39 ve 0,38 olarak bulunmuştur. Öğrencilerin arkadaşlarının performanslarını puanladığı akran puanları incelendiğinde; logit değerleri haftalara göre sırayla 0,12, 0 ve -0,26 olarak bulunmuştur. Öğretmen puanlamaları incelendiğinde; logit değerleri haftalara göre sırayla -0,30, -0,31 ve -0,31 olduğu görülmüştür. Puanlayıcı türünün puanlama katılık/cömertliklerini gösteren logit değerlere

ait standart hata (RMSE) değeri 0,07 olarak bulunmuştur. Ayırma indeksi 4,07 ve puanlayıcı türlerinin puanlama katılık/cömertlik sıralamasının güvenilirliğini gösteren güvenilirlik katsayısı 0,94 olarak yeterince yüksek olduğu bulunmuştur. Ayırma indeksi kullanılarak hesaplanan $(4G+1)/3$ formülünde görev yüzeyi için ayırma oranı yaklaşık dokuz çıkmıştır. Sabit etkiye ait “Üç haftalık performans görevi uygulamasının verilerini içeren analizdeki puanlayıcı türlerinin puanlama katılık/cömertlikleri arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2 = 196,8$, $sd=8$, $p=0.00$) yokluk hipotezi reddedilmiştir. Dolayısıyla üç haftalık performans görevi uygulamasının verilerini içeren analizdeki puanlayıcı türlerinin puanlama katılık/cömertlikleri arasında manidar farklılık bulunmaktadır. Ki-kare değerinin anlamlı çıkması, ayırma indeksinin, ayırma oranının ve güvenilirliğin yüksek olması görevlerin puanlanmasında halo etkisinin olmadığına işaretler. Tablo 13’te görüldüğü üzere iç uyum ve dış uyum değerlerinin 0,83 ile 1,39 arasında yer alması verilerin modele çok iyi uyum sağladığına işaret etmektedir. Uyum istatistiklerinin iyi olması da puanlayıcıların tutarlı puanlama yapmış olduğu anlamına gelmektedir.

Alt problem 4: d) Üç görevin bulunduğu analizde görevlere ait ölçüm raporundaki istatistikler nedir?

Üç performans görevi uygulamasının verilerini içeren analizde görevlerin güçlük düzeylerine ilişkin istatistikler ve uyum indeksleri Tablo 14’te verilmiştir.

Tablo 14. Üç Uygulamannın Verilerini İçeren Analizde Görevlere Ait Ölçüm Raporu

N	P. Türü	Gözlenen		Yansız		Model		Uygunluk İçİ		Uygunluk Dışı	
		Ort.	Ort.	Ölçüm	Hata	Kareler	Z	Kareler	Z		
3	Görev 3	2,72	3,17	0,15	0,03	1,11	2,2	1,09	1,5		
1	Görev 1	2,70	3,16	0,14	0,03	0,96	-0,9	0,92	-1,6		
2	Görev 2	2,22	2,68	-0,29	0,03	0,98	-0,5	1,01	0,1		
Ortalama		2,55	3	0	0,03	1,01	0,3	1	0		
Std. Sapma		0,23	0,23	0,20	0	0,07	1,4	0,07	1,3		
RMSE (Model)= 0,03			Std sapma= 0,20			Ayırma İndeksi= 6,41					
Güvenirlik= 0,98											
Tamamı aynı ki-kare değeri= 138,9				s.d.= 2		p= 0.00					
Rasgele normal ki- kare değeri= 2				s.d.= 1		p= 0,16					

Tablo 14’te görüldüğü üzere görev yüzeyini oluşturan “Görev 1”, “Görev 2” ve “Görev 3” olmak üzere üç tane performans görevinin güçlük düzeyini gösteren logit değerleri sırayla 0,14, -0,29 ve 0,15 olarak bulunmuştur. Puanlayıcı türünün puanlama katılık/cömertliklerini gösteren logit değerlere ait standart hata (RMSE) değeri 0,03 olarak bulunmuştur. Ayırma indeksi 6,41 bulunmuştur. Ayırma indeksinin yüksek olması yine görevler için de istenilen bir durumdur. Performans görevlerine ilişkin güçlük düzeyinin sıralamasının ne derece güvenilir olduğunu gösteren güvenilirlik katsayısının ise 0,98 olarak yeterince yüksek olduğu söylenebilir. Sabit etkiye ait “Üç uygulamanın verilerini içeren analizde performans görevlerinin güçlük düzeyleri arasında anlamlı bir farklılık vardır” hipotezi ki-kare ile test edildiğinde ($X^2 = 138,9$, $sd=2$, $p=0.00$) yokluk hipotezi reddedilmiştir. Dolayısıyla üç performans görevi uygulamasının verilerini içeren analizde performans görevlerinin güçlük düzeyleri arasında manidar farklılık bulunmaktadır. Tablo 14’te iç uyum ve dış uyum değerlerinin 0,92 ile 1,11 arasında yer alması verilerin modele çok iyi uyum sağladığına işaret etmektedir.



BÖLÜM IV

SONUÇ, TARTIŞMA ve ÖNERİLER

Bu bölümde, araştırmanın bulgularına dayalı olarak bulunan sonuçlar, bulguların önemi ve etkileri üzerine tartışma ve çalışmanın sonuçlara dayalı gelecekte yapılabilecek çalışmalar için önerilere yer verilmiştir.

4.1.Sonuçlar ve Tartışma

Öğrencilerin problem çözme becerilerini ölçmeye yönelik birinci, ikinci ve üçüncü performans görevi uygulamalarında; öz, akran ve öğretmen puanlarının çok yüzeyle Rasch ölçme modeline göre FACET analizi için model veri uyumu incelenerek verinin model için uygun olduğu sonucuna varılmıştır.

Öz, akran ve öğretmen puanlamaları ile problem çözme becerisine ilişkin performansı ölçülen 57 öğrencinin FACET analizi ile analize alınan üç uygulamada da hesaplanan ayırma indeksi ve güvenilirlik katsayı değerlerinin oldukça yüksek olduğu belirlenmiştir. Öğrenci yüzeyine ait güvenilirlik katsayısı KR-20, Cronbach alfa ve genellenebilirlik katsayısına eş değer olduğuna göre, birinci, ikinci ve üçüncü uygulamalarda öğrencilerin problem çözme becerilerine ilişkin performanslarının öz, akran ve öğretmen puanlamalarının, iç tutarlılık anlamında güvenirlüğün çok yüzeyle Rasch ölçme modeline göre sağlandığı sonucuna varılmıştır. Ayırma indekslerinin yüksek olması, ki-kare değerleri öğrencilerin belirtilen üç uygulamada da gösterdiği performanslarda anlamlı olarak farklılaşması ve hesaplanan ayırma indeksi oranlarına göre öğrencilerin performans olarak farklı beceri düzeylerine ayrılabilceğinin görülmesi yapılan puanlamalarda puanlayıcının merkeze yönelme etkisinin olmadığı sonucuna işaretir.

Birinci uygulamada öğrencilerin genel olarak logit cetvelde pozitif alanda yer aldığı için genel olarak başarılı olduğu, problem çözme becerisi performansı en yüksek öğrencinin 7

numaralı öğrenci olduğu ve performansı en düşük öğrenci ise 21 numaralı öğrencinin olduğu görülmüştür (Şekil 3). İkinci uygulamada yine öğrencilerin genel olarak başarılı olduğu, en yüksek performansa sahip öğrencinin 41 numaralı öğrenci olduğu ve en düşük performansa sahip öğrencinin 28 numaralı öğrenci olduğu belirlenmiştir (Şekil 5). Üçüncü uygulamada da öğrencilerin genel olarak başarılı olduğu görülmüş, en yüksek performans gösteren öğrenci 24 numaralı öğrencinin olduğu belirlenirken en düşük performans gösteren öğrencinin 38 numaralı öğrenci olduğu gözlenmiştir (Şekil 7). Üç uygulamanın verilerini kapsayan genel analizde öğrencilerin genel olarak başarılı olduğu gözlenmiş, problem çözme becerisinde en yüksek performansı gösteren 7 numaralı öğrenci olurken en düşük performansı gösteren 38 numaralı öğrenci olmuştur (Şekil 9). Grupların başarılı görünmesinin yapılan puanlamalarda öğrenci puanlayıcıların puanlamalarında cömert davranmalarının da etkisinin olabileceği düşünülmektedir.

Problem çözme aşamalarına ilişkin birinci analizde, en kolay yapılan aşama “Uygulama” olurken en zor yapılan aşama “Kontrol etme” olmuştur. “Anlama” ile “Çözüm yolu” aşamaları yaklaşık orta zorlukta bulunmuştur. İkinci analizde, en kolay yapılan aşamanın “Anlama” aşaması olduğu en zor yapılan aşamanın ise “Kontrol etme” olduğu gözlenmiştir. “Uygulama” ile “Çözüm yolu” aşamalarının yaklaşık orta düzeyde zorluk/kolaylık düzeyinde olduğu söylenebilir. Üçüncü analizde ise, “Anlama” ile “Uygulama” aşamalarının en kolay yapılan aşamalar, “Çözüm yolu” ile “Kontrol etme” aşamalarının en zor yapılan olduğu gözlenmiştir. Yapılan dört analizde de “Kontrol etme” aşamasının logit cetvelde en altta yer alması, öğrencilerin çözümü kontrol etme değerlendirme davranışlarının düşük olduğuna işaret eder. Üç uygulamanın verilerini içeren genel analizde, en kolay yapılan aşamaların “Anlama” ile “Uygulama” olduğu görülürken, en zor yapılan aşamaların ise “Çözüm yolu” ile “Kontrol etme” aşamaları olduğu görülmüştür. Aşama yüzeylerine ait ayırma indekslerinin büyük olması istenilen bir durumdur ve dört analiz için de yüksek çıkmıştır. Bu da puanlama anahtarındaki problem çözme aşamalarının aynı kapsamı yoklamadığına işaret olarak düşünülebilir. Puanlama anahtarının aşama yüzeyine ait ki-kare değeri de her bir aşamanın istatistiksel olarak anlamlı farklılaşması da bu durumu destekleyen bir durumdur.

Üç uygulama verilerini içeren genel analizde bir yüzey olarak ele alınan görev yüzeyinde görevler arasında logit değer olarak büyük bir fark bulunmazken kolaydan zora doğru sıralaması “Görev 1”, “Görev 3”, “Görev 2” şeklindedir. Ki-kare değeri görevlerin istatistiksel olarak birbirinden farklılaştığına işaret etmektedir. Ki-kare değerinin

farklılaşması, ayırma indeksinin ve ayırma oranının da yüksek çıkması görevlerin farklı zorluklarda olması ve yapılan puanlamalarda halo etkisi olmadığı sonucu desteklemektedir.

Puanlayıcı türünün katılık/cömertliğine ilişkin birinci, ikinci, üçüncü ve üç uygulama verilerini kapsayan genel analizde en cömert puanlayıcıların öz puanlayıcılar olduğu gözlenmiştir. Öz değerlendirme puanlamalarının öğretmen puanlamalarına göre daha cömert puanlanması alan yazındaki farklı çalışmaların (Farrokhi vd. 2011; Farrokhi vd. 2012; Karakaya 2015; Karakaya vd. 2015) sonuçları ile örtüşmektedir. Buradan öğrencilerin kendilerine yüksek not verme eğiliminde oldukları söylenebilir. Fakat Matsuno (2006) çalışmasında öz, akran ve öğretmen puanlamalarını karşılaştırdığında öğrencilerin öz puanlamalarının katı, akran puanlamalarının cömert ve öğretmen puanlamalarının da ne katı ne de cömert olarak gözlemlemiştir. Bu çalışmada, akran puanlayıcıların birinci, ikinci ve üçüncü analizde de öz puanlayıcılara göre katı, öğretmen puanlayıcılara göre daha cömert puanlama yaptıkları görülmüştür. Bu sonuç Farrokhi, Esfandiari ve Schaefer (2012)'ın yaptıkları çalışma sonuçları ile benzerlik taşımaktadır. Yüzüak, Yüzüak ve Kaptan (2015) da çalışmalarında, akranların öğretmenlere göre daha cömert puanlama yaptıkları ve genel analiz sonuçları dikkate alındığında tutarlı bir puanlama yapıldığı sonucuna varmışlardır. Ancak, akran puanlamaları ile öğretmen puanlamalarını içeren çalışmalara bakıldığında akran puanlayıcıların öğretmene göre daha katı bir puanlama yaptıkları çalışmalar da bulunmaktadır (Karakaya, 2015; Nakamura, 2002).

Yapılan çalışmada dikkate değer bir nokta da birinci, ikinci ve üçüncü analizlerde akran puanlarının logit ölçüleri azalarak öğretmen katılığına yaklaşma durumudur. Nitekim tüm verilerin bulunduğu genel analizde akran puanlayıcıların süreç içerisinde logit değerlerine bakılarak hafta hafta öğretmen puanlayıcıların katılık/cömertliğine yaklaştığı gözlenmiştir. Burada öğretmen puanlayıcıların gerçek öğrenci performanslarına en yakın değerlendirme yaptığı düşünüldüğünde, akran puanlayıcıların katılığının öğretmen puanlayıcılara yaklaşması öğrencilerin akranlarını değerlendirmede bir gelişim gözlendiği çıkarımı yapılabilir. Öğrencilerin uygulama tecrübesi arttıkça daha objektif değerlendirmeye sahip olabileceği yorumlanabilir. Öğrenci ve öğretmen puanlamalarının birlikte kullanıldığı güvenilirlik çalışmalarında puanlayıcı güvenilirliğinin sağlandığı görülmüş ve özellikle sınıf içi değerlendirmelerde pratik bir yöntem olarak bazı çalışmalarda önerilmiştir (Holster, 2012; Karakaya 2015; Karakaya, Sarıtaş ve Salmaner, 2015; Yüzüak, Yüzüak ve Kaptan, 2015).

Çalışmada öğretmen puanlayıcılara bakıldığında üç uygulamada da en katı puanlama yapan puanlayıcılar olmuştur. Üç uygulamanın verilerini içeren genel analizde öğretmen puanlayıcıların hafta hafta incelendiğinde yaklaşık aynı katılığa puanlama yaptıkları gözlenmiştir. Uygulama sürecinde genel olarak puanlayıcı türlerinden öz ve öğretmen puanlayıcıların katılık/cömertliklerinde büyük bir farklılık gözlenmezken akran puanlayıcıların zaman içerisinde yapılan uygulamalar ve bilgilendirmeler ile öğretmen katılığına yaklaştığı yorumlanabilir. Baştürk (2008) çalışmasında, öğretmen ve akran değerlendirmelerin üç neden dolayı daha iyi olabileceğini belirtmiştir: İlki öğrencilerin öğretmenin öğrenci çalışmalarının seviyesini değerlendirmek üzere oluşturduğu değerlendirme ölçütlerini tanınması, ikinci olarak öğrencilerin sınıf içerisinde hem performansı sergileme hem de performansı puanlama olarak iki görev ile aktif bir katılımcı olmasıdır. Üçüncü olarak ise; öğrenciler gerçek sınıf içindeki durumları değerlendirdiğinde içerik ile bağlantılı konuları ve dersin amacını daha iyi öğrenebilirler. Karmaşık ve gelişmesi zaman alan bir beceri olan problem çözme becerisi için de sınıf içinde öz ve akran uygulamalarının öğrencileri olumlu etkileyeceği düşünülmektedir.

Puanlayıcı türü yüzeyine ait güvenilirlik katsayısının yüksek olması ile beraber ve bu yüzeye ait ki-kare değeri, öz akran ve öğretmen puanlayıcıların puanlamalarının istatistiksel olarak anlamlı farklılaştığını göstermektedir. Puanlayıcılar için ki-kare değerinin anlamlı çıkması istenilen bir durum değildir. Fakat puanlayıcı türüne bakıldığında öğretmen puanlayıcıların yanında öz ve akran puanlayıcıların aynı yüzeyde bulunmuş olmasının bu duruma zemin oluşturmuş olabileceği düşünülmektedir. Puanlayıcı yüzeyine ait güvenilirlik katsayısının yüksek fakat ki-kare değerinin anlamlı farklılaştığından dolayı değerlendirmelerin istatistiksel olarak anlamlı farklılaştığı durumlara yönelik öneriler getirilmiştir. Bu durumda Karakaya (2015) tarafından puanlayıcılara bu tür uygulamalar ve puanlama anahtarları hakkında eğitim verilmesi ve öz ile akran değerlendirmelerin sınıf içinde daha çok önem verilerek kullanılması önerilmiştir. Çalışma kapsamında puanlayıcılara en başta eğitim verilmiş sonraki haftalarda da genel dönütler sağlanmıştır. Fakat problem çözme becerilerinin karmaşık bir beceri olması, zaman sınırlaması olması ve öğrencilerin kendini ve akranını değerlendirmesinin alışılmış bir durum olmaması puanlayıcı türlerinin farklılaşmasına sebep olmuş olabilir.

Tüm bu sonuçlardan da yararlanarak ÇYRÖM'nin problem çözme becerilerinin öz, akran ve öğretmen puanları ile değerlendirilmesinde kullanılan her bir yüzey ve her yüzeyin her

elemanına ilişkin ayrıntılı bilgiler sağlaması katkısı ile sınıf içinde de kullanımı elverişli bir ölçme modeli olduğu sonucuna varılmıştır.

4.2.Öneriler

4.2.1.Araştırma Sonuçlarından Çıkarılan Öneriler

1. Çalışmada problem çözme becerilerinin ölçülmesinde; öz, akran ve öğretmen puanlayıcıların uyumu ve yüksek çıkmış ve öğrencileri yüksek güvenilirlikle değerlendirdikleri görülmüştür. Bu uyumun daha da artırılması için öğrenciler ve öğretmenlere puanlama yöntemleri ve ölçme araçları hakkında eğitim verilmesi sağlanarak puanlayıcılar arası uyum daha da artırılabilir.
2. Uygulama sürecinde öğrenci puanlarının uyumu özellikle akran puanlamalarının katılık/cömertliğinin öğretmen puanlamalarına olan yakınlığı göz önünde bulundurulunca öğrencilerin bu tür sınıf içi değerlendirmelerde yer alması önerilmektedir.
3. Çalışmada sadece dört uygulama ile öğrenci puanlamalarında öğretmen puanlamalarına yaklaşma görülmüştür. Öğrencilerin öz ve akran puanlamalarında daha nesnel bir puanlama yapabilmeleri adına, öz ve akran değerlendirme uygulamalarının daha geniş zamanda, sayıca daha fazla yapıldığı bir sürecin planlanarak izlenmesi sağlanabilir.
4. Çalışmada yapılan dört analizde de öğrenciler için DPA yer alan problem çözme aşamaları arasında en zor aşama “kontrol etme ve değerlendirme” aşaması olarak belirlenmiştir. Öz ve akran değerlendirme uygulamalarının kontrol etme, değerlendirme davranışlarının gelişimini sağlamak ve çalışmalara karşı eleştirel bir bakış açısı kazandırmak adına sınıf içinde kullanımı artırılabilir.
5. Çalışma sonuçlarını sebep sonuç olarak daha iyi açıklamak üzere devamında nitel bir çalışma yapılarak desteklenebilir.

4.2.2.İleride Yapılabilecek Çalışmalara İlişkin Öneriler

1. Öğrenci çalışmalarının değerlendirilmesinde öğrencilerin puanlama sürecine dâhil olması ile öğrencinin ilgili beceride gösterdiği performansın değişimi izlenebilir.

2. Problem çözüme becerilerinin değerlendirilmesinde çok yüzeyle Rasch ölçme modeli kullanılarak fakat farklı yüzeyle seçilerek çalışma yürütülebilir.
3. Problem çözüme becerisi dışında farklı disiplinlere ait üst düzey bilişsel becerilerin değerlendirilmesinde uygulanan performans ölçmeleri için benzer bir güvenilirlik çalışmaları yapılabilir.
4. Performansa dayalı durum belirlemede görevleri farklı zorluklarda hazırlayarak öz ve akran puanlamalarında görevlere farklı davranış sergileyip sergilemeyecekleri incelenebilir.



KAYNAKLAR

- Akın, Ö. & Baştürk, R. (2012). Keman eğitiminde temel becerilerin Rasch ölçme modeli ile değerlendirilmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 31(1), 175-187.
- Alharby, E.R (2006). *A comparison between two scoring methods, holistic vs. analytic, using two measurement models, the generalizability theory and the many-facet Rasch measurement, within the context of performans assessment*. Doktora Tezi, The Pennsylvania State University.
- Altun, M. (2002). *Eğitim Fakülteleri ve İlköğretim Öğretmenleri için Matematik Öğretimi*. Bursa: Alfa.
- Arslan A. (2012). *6 yaş grubu çocuklarda genel problem çözme beceriler değerlendirme ölçeğinin geliştirilmesi, geçerlilik ve güvenirlik çalışmaları*. Yüksek Lisans Tezi, Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Atılğan, H. (2004). *Genellenabilirlik kuramı ve çok değişkenlik kaynaklı Rasch modelinin karşılaştırılmasına ilişkin bir araştırma*. Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Baird, J.A., Hayes, M., Johnson, R., Johnson, S. & Lamprianou, I. (2013). Marker effects and examination reliability a comparative exploration from the perspectives of generalizability theory, Rasch modelling and multilevel modelling. 12 Ağustos 2015 tarihinde <http://webarchive.nationalarchives.gov.uk/20141031163546/http://www.ofqual.gov.uk/files/2013-01-21-marker-effects-and-examination-reliability.pdf> sayfasından erişilmiştir.
- Bal, A. P., & Doğanay, A. (2010). İlköğretim beşinci sınıf matematik öğretiminde ölçme-değerlendirme sürecinde yaşanan sorunların analizi. *Kuram ve Uygulamada Eğitim Yönetimi Dergisi*, 16(3), 373-398.

- Baştürk, R. (2008). Applying the many – facet Rasch model to evaluate powerpoint presentation performance in higher education. *Assesment and Evaluation in Higher Education*, 33 (4), 431 – 444. 09 Aralık 2014 tarihinde <http://www.tandfonline.com/doi/pdf/10.1080/02602930701562775> sayfasından erişilmiştir.
- Baştürk, R. (2010). Bilimsel araştırma ödevlerinin çok yüzeyli Rasch ölçme modeli ile değerlendirilmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(1), 51-57.
- Baykul, Y. (1999). *İlköğretimde matematik öğretim*. Ankara: Anı
- Baykul, Y. (2006). *İlköğretimde matematik öğretimi (1- 5 sınıflar)*. Ankara: Pegem Akademi.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. İkinci Baskı. Ankara: Pegem Akademi.
- Berberoğlu, G. (1988). *Seçme amacıyla kullanılan testlerde Rasch modelinin katkıları*. Doktora Tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Bingham, A. (2004). *Çocuklarda problem çözme yeteneklerinin geliştirilmesi* (A. Ferhan Oğuzhan, Çev.). İstanbul: Milli Eğitim Bakanlığı.
- Bozkurt, E. & Demir, R. (2012). Öğrenci görüşleriyle akran değerlendirme: Bir örnek uygulama. *İlköğretim Online*, 11(4), 966-978. [Online]: 12 Aralık 2014 tarihinde <http://ilkogretim-online.org.tr> sayfasından erişilmiştir.
- Büyükkıdık, S. (2012). *Problem çözme becerisinin değerlendirilmesinde puanlayıcılar arası güvenilirliğin klasik test kuramı ve genellenebilirlik kuramına göre karşılaştırılması*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Büyüköztürk, Ş., Çakmak, K. E., Akgün, E. Ö., Karadeniz, Ş. & Demirel, F. (2013). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi (14. Baskı).
- Brookhart, Susan M. (2013). *How to create and use rubrics for formative assessment and grading*. Virginia USA: ASCD Alexandria.

- Brown, William L., O’Gorman, K. & Du, Y. (1996, Nisan). *The reliability and validity of mathematics performance assessment*. “Annual Meeting of the American Educational Research Association” de sunulmuş bildiri, NY, New York.
- Crocker, L.& Algina, J. (1986). *Introduction to classical & Modern test theory*. USA:Harcourt Brace Javanovich College.
- Çakıcı Eser, D. ve Gelbal, S. (2013). Genellenebilirlik kuramı ve lojistik regresyona dayalı hesaplanan puanlayıcılar arası tutarlığın karşılaştırılması. *Kastamonu Eğitim Dergisi*, 21 (2), 421-438.
- Farrokhi F., Esfandiari R. & Dalili, M. V. (2011). Applying the many-facet rasch model to detect centrality in self-assessment, peer-assessment and teacher assessment. *World Applied Sciences Journal* 15 (Innovation and Pedagogy for Lifelong Learning): 70-77.
- Farrokhi F., Esfandiari R. & Schaefer E. (2012). A many-facet Rasch measurement of differential rater severity/leniency in three types of assessment. *JALT Journal*, 34(1), 79-102.
- Gelbal, S. & Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* 33, 135-145.
- Güler, N. (2008). *Klasik test kuramı genellenebilirlik kuramı ve Rasch modeli üzerine bir araştırma*. Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Güler, N. & Gelbal, S. (2010). A study based on classic test theory and many facet Rasch model. *Eurasian Journal of Educational Research*, 38, 108-125.
- Güler, N. (2014). Analysis of open-ended statistics questions with many facet Rasch model. *Eurasian Journal of Educational Research*, 55, 73-90.
- Gündüz Sefer, D. (2006). *Matematik dersinde problem çözme becerilerinin dereceli puanlama anahtarı kullanılarak değerlendirilmesi*, Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Hetherman, S. C. (2004). *An application of multi faceted rasch measurement to monitor effectiveness of the written composition in english in the New york City Department of Education*. Doktora Tezi, Teacher College, Colombia University, Colombia.

- Holster, T. A., (2012). Many-faceted Rasch analysis of student peer assessment. 76,69-86
- İlhan, M. (2015). *Standart ve solo taksonomisine dayalı rubrikler ile puanlanan açık uçlu matematik sorularında puanlayıcı etkilerinin çok yüzeyli Rasch ölçme modeli ile incelenmesi*. Doktora Tezi, Gaziantep Üniversitesi, Eğitim Bilimleri Enstitüsü, Gaziantep.
- Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130–144.
- Karakaya, İ. (2015). Comparison of self, peer and instructor assessments in the portfolio assessment by using many facet Rasch model. *Journal of Education and Human Development*, 4(2), 182-192.
- Karakaya, İ., Sarıtaş, S. ve Salmaner, R. (2015, Mayıs). *Assessment of performance-based tasks within the context of statistics lesson with multi-faceted rasch model*. “The International Congress on Education for the Future: Issues and Challenges (ICEFIC 2015)” te sunulmuş bildiri, Ankara Üniversitesi, Ankara.
- Karasar, N. (1998). *Bilimsel araştırma yöntemleri*. Ankara: Nobel.
- Karataş, İ. & Güven, B. (2003). Problem çözme davranışlarının değerlendirilmesinde kullanılan yöntemler: Klinik mülakatın potansiyeli. *İlköğretim- Online* 2 (2), 2-9. [Online]: 12 Aralık 2014 tarihinde <http://ilkogretim-online.org.tr> sayfasından erişilmiştir.
- Kılıç, S. (2009). *Ölçümlerin uyumluluğu ve tıptaki uygulamalar*. Yüksek Lisans Tezi, Çukurova Üniversitesi, Sağlık Bilimleri Enstitüsü, Adana.
- Kim, S. C. & Wilson, M. (2009). A comparative analysis of the ratings in performance assessment using generalizability theory and the many-facet Rasch model. *Journal of Applied Measurement*, 10(4), 408-423.
- Kuehl, G., Sofronas, K. & Lau, A. (2015). Pre-service and in-service teachers’ rubric assessments of mathematical problem solving. proceedings 2014. Paper 1.
- Kutlu, Ö., Doğan, C. D. & Karakaya, İ. (2010). *Öğrenci başarısının belirlenmesi - performansa ve portfolyoya dayalı durum belirleme*. Ankara: Pegem Akademi, (3. Baskı).

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563–575.

Linacre, J. M. (1993, Nisan). *Generalizability theory and many facet rasch measurement*. “Annual Meeting Of The American Educational Research Association” de sunulmuş bildiri, Georgia, Atlanta. 30.11.2014 tarihinde <http://files.eric.ed.gov/fulltext/ED364573.pdf> sayfasından erişilmiştir.

Linacre, J. M. (2014). A user's guide to FACETS Rasch-model computer programs.

Macmillan, P. D. (2000). Classical generalizability and multifaceted rasch detection interrater variability in large, sparse data set. *Journal of Experimental Education*. 68, 2, 167-190.

Matsuno, S. (2006). Self, peer, and teacher assessment in Japanese university efl writing classrooms. Doktora Tezi. Temple University.

MEB Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü, (2011). *PISA 2012 Türkiye*. Ankara: Eğitek. 12 Aralık 2014 tarihinde

<http://pisa.meb.gov.tr/wp-content/uploads/2013/07/PISA-kitab%C4%B1.pdf> sayfasından erişilmiştir.

MEB (2009). *İlköğretim matematik dersi (6- 8. sınıflar) öğretim programı ve kılavuzu*. Ankara Devlet Kitapları.

MEB Talim ve Terbiye Kurulu Başkanlığı, (2013). *Ortaokul matematik dersi (5,6,7, ve 8. Sınıflar) Öğretim Programı*.

Morgan, C. T. (2006). *Psikolojiye giriş*. (H. Arıncı vd. Çev.). Ankara: Meteksan.

Moskal, B. M. (2000). Scoring rubrics: what, when and how?. *Practical Assessment, Research and Evaluation*, 7(3), 70-80.

Nakamura, Y. (2000). Many facet rasch based analysis of communicative language testing results. *Journal of communication students*, 12, 3-13.

Nakamura, Y. (2002). Teacher assessment and peer assessment in practise. *Educational Studies*, 44.

National Council of Teachers of Mathematics NCTM, (2000). *Principles and standards for school mathematics*. Reston, VA: N C T M,

- OECD, (2012). PISA 2012 Results: Creative problem solving: students' skills in tackling real-life problems (Volume V) 25 Haziran 2015 tarihinde <http://www.oecd.org/pisa/keyfindings/PISA-2012-PS-snapshot-performance.pdf> sayfasından erişilmiştir.
- OECD, (2012). "PISA 2012 results: creative problem solving (volume v): students' skills in tackling real-life problems" assessing problem-solving skills in PISA 2012. 12 Aralık 2014 tarihinde <http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm> sayfasından erişilmiştir.
- Orbeyi, S. & Güven B. (2008). Yeni ilköğretim matematik dersi öğretim programı'nın değerlendirme ögesine ilişkin öğretmen görüşleri. *Eğitimde Kuram ve Uygulama*, 4 (1):133-147.
- Özmen-Hızarcıoğlu, B. (2013). *Problem çözme sürecinde dereceli puanlama anahtarı (rubrik) kullanımında puanlayıcı uyumunun incelenmesi*. Yüksek Lisans Tezi, Abant İzzet Baysal Üniversitesi Eğitim Bilimleri Enstitüsü, Bolu.
- Parlak, B. & Doğan, N. (2014). Dereceli puanlama anahtarı ve puanlama anahtarından elde edilen puanların uyum düzeyleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* 29(2), 189-197.
- Polya, G. (1973). *How to solve it? A new aspect of mathematical method*. Second edition. New Jersey: Princeton University.
- Punch, K. F. (1998). *Introduction to social research: Quantitative and qualitative approaches*. Thousand Oaks, CA: Sage.
- Romagnano, L. (2001). The myth of objectivity in mathematics assessment. *Principles and Standards for School Mathematics (NCTM)* 94(1), 22.
- Semerci, Ç. (2011a). Analyzing micro teaching applications with many-facet Rasch measurement model. *Education & Science*, 36(161), 14-25.
- Semerci, Ç. (2011b). Doktora yeterlikler çerçevesinde öğretim üyesi, akran ve öz değerlendirmelerin Rasch ölçme modeliyle analizi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2(2), 164-17.

- Shavelson, R. J., & Webb, M. N. (1991). *Generalizability theory: A Primer*. SAGE Publication, Inc., California.
- Smith, V. E. & Kulikowich, J. M. (2004). An application of generalizability theory and many-facet Rasch measurement using a complex problem-solving skills assessment. *Educational and Psychological Measurement*, 64, 617- 639.
- Şanlı, N. (2005). Çocukların problem çözme becerisini geliştirmek için. *Çocuk Çocuk Dergisi*, (52); 20-21.
- Tekin, H. (1991). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı (6. Baskı)
- Toluk, Z. & Olkun, S. (2002). Problem solving in turkish mathematics education: primary school mathematics textbooks. *Educational Sciences: Theory & Practice*, 2(2).
- Topping, K. (1998) Peer assessment between students and in colleges and universities, *Review of Educational Research*, 68(3), 249-276
- Topses, G. (2003). *Öğrenme ve öğretim süreçlerinin psikolojisi. Gelişim ve öğrenme psikolojisi*. Ankara: Nobel.
- Umay, A. (1996). Matematik eğitimi ve ölçülmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 12, 145-149.
- Uysal, K. (2008). *Öğrencilerin ölçme değerlendirme sürecine katılımı: akran değerlendirme ve öz değerlendirme*. Yüksek Lisans Tezi, Abant İzzet Baysal Üniversitesi, Sosyal Bilimler Enstitüsü, Bolu.
- Uzun, A. (2010). *İlköğretim 4 ve 7. sınıf öğrencilerinin öz değerlendirme uygulamalarına yönelik tutumları*. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, Eğitim Bilimleri Enstitüsü, İzmir.
- Weigle, S. C. (1999). Investigating rater/prompt interactions in writing assessment: quantitative and qualitative approaches. *Assessing Writing*, 6(2),145-178.
- White, J. (1999). Thinking about assessment, *Journal of Philosophy of Education*, 33(2), 201.
- William, D. (2003). National curriculum assessment: how to make it better. *Research Papers in Education*, 18(2), 129-136.

- Wolfe, E. W. & Chiu, Chris W. T. (1997, Mart). *Detecting rater effects with a multi-faceted rating scale model*. “Annual Meeting of the National Council on Measurement in Education” de sunulmuş bildiri, IL, Chicago.
- Yayan, B. (2010). *Student and teacher characteristics related to problem solving skills of the sixth grade Turkish students*. Doktora Tezi, Orta Doğu Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Yıldırım, R. (2003). *Yaratıcılık ve yenilik*. Ankara: Sistem (4. Baskı).
- Yurdugül, H. (2005, Eylül). *Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması*. XIV. Ulusal Eğitim Bilimleri Kongresi’nde sunulmuş bildiri, Pamukkale Üniversitesi, Denizli.
- Yüzüak, A. V., Yüzüak, B. & Kaptan, F. (2015). Performans görevinin akran gruplar ve öğretmen yaklaşımları doğrultusunda çok-yüzeyle Rasch ölçme modeli ile analizi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 1-11.

EKLER



EK 1. Uzman Değerlendirme Formları

EK 1.a Performans Görevi Uzman Değerlendirme Formu

Değerli Uzmanım;

“Problem Çözme Becerilerinin Değerlendirilmesinde Öz, Akran ve Öğretmen Puanlarının Çok Yüzeyle Rasch Ölçme Modeli İle İncelenmesi” isimli yüksek lisans tez çalışmamızda kullanmak üzere performans görevleri geliştirmekteyiz. **6. sınıf** öğrencilerinin problem çözme becerilerini ölçmeye yönelik geliştirilen bu ölçme araçlarının geçerlik ve güvenilirliğine yönelik kanıt elde etmek için görüşlerinize başvurmak istiyoruz.

Ölçme araçları için aşağıda belirtilen ölçütler açısından “**Uygun, Düzeltmeli, Uygun Değil**” olmak üzere üç düzey bulunmaktadır. Ölçme araçlarını ölçütlere uygunluk açısından değerlendirerek ilgili kutucuğa “**X**” işareti koyunuz. Uzman formunda olmayan ve belirtmek istediğiniz ölçüt varsa önerilerinizi belirtiniz.

Değerli katkılarınız için şimdiden teşekkür ederiz.

Tez Danışmanı: Doç. Dr. İsmail KARAKAYA

Tez Danışan: Arş. Gör. Seyhan SARITAŞ

Gazi Üniversitesi
Gazi Eğitim Fakültesi
Eğitim Bilimleri Bölümü
Eğitimde Ölçme ve Değerlendirme AD

Gazi Üniversitesi
Gazi Eğitim Fakültesi
Eğitim Bilimleri Bölümü
Eğitimde Ölçme ve Değerlendirme AD

Değerlendiren Uzmanın;

Adı soyadı:.....

Unvanı:.....

Deneyimi:.....

.....

.....

Eğitim Düzeyi:.....

PERFORMANS GÖREVLERİ İÇİN DEĞERLENDİRME FORMU

ÖLÇÜTLER	Problem	DÜZEYLER		
		Uygun	Düzeltilmeli	Uygun Değil
Yaşamda önemli olan olaylara ya da sınıf içi öğrenmelere uygundur. Öneriler:	1.			
	2.			
	3.			
	4.			
Yaratıcılık, problem çözme, araştırma yapma, analitik düşünme gibi üst düzey zihinsel süreçlerden bir ya da bir kaçına uygundur. Öneriler:	1.			
	2.			
	3.			
	4.			
Öğrencilerin dil düzeyine uygun, anlaşılır ve ilgi çekicidir. Öneriler:	1.			
	2.			
	3.			
	4.			
Problem durumu belirlenen zihinsel süreçlere ve öğrencinin sınıf düzeyine uygundur. Öneriler:	1.			
	2.			
	3.			
	4.			
Problemi çözmek için belirlenen süre öğrencinin sınıf düzeyine uygundur. Öneriler:	1.			
	2.			
	3.			
	4.			

Genel Görüş ve Öneriler:

EK 1.b Dereceli Puanlama Anahtarı Uzman Değerlendirme Formu

Değerli Uzmanım;

“Problem Çözme Becerilerinin Değerlendirilmesinde Öz, Akran ve Öğretmen Puanlarının Çok Yüzeyle Rasch Ölçme Modeli İle İncelenmesi” isimli yüksek lisans tez çalışmamızda kullanmak üzere ölçme araçları geliştirmekteyiz. **6. sınıf** öğrencilerinin problem çözme becerilerini değerlendirmeye yönelik geliştirilen dereceli puanlama anahtarının geçerlik ve güvenilirliğine yönelik kanıt elde etmek için görüşlerinize başvurmak istiyoruz.

Ölçme araçları için aşağıda belirtilen ölçütler açısından “**Uygun, Düzeltmeli, Uygun Değil**” olmak üzere üç düzey bulunmaktadır. Ölçme araçlarını ölçütlere uygunluk açısından değerlendirerek ilgili kutucuğa “**X**” işareti koyunuz. Uzman formunda olmayan ve belirtmek istediğiniz ölçüt varsa önerilerinizi belirtiniz.

Değerli katkılarınız için şimdiden teşekkür ederiz.

Tez Danışmanı: Doç. Dr. İsmail KARAKAYA

Tez Danışan: Arş. Gör. Seyhan SARITAŞ

Gazi Üniversitesi
Gazi Eğitim Fakültesi
Eğitim Bilimleri Bölümü
Eğitimde Ölçme ve Değerlendirme AD

Gazi Üniversitesi
Gazi Eğitim Fakültesi
Eğitim Bilimleri Bölümü
Eğitimde Ölçme ve Değerlendirme AD

Değerlendiren Uzmanın;

Adı soyadı:.....

Unvanı:.....

Deneyimi:.....

.....

.....

Eğitim Düzeyi:.....

DERECELİ PUANLAMA ANAHTARI UZMAN DEĞERLENDİRME FORMU

ÖLÇÜTLER	DÜZEYLER		
	Uygun	Düzeltilmeli	Uygun Değil
Performansı değerlendirmek için belirlenen problem çözme aşamaları amacına uygun ve yeterlidir. Öneriler:			
Puanlama anahtarında bulunan problem çözme aşamaları ayrı ayrı tanımlanmış ve her aşama öğrencinin kullandığı bir beceri hakkında bilgi vermiştir. Öneriler:			
Her düzeyde istenilen performanslar tanımlanırken olabildiğince az sayıda görelî ifade kullanılmıştır. Öneriler:			
Dil ve anlatım Türkçeye uygundur. Öneriler:			
Öğrencilerin dil düzeyine uygun, anlaşılır ifadeler kullanılmıştır. Öneriler:			
Performansın yeterliliği için belirlenen düzey sayısı duyarlılık için uygundur. Öneriler:			

Genel Görüş ve Öneriler:

EK 2. Uygulanan Performans Görevleri

EK 2.a Deneme Haftasında Uygulanan Performans Görevi



Sevgili öğrenciler,

Aşağıdaki problem durumu, sizlerin problem çözme becerinizi belirlemek amacıyla oluşturulmuştur. Problem durumunu okuduktan sonra cevaplarınızı problem çözme süreçlerini dikkate alarak ilgili boşluklara yazılı olarak ifade etmelisiniz. Puanlamanız dereceli puanlama anahtarı ile yapılacaktır. Çalışmanın sonuçları bilimsel bir çalışmada kullanılacak olup içtenlikle cevaplamanız beklenmektedir. Katılımınız için teşekkürler. 😊 Süre: 15 dk

Düzyey: 6. sınıf

Kazanım: Doğal sayılarla dört işlem yapmayı gerektiren problemleri çözer.

Beklenen performans: Problem çözme, akıl yürütme, ilişkilendirme



PROBLEM DURUMU:

Her cumartesiye ailece beraber geçiren Öztürk ailesinin, bu cumartesi planını ve harcama bütçesini 12 yaşındaki 6. sınıf öğrencisi Mustafa hazırlayacaktır. Bütçe için 100 TL'si olan Mustafa; 2. sınıfta olan kardeşi Uğur, annesi ve babası ile gidebileceği etkinlik ve etkinlik ücretlerini araştırarak aşağıdaki tabloyu hazırlamıştır. Evden herhangi bir etkinliğe gidiş ve etkinlikten etkinliğe geçerken araç kullanmaları gerekmektedir. Gün sonunda yine ailenin eve dönmüş olmaları gerektiğine göre Mustafa'nın şartlarında siz olsaydınız cumartesi gününü bütçeyi aşmadan nasıl planlardınız?



Etkinlikler		Ücret	
Bilim ve Kültür Parkı	0-11 yaş	5 TL	
	12 yaş ve üzeri	10 TL	
Sinema	Öğrenci	12 TL	
	Tam	18 TL	
Yemek	Çocuk menü	7 TL	
	Tam Porsiyon	10 TL	
	1.5 Porsiyon	14 TL	
Tiyatro	Öğrenci	7 TL	
	Tam	10 TL	
Yol	Taksi	Bir yerden bir yere geçiş için 20 TL	
	Toplu taşıma	Öğrenci	2 TL
		Tam	3 TL

(Her bir etkinliğe gidiş ve ya geliş için taksii ücretinin sabit olduğunu var sayıyoruz!)

❖ Problemi dikkatlice okuyarak ne anladığınızı yazınız. (İsterseniz şekil, görsel kullanabilirsiniz.)

❖ Problemin çözümüne yönelik uygulamanızı sebebi ile açıklayınız.

❖ Problemin çözümü için seçtiğiniz yolu uygulayınız.

❖ Yaptığınız çözümü kontrol edin ve değerlendiriniz.

EK 2.b Birinci Uygulamadaki Performans Görevi

Sevgili öğrenciler,

Aşağıdaki problem durumu sizlerin problem çözme becerinizi belirlemek amacıyla oluşturulmuştur. Problem durumunu okuduktan sonra cevaplarınızı problem çözme süreçlerini dikkate alarak ilgili boşluklara yazılı olarak ifade etmelisiniz. Puanlamanız dereceli puanlama anahtarı ile yapılacaktır. Çalışmanın sonuçları bilimsel bir çalışmada kullanılacak olup içtenlikle cevaplamanız beklenmektedir. Katılımınız için teşekkürler. ☺ Süre:15 dk

Düzyey:6. Sınıf

Kazanım: Kesirlerle işlem yapmayı gerektiren problemleri çözer.

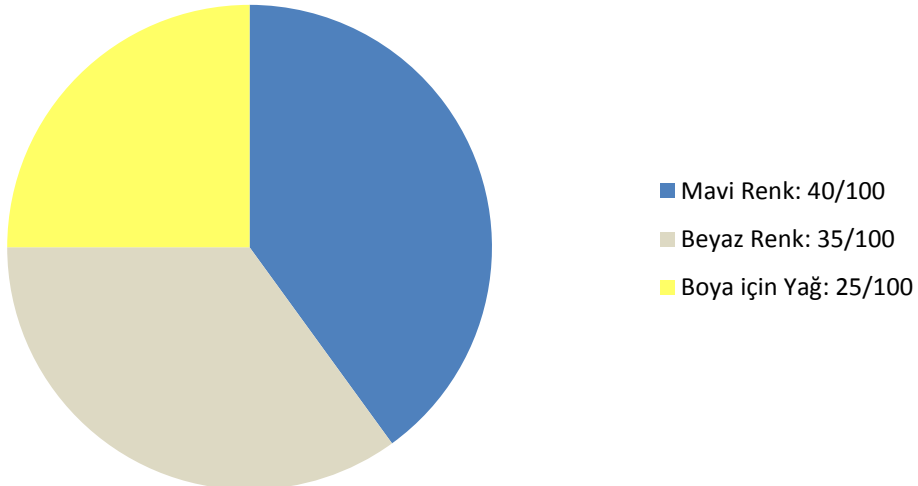
Beklenen performans: Problem çözme, akıl yürütme, ilişkilendirme



PROBLEM DURUMU:

Ceren ile Celal okulun bahçe duvarını boyama görevini almışlardır. Duvarın rengine karar verilebilmesi için okulda yapılan oylamada öğrencilerin en çok isteği renk gök mavisi rengi olmuştur. Ancak okulun deposunda mavi ve beyaz renkte boya ile boyayı açmak için kullanılan yağdan bol miktarda bulunmaktadır. Ellerindeki boyaları kullanarak gök mavisi renginde boya üretmek isteyen Ceren ile Celal'in grafikte verilen oranları kullanmaları gerekmektedir. Bahçe duvarını boyamak için 200 lt boyaya ihtiyaç olduğuna göre kullanılması gereken mavi boya, beyaz boya ve yağ miktarlarını bulunuz.

BOYA KARIŞIMI



❖ Problemi dikkatlice okuyarak ne anladığınızı yazınız. (İsterseniz şekil, görsel kullanabilirsiniz.)

❖ Problemin çözümüne yönelik uygulamanızı sebebi ile açıklayınız.

❖ Problemin çözümü için seçtiğiniz yolu uygulayınız.

❖ Yaptığınız çözümü kontrol edin ve değerlendiriniz.

EK 2.c İkinci Uygulamadaki Performans Görevi

Sevgili öğrenciler,

Aşağıdaki problem durumu sizlerin problem çözme becerinizi belirlemek amacıyla oluşturulmuştur. Problem durumunu okuduktan sonra cevaplarınızı problem çözme süreçlerini dikkate alarak ilgili boşluklara yazılı olarak ifade etmelisiniz. Puanlamanız dereceli puanlama anahtarı ile yapılacaktır. Çalışmanın sonuçları bilimsel bir çalışmada kullanılacak olup içtenlikle cevaplamanız beklenmektedir. Katılımınız için teşekkürler. 😊 Süre:15 dk

Düzyey:6. sınıf

Kazanım: Kesirlerle işlem yapmayı gerektiren problemleri çözer.

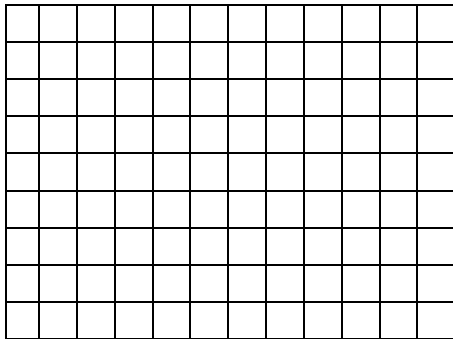
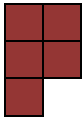
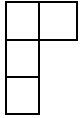
Beklenen performans: Problem çözme, akıl yürütme, ilişkilendirme



PROBLEM DURUMU :

Deniz, anneler günü için eni 9, boyu 12 birim olan mutfak masasının yüzeyini elindeki el işi kâğıtlarından yapılmış iki farklı renk ve biçimdeki model ile kaplamayı planlamaktadır. Deniz bu iki modeli eşit sayıda kullanmak istediğine göre masayı tamamen kapladığında masanın kaçta kaç beyaz renkte olacaktır?

(Deniz'in elindeki iki model ile masanın boyutları aynı tür birimlerden oluşmaktadır.)



9 birim

12 birim



❖ Problemi dikkatlice okuyarak ne anladığınızı yazınız. (İsterseniz şekil, görsel kullanabilirsiniz.)

❖ Problemin çözümüne yönelik uygulamanızı sebebi ile açıklayınız.

❖ Problemin çözümü için seçtiğiniz yolu uygulayınız.

❖ Yaptığınız çözümü kontrol edin ve değerlendiriniz.

EK 2.c Üçüncü Uygulamadaki Performans Görevi

Sevgili öğrenciler,

Aşağıdaki problem durumu, sizlerin problem çözme becerinizi belirlemek amacıyla oluşturulmuştur. Problem durumunu okuduktan sonra cevaplarınızı problem çözme süreçlerini dikkate alarak ilgili boşluklara yazılı olarak ifade etmelisiniz. Puanlamanız dereceli puanlama anahtarı ile yapılacaktır. Çalışmanın sonuçları bilimsel bir çalışmada kullanılacak olup içtenlikle cevaplamanız beklenmektedir. Katılımınız için teşekkürler. 😊 Süre:15 dk

Düzyey:6. sınıf

Kazanım: Doğal sayılarla dört işlem yapmayı gerektiren problemleri çözer.

Beklenen performans: Problem çözme, akıl yürütme, ilişkilendirme



PROBLEM DURUMU:

Aşağıda bir diyetisyenin 10-15 yaş grubundaki öğrencilerin kalori ihtiyaçları ile ilgili ifadeleri yer almaktadır. Buna göre tabloda verilen besin kalori değerlerini ve diyetisyenin önerilerini dikkate alarak 1 günlük (3 Öğün; kahvaltı, öğle ve akşam yemeği) beslenme planı oluşturunuz ve toplam kalori miktarını hesaplayınız.

Merhaba sevgili öğrenciler! Beslenirken günlük kaç kaloriye ihtiyacımız var biliyor musunuz? Biz diyetisyenlere göre 10- 15 yaş grubundakilerin gelişimleri için ihtiyacı olan günlük kalori miktarı ortalama 2000 ile 2300 kalori arasında olmalıdır.



1 Porsiyon için Besin Kalori (Kkcal) Değerleri

BESİN	Kalori Değeri	BESİN	Kalori Değeri	BESİN	Kalori Değeri	BESİN	Kalori Değeri
Gözleme	490	Et suyu çorba	291	Yoğurt	130	Puding/ Muhallebi	175
Tost	267	Sebze çorbası	85	Ayran	90	Ayva tatlısı	230
Yumurta	90	Bulgur pilavı		Nar suyu	120	Sütlaç	140
Beyaz peynir	110	Tavuk ızgara	130	Portakal suyu	100	Çikolatalı kek	190
Yeşil zeytin	73	Patates Kızartması	319	Elma	68	Meyveli kek	138
Reçel	54	Kuzu ızgara	148	Mandalina	56	Kurabiye	350
Bal	64	Kıymalı pide	500	Mantı	490	Süt	91
Siyah zeytin	116	Pirinç pilavı	326	Bezelye	69	Şekersiz çay	1
Simit	434	Ispanak	78	Çikolata	136	1 küp şeker	9
Domates çorbası	189	Hamsi	177	1dilim ekmek	90	1 dilim kepekli	60

Çözüm için isterseniz aşağıdaki tabloyu kullanabilirsiniz.

ÖĞÜNLER	KALORİ MİKTARI
Kahvaltı:	
Öğle:	
Akşam:	
Toplam kalori miktarı:	

❖ Problemi dikkatlice okuyarak ne anladığınızı yazınız. (İsterseniz şekil, görsel kullanabilirsiniz.)

❖ Problemin çözümüne yönelik uygulamanızı sebebi ile açıklayınız.

❖ Problemin çözümü için seçtiğiniz yolu uygulayınız.

❖ Yaptığınız çözümü kontrol edin ve değerlendiriniz.

EK 3. Dereceli Puanlama Anahtarları

EK 3.a Öğretmene Yönelik Dereceli Puanlama Anahtarı

Problem Çözme Becerisine İlişkin Dereceli Puanlama Anahtarı (Rubrik)

Değerli Öğretmenim,

Aşağıda problem çözme becerisini değerlendirmeye yönelik dereceli puanlama anahtarı bulunmaktadır. Dereceli puanlama anahtarında, öğrencilerin problem çözme sürecinin aşamaları ve performanslarını gösterebilecek "0", "1", "2", "3", "4" harfleri ile gösterilen düzeyler belirlenmiştir. Problem çözme sürecinin tüm alt başlıklarına karşılık gelen her bir düzeyi gösteren performans özellikleri açıklanmıştır. Sizden öğrencilerin gösterdikleri performanslara karşılık gelen düzeydeki rakamları yazarak dereceli puanlama anahtarına göre puanlamanız istenmektedir.

Katkılarınız için şimdiden teşekkürler.

Seyhan SARTIŞ

AŞAMALAR	PERFORMANSIN DÜZEYLERİ				Başarı Puanı	
	0	1	2	3		4
Problemi Anlama	Cevap yok.	Problemin özetini yanlış yapmış. Problemden verilenleri, problemin koşulunu ve istenileni yanlış yazmış.	Problemin özetini yüzeysel yapmış. Problemin verilenleri, koşulu ve istenilenden birini doğru yazmış.	Problemin özetini doğru ama eksik yapmış. Problemin verilenleri, koşulu ve istenilenin ikisini doğru yazmış.	Problemi eksiksiz olarak özetlemiş. Problemin verilenleri, koşulunu ve istenileni doğru yazmış.	
Çözüm ile İlgili Stratejinin Seçimi	Cevap yok.	Problemin çözümüne yönelik yanlış gerekçelerle yanlış bir çözüm yolu seçmiş.	Problemin çözümünü doğru seçmiş fakat istenilene nasıl ulaşacağını yazmamış.	Problemin çözümünü doğru seçmiş, istenilene nasıl ulaşacağını yazmış fakat gerekçesini yanlış yazmış.	Problemin çözümüne yönelik doğru bir çözüm seçerek gerekçesini açıklamış.	
Seçilen Çözüm Stratejisini Uygulama	Cevap yok.	Problemin çözümüne ilişkin işlemlerin tümünde hata yapmış veya doğru cevaba şans eseri ulaşmış.	Problemin çözümüne ilişkin işlemlerde en az iki veya daha fazla hata yapmış.	Problemin çözümüne ilişkin işlemlerde sadece bir yerde hata yapmış.	Problemin çözümüne ilişkin işlemlerin tamamını hatasız yaparak doğru cevaba ulaşmış.	
Çözümü Kontrol Etme Ve Değerlendirme	Cevap yok.	Çözümün doğruluğunu ve değerlendirmesini yanlış yapmış.	Çözümün doğruluğunu kısmen yapıp yorumu yanlış yapmış.	Çözümün doğruluğunu yapmamış veya yanlış yapmış.	Çözümün doğruluğunu ve değerlendirmesini doğru yapmış.	
Toplam Başarı Puanı						

EK 3.b Öğrenciye Yönelik Dereceli Puanlama Anahtarı

Problem Çözme Becerisine İlişkin Dereceli Puanlama Anahtarı

Sevgili Öğrenci,

Aşağıda problem çözme becerinizi değerlendirmeye yönelik dereceli puanlama anahtarı bulunmaktadır. Dereceli puanlama anahtarında, öğrencilerin problem çözme sürecinin aşamaları ve performanslarını gösterebilecek "O", "D", "C", "B", "A" harfleri ile gösterilen düzeyler belirlenmiştir. Problem çözme sürecinin tüm alt başlıklarına karşılık gelen her bir düzeyi gösteren performans özellikleri açıklanmıştır. Sizden gösterilen performanslara karşılık gelen düzeydeki harfleri yazarak dereceli puanlama anahtarına göre puanlamanız istenmektedir.

Katkılarınız için şimdiden teşekkürler.

Seyhan SARITAŞ

AŞAMALAR	PERFORMANSIN DÜZEYLERİ						Öz	Akran
	"O"	Daha çok çaba! "D"	Ha Gayret! "C"	Daha iyi olabilir! "B"	Mükemmel! "A"			
Problemi Anlama	Cevap yok.	Problemin özetini yanlış yapmış. Problemde verilenleri, problemin koşulunu ve istenileni yanlış yazmış.	Problemin özetini yüzeysel yapmış. Problemin verilenleri, koşulu ve istenilenden birini doğru yazmış.	Problemin özetini doğru ama eksik yapmış. Problemin verilenleri, koşulu ve istenilenin ikisini doğru yazmış.	Problemi eksiksiz olarak özetlemiş. Problemin verilenleri, koşulunu ve istenileni doğru yazmış.			
Çözüm ile İlgili Stratejinin Seçimi	Cevap yok.	Problemin çözümüne yönelik yanlış gerekçelerle yanlış bir çözüm yolu seçmiş.	Problemin çözümünü doğru seçmiş fakat istenilene nasıl ulaşacağını ve gerekçesini yazmamış.	Problemin çözümünü seçmiş, istenilene nasıl ulaşacağını yazmış fakat gerekçesini yanlış yazmış.	Problemin çözümüne yönelik doğru bir çözüm seçerek gerekçesini açıklamış.			
Seçilen Çözüm Stratejisini Uygulama	Cevap yok.	Problemin çözümüne ilişkin işlemlerin tümünde hata yapmış veya doğru cevaba şans eseri ulaşmış.	Problemin çözümüne ilişkin işlemlerde en az iki veya daha fazla hata yapmış.	Problemin çözümüne ilişkin işlemlerde sadece bir yerde hata yapmış.	Problemin çözümüne ilişkin işlemlerin tamamını hatasız yaparak doğru cevaba ulaşmış.			
Çözümü Kontrol Etme Ve Değerlendirme	Cevap yok.	Çözümün sağlanmasını ve değerlendirilmesini yanlış yapmış.	Çözümün sağlanmasını kısmen yapıp yorumu yanlış yapmış.	Çözümün sağlanmasını yapmış, yorumunu yapmamış veya yanlış yapmış.	Çözümün sağlanmasını ve değerlendirilmesini doğru yapmış.			
Toplam Başarı Puanı								

EK 4. Araştırma İzni



T.C.
ANKARA VALİLİĞİ
Milli Eğitim Müdürlüğü

Sayı : 14588481/605.99/971568

28/01/2015

Konu: Araştırma izni

GAZİ ÜNİVERSİTESİNE
(Eğitim Bilimleri Enstitüsü)

İlgi: a) MEB Yenilik ve Eğitim Teknolojileri Genel Müdürlüğünün 2012/13 nolu Genelgesi.
b) 05/01/2015 tarihli ve 6400 sayılı yazınız.

Üniversiteniz Eğitim Bilimleri Enstitüsü Yüksek Lisans Öğrencisi Seyhan SARITAŞ'ın "Problem çözme becerilerinin değerlendirilmesinde öz, akran ve öğretmen puanlarının çok yüzeyle rasch ölçme modeli ile incelenmesi" başlıklı tezi kapsamında çalışma yapma talebi Müdürlüğümüzce uygun görülmüş ve araştırmanın yapılacağı İlçe Milli Eğitim Müdürlüğüne bilgi verilmiştir.

Anket formunun (7 sayfa) araştırmacı tarafından uygulama yapılacak sayıda çoğaltılması ve çalışmanın bitiminde iki örneğinin (cd ortamında) Müdürlüğümüz Strateji Geliştirme (1) Şubesine gönderilmesini arz ederim.

Müberra OĞUZ
Müdür a.
Şube Müdürü

Güvenli Elektronik İmza
Aslı ile Aynıdır.

28.1.01.../2015...

SUBAŞI
Şef

Atatürk Blv. 06648 Kızılay/ANKARA
Elektronik Ağ: www.meb.gov.tr
e-posta: adsoyad@meb.gov.tr

Ayrıntılı bilgi için: Ad SOYAD Ünvan
Tel: (0 312) XXX XX XX
Faks: (0312) XXX XX XX

Bu evrak güvenli elektronik imza ile imzalanmıştır. <http://evraksorgu.meb.gov.tr> adresinden ac2d-36f6-3b6e-9f9b-3f9b kodu ile teyit edilebilir