



**ÇOCUKLAR İÇİN ANİMASYONLU BİLİMSEL YARATICILIK TESTİ KANIT
DEĞERLENDİRME BİLEŞENİNİN GELİŞTİRİLMESİ VE PSİKOMETRİK
NİTELİKLERİNİN İNCELENMESİ**

Doktora Tezi

Mehmet Arda ATAKAYA

Eskişehir 2025

**ÇOCUKLAR İÇİN ANİMASYONLU BİLİMSEL YARATICILIK TESTİ KANIT
DEĞERLENDİRME BİLEŞENİNİN GELİŞTİRİLMESİ VE PSİKOMETRİK
NİTELİKLERİNİN İNCELENMESİ**

Mehmet Arda ATAKAYA

DOKTORA TEZİ

Özel Yetenekliler Eğitimi Programı / Özel Eğitim Anabilim Dalı

Danışman: Prof. Dr. Uğur SAK

Eskişehir

Anadolu Üniversitesi

Lisansüstü Eğitim Enstitüsü

Haziran 2025

Bu tez çalışması BAP Komisyonunca kabul edilen 2206E155 nolu proje kapsamında desteklenmiştir.

JÜRİ VE ENSTİTÜ ONAYI

Mehmet Arda ATAKEYA'nın "Çocuklar İçin Animasyonlu Bilimsel Yaratıcılık Testi Kanıt Değerlendirme Bileşeninin Geliştirilmesi ve Psikometrik Niteliklerinin İncelenmesi" başlıklı tezi 30/05/2025 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca Özel Eğitim Anabilim Dalı Özel Yetenekliler Eğitimi Programında Doktora tezi olarak kabul edilmiştir.

	Ünvanı Adı Soyadı	İmza
Üye (Danışman)	: Prof. Dr. Uğur SAK	
Üye	: Prof. Dr. Mustafa Serdar KÖKSAL	
Üye	: Doç. Dr. Murat Doğan ŞAHİN	
Üye	: Doç. Dr. Nazmiye Nazlı Ateşgöz	
Üye	: Doç. Dr. Esra KANLI DENİZCİ	

ÖZET

ÇOCUKLAR İÇİN ANİMASYONLU BİLİMSEL YARATICILIK TESTİ KANIT DEĞERLENDİRME BİLEŞENİNİN GELİŞTİRİLMESİ VE PSİKOMETRİK NİTELİKLERİNİN İNCELENMESİ

Mehmet Arda ATAKAYA

Özel Eğitim Anabilim Dalı

Özel Yetenekliler Eğitimi Programı

Anadolu Üniversitesi, Lisansüstü Eğitim Enstitüsü, Haziran 2025

Danışman: Prof. Dr. Uğur SAK

Çalışmada anasınıfı, birinci ve ikinci sınıf öğrencilerinin bilimsel yaratıcılığını ölçmek için geliştirilmiş olan Çocuklar için Animasyonlu Bilimsel Yaratıcılık Testi' ne (ÇABİYAT) kanıt değerlendirme bileşeni eklenmiş ve revize edilen testin (ÇABİYAT-R) psikometrik özellikleri incelenmiştir. ÇABİYAT-R hipotez geliştirme, deney tasarlama ve kanıt değerlendirme alt testlerinden oluşmakta ve her bir alt testte 3'er madde olmak üzere toplam 9 madde yer almaktadır. 252 kişinin katıldığı araştırmada, yapı geçerliği kapsamında gerçekleştirilen EGA sonuçlarında ön görülen üç faktörlü yapı elde edilmiştir. AYEM ve DFA sonucunda ÇABİYAT-R'nin üç faktörlü yapısı doğrulanmış, iki yaklaşım için de mükemmel uyum indeksleri elde edilmiştir. AVE değerlerinin .518 ile .548; CR değerlerinin .756 ile .784 ve $\sqrt{AVE}$ değerlerinin ise .72 ile .74 arasında değiştikleri görülmüştür. Sonuçlar ÇABİYAT-R'nin yapı geçerliğinin iyi düzeyde sağlandığını göstermiştir. ÇABİYAT-R alt test ve toplam test puanlarının üst sınıflar lehine anlamlı şekilde daha yüksek oldukları görülmüştür. SAGES 2 bataryası ile ÇABİYAT-R testi arasında .50 üzerinde anlamlı korelasyon katsayıları hesaplanmış ve testin uyum geçerliğinin iyi düzeyde olduğu bulunmuştur. Madde yanlılık analizleri ölçme değişmezliği modelleri ile test edilmiş ve ÇABİYAT-R maddelerinin kız ya da erkekler lehine herhangi bir yanlılık barındırmadığı tespit edilmiştir. Güvenirlik analizleri kapsamında hesaplanan Cronbach Alpha ve McDonald's Omega katsayılarının .765 ile .784 arasında değiştikleri, kodlayıcılar arası güvenirliliğin ise $r_{ICC}=.998$ mükemmel düzeyde olduğu görülmüştür.

Anahtar Sözcükler: Bilimsel yaratıcılık, TOSCAC, ÇABİYAT, kanıt değerlendirme, açımlayıcı yapısal eşitlik modellemesi, açımlayıcı grafik analizi

ABSTRACT
**DEVELOPMENT OF THE EVIDENCE EVALUATION COMPONENT FOR
THE ANIMATED SCIENTIFIC CREATIVITY TEST FOR CHILDREN AND
EXAMINATION OF ITS PSYCHOMETRIC PROPERTIES**

Mehmet Arda ATAKAYA

Department of Special Education

Gifted Education Program

Anadolu University, Graduate School, June 2025

Advisor: Prof. Dr. Uğur SAK

In the study, animated scientific creativity test for K2 students was revised by adding an evidence evaluation component and the psychometric properties of the revised test (TOSCAC-R) were examined. TOSCAC-R consists of three subtests: hypothesis development, experiment design, and evidence evaluation, with each subtest containing three items, resulting in a total of nine items. In the study, which included 252 participants, an Exploratory Graph Analysis (EGA) was conducted to assess construct validity, and the expected three-factor structure was obtained. Both Exploratory Structural Equation Modeling (ESEM) and Confirmatory Factor Analysis (CFA) confirmed the three-factor structure of TOSCAC-R, with excellent fit indices for both approaches. The Average Variance Extracted (AVE) values ranged from .518 to .548, Composite Reliability (CR) values ranged from .756 to .784, and $\sqrt{\text{AVE}}$ values ranged from .72 to .74. These results demonstrated that TOSCAC-R has a good level of construct validity. It was observed that subtest and total test scores of TOSCAC-R were significantly higher in favor of upper-grade students. Significant correlation coefficients above .50 were found between the SAGES-2 battery and TOSCAC-R, indicating a good level of criterion validity. Item bias analyses were conducted using measurement invariance models to examine gender differences, and it was determined that TOSCAC-R items did not exhibit bias in favor of either boys or girls. As part of the reliability analyses, Cronbach's Alpha and McDonald's Omega coefficients ranged between .765 and .784, while interscorer reliability was $r_{\text{ICC}} = .998$, indicating excellent reliability.

Keywords: Scientific creativity, TOSCAC, ÇABİYAT, evidence evaluation, exploratory structural equation modelling, exploratory graph analysis

TEŞEKKÜR

Danışmanım, değerli Hocam Prof. Dr Uğur Sak’a akademik rehberliği, süreç içerisindeki yapıcı eleştirileri, bilgi ve deneyimiyle bana yol gösterdiği için içten teşekkür ederim. Tez izleme komitelerinde istatistiksel analiz sürecine getirdiği özgün yaklaşımlar ve önerdiği farklı yöntemler ile çok büyük katkılar sağlayan değerli Hocam Doç. Dr. Murat Doğan Şahin’e gönülden teşekkürlerimi sunarım. Çalışmamın başından itibaren desteğini hiç esirgemeyen, tez izleme komitelerindeki katkılarından dolayı Doç. Dr. N. Nazlı Ateşgöz’e teşekkürü bir borç biliyorum. Yoğunluklarına rağmen savunma komitemde yer alan ve çalışmama anlamlı katkılar sağlayan Prof. Dr. Mustafa Serdar Köksal ve Doç. Dr. Esra Kanlı Denizci’ye teşekkürlerimi sunarım.

Yüksek lisans sürecimden bugüne değin akademik kariyerime çok büyük katkılar sağlayan, özgün ve eleştirel fikirleriyle çalışmama destek olan Doç. Dr. M. Bahadır Ayas’a içten teşekkür ederim. Bana bu süreçte moral ve motivasyon kaynağı olan Hocam Doç. Dr. Bilge Bal Sezerel, oda arkadaşım Arş. Gör. Dr. Selin Bozbey Esmeroğlu, ile sevgili arkadaşım Arş. Gör. Ferhat Köprü’ye, madde hazırlık ve seçim sürecinde ÜYEP öğretmenlerinden Hakan Duman ve Osman Gürel’e, uygulamalar süresince desteklerini esirgemeyen Serkan Bayar’a, beni günlerce sabırla dinleyip bıkmadan usanmadan işini en iyi şekilde yapan değerli kardeşim Kadir Gökçek’e sonsuz teşekkür ederim.

Hayatımın her noktasında bana inanıp güvenen, varlığını daima hissettiren ve bugünlere gelmemde çok büyük katkısı olan ne yazık ki sürecin tamamlandığını göremeden aramızdan ayrılan kıymetli babama büyük bir minnet borçluyum. Her düştüğümde beni ayağa kaldıran, hüznüm ile hüzünlenip, sevincim ile mutlu olan sevgili anam, ablam ve yeğenim Naz’a gönülden teşekkür ederim. Hayatıma girerek bana umut ve yaşam kaynağı olan ve büyük hayaller paylaştığım sevgilim Deniz’in bu sürecin bir parçası olmasından duyduğum büyük mutluluğu burada ifade etmek isterim. Yaşamının 7 yılını devirmiş köpeğim Odin’in çok uzun yıllar sağlıklı ve yanımızda kalması dileği ile.

Mehmet Arda ATA KAYA

Eskişehir, 2025

Tarih:

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Anadolu Üniversitesi tarafından kullanılan “bilimsel intihal tespit programı”yla tarandığını ve hiçbir şekilde “intihal içermediğini” beyan ederim.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara razı olduğumu bildiririm.

İmza

Mehmet Arda ATAKEYA

ÜRETKEN YAPAY ZEKÂ KULLANIM BEYANI

Bu tezi hazırlarken verilerin analizinde ChatGPT üretken yapay zekâ programından destek aldığımı beyan ederim. Üretken yapay zekâ programlarından aldığım bilgilerin doğruluğunu kontrol ettiğimi bildiririm. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

MEHMET ARDA ATAĞAYA

İÇİNDEKİLER

	<u>Sayfa</u>
1. GİRİŞ	1
1.1. Problem	1
1.2. Araştırmanın Amacı	6
1.3. Araştırmanın Önemi	7
1.4 Varsayımlar	7
1.5 Sınırlılıklar	8
1.6 Tanımlar	8
2 ALANYAZIN	9
2.1 Yaratıcılık	9
2.1.1 Zihin yapı kuramı modeli	10
2.1.2 Bileşensel yaratıcılık modeli	12
2.1.3 Sistemler modeli	13
2.2 Bilimsel Yaratıcılık	14
2.2.1 Bilimsel yaratıcılığın ikili arama modeli	16
2.2.1.1 Hipotez geliştirme ile ilgili araştırmalar	19
2.2.1.2 Deney tasarlama ile ilgili araştırmalar	21
2.2.1.3 Kanıt değerlendirme ile ilgili araştırmalar	24
2.3 Bilimsel Yaratıcılığın Gelişimi	30
2.3.1 Bilgi	30
2.3.2 Yaş	32
2.3.3 Sınıf	34
2.4 Bilimsel Yaratıcılık ve Cinsiyet	36
2.5 Bilimsel Yaratıcılığın Ölçülmesi	38
2.5.1 Öncül bilimsel yaratıcılık testleri	39
2.5.2 Kuramsal temelli bilimsel yaratıcılık testleri	40

2.5.2.1 Ortaokullar için bilimsel yaratıcılık testi.....	42
2.5.2.2 Beşinci sınıflar için bilimsel yaratıcılık testi.....	45
2.5.2.3 Okul öncesi öğrenciler için görsel bilimsel yaratıcılık testi -FSCT ...	46
2.5.2.4 Kimya alanında geliştirilmiş bilimsel yaratıcılık testi- CSCT	47
2.5.2.5 Yaratıcı bilimsel çağrışımlar testi (Y-BÇT)	49
2.5.2.6 Çoğul düşünmeye dayalı problem çözme yetenek testi- DPAS.....	51
2.5.2.7 Bilimsel üretkenlik testi (BÜT-6-8).....	53
2.5.2.8 Bilimsel üretkenlik testi (BÜT 3-5).....	55
2.5.2.9 Çocuklar için animasyonlu bilimsel yaratıcılık testi- ÇABİYAT	58
3 YÖNTEM	63
3.1 Araştırma Modeli	63
3.2 Çalışma Grubu	63
3.3 Veri Toplama Araçları	64
3.3.1 ÇABİYAT-R	64
3.3.1.1 ÇABİYAT-R'nin içeriği	64
3.3.1.2 ÇABİYAT-R'nin uygulama süreci.....	67
3.3.1.3 ÇABİYAT-R puanlama yöntemi.....	69
3.3.1.4 ÇABİYAT-R' nin revize edilme süreci.....	71
3.3.1.4.1 ÇABİYAT-R' nin teorik altyapısı.....	72
3.3.1.4.2 Test madde havuzunun oluşturulması.....	74
3.3.1.4.3 Havuzdan madde seçimi	77
3.3.1.4.4 Animasyonların oluşturulması.....	79
3.3.1.4.5 Ön değerlendirmenin uygulanması	81
3.4 Veri Toplama Süreci.....	84
3.5 Verilerin Analizi ve Yorumlanması.....	85
3.5.1 Madde analizleri	86
3.5.2 Geçerlik analizleri	86

3.5.2.1 Yapı geçerliği	87
3.5.2.1.1 Açımlayıcı grafik analizi- EGA	87
3.5.2.1.2 Açımlayıcı yapısal eşitlik modellemesi	92
3.5.2.1.3 Yakınsama ve Iraksama Geçerlikleri	95
3.5.2.2 Ölçüt geçerliği	96
3.5.2.3 Uyum geçerliği.....	96
3.5.2.3 Bilinen grup geçerliği.....	96
3.5.2.4 Maddelerin cinsiyet bağlamında yanlılık analizi	97
3.5.3 Güvenirlik analizleri	98
4 BULGULAR.....	100
4.1 Madde Analizlerine İlişkin Bulgular	100
4.2 ÇABİYAT-R' nin Geçerliği.....	109
4.2.1 ÇABİYAT-R' nin yapı geçerliği.....	109
4.2.1.1 Açımlayıcı grafik analizi bulguları.....	109
4.2.1.2 Açımlayıcı yapısal eşitlik modellemesi bulguları.....	114
4.2.1.3 Yakınsama ve ıraksama geçerliklerine ilişkin bulgular.....	118
4.2.2 Ölçüt geçerliğine ilişkin bulgular	120
4.2.2.1 Uyum geçerliğine ilişkin bulgular	120
4.2.2.2 Bilinen grup geçerliğine ilişkin bulgular	121
4.2.3 Cinsiyete göre ölçme değişmezliği analizi bulguları.....	122
4.3 ÇABİYAT-R'nin Güvenirliği.....	124
4.3.1 ÇABİYAT-R'nin iç tutarlılığına ilişkin bulgular	124
4.3.2 Kodlayıcılar arası güvenirlik bulguları.....	127
5 SONUÇ, TARTIŞMA ve ÖNERİLER.....	129
5.1 ÇABİYAT-R'nin Madde Analizlerine İlişkin Tartışma ve Sonuç	129
5.2 ÇABİYAT-R'nin Geçerliğine İlişkin Tartışma ve Sonuç.....	131
5.2.1 ÇABİYAT-R'nin yapı geçerliğine ilişkin tartışma ve sonuç	131

5.2.1.1 ÇABİYAT-R'nin açıklayıcı grafik analizine (EGA) ilişkin tartışma ve sonuç.....	132
5.2.1.2 ÇABİYAT-R' nin doğrulayıcı faktör (DFA) ve açıklayıcı yapısal eşitlik modellemesi (AYEM) analizlerine ilişkin tartışma ve sonuç	134
5.2.1.3 Yakınsama ve ıraksama geçerliklerine ilişkin tartışma ve sonuç	138
5.2.2 ÇABİYAT-R'nin Ölçüt Geçerliğine İlişkin Tartışma ve Sonuç.....	139
5.2.2.1 ÇABİYAT-R'nin uyum geçerliğine ilişkin tartışma ve sonuç	140
5.2.2.2 ÇABİYAT-R'nin bilinen gruplar geçerliğine ilişkin tartışma ve sonuç	141
5.2.3 ÇABİYAT-R'nin Cinsiyete Göre Ölçme Değişmezliğine İlişkin Tartışma ve Sonuç	142
5.3 ÇABİYAT-R'nin Güvenirliğine İlişkin Tartışma ve Sonuç.....	145
5.3.1 ÇABİYAT-R'nin iç tutarlılığına ilişkin tartışma ve sonuç.....	145
5.3.2 Kodlayıcılar arası güvenirlige ilişkin tartışma ve sonuç.....	147
5.4 Öneriler	148
5.4.1 Eğitim uygulamalarına yönelik öneriler	148
5.4.2 İleriki araştırmalara yönelik öneriler.....	148
KAYNAKÇA	150
EKLER	
ÖZGEÇMİŞ	

TABLÖLAR DİZİNİ

Sayfa

Tablo 3.1. Katılımcıların cinsiyet ve sınıf düzeylerine göre dağılımları.....	61
Tablo 3.2. ÇABİYAT-R maddelerine ilişkin bilgiler	63
Tablo 3.3. Maddelerin animasyon süreleri	66
Tablo 3.4. ÇABİYAT-R madde havuzu oluşturulma süreci	73
Tablo 3.5. Anaokulu öğretmenleri madde değerlendirme kriteri	75
Tablo 3.6. Madde seçim süreci ve disiplinlere göre dağılımları	76
Tablo 4.1. Maddelere ilişkin betimsel istatistik bulguları	97
Tablo 4.2. Sınıf düzeylerine göre kanıt değerlendirme maddelerinin ayırt edicilikleri .	99
Tablo 4.3. Basit doğrusal regresyon analizi sonuçları.....	100
Tablo 4.4. Cinsiyete göre kanıt değerlendirme madde ortalamaları ve anlamlılık değerleri	101
Tablo 4.5. Maddeler arası korelasyon katsayı sonuçları	103
Tablo 4.6. Madde güvenirlik analizi sonuçları.....	104
Tablo 4.7. Kanıt değerlendirme maddelerinin seçimine ilişkin alınan kararlar	105
Tablo 4.8. ÇABİYAT-R EGA modeli kenar matris ağırlıkları (w)	109
Tablo 4.9. BootEGA betimsel istatistik sonuçları	109
Tablo 4.10. BootEGA analizi faktör sayısı frekans dağılımları	110
Tablo 4.11. BootEGA analizi faktörler arası madde kararlılık katsayı sonuçları	110
Tablo 4.12. Uyum kabul ölçütleri ve model uyum indeks sonuçları.....	112
Tablo 4.13. Üç faktörlü DFA ve AYEM modelleri faktör yükleri, açıklanan ve artık varyans değerleri.....	113
Tablo 4.14. Modellere ilişkin AVE ve CR katsayı sonuçları.....	115
Tablo 4.15. ÇABİYAT-R ve SAGES 2 uyum geçerliği katsayı sonuçları.....	117
Tablo 4.16. Alt ölçek ve toplam test puanlarına göre sınıf düzeyinde yapılan ANOVA analizi sonuçları.....	118

Tablo 4.17. Ölçme değişmezliği modellerine ilişkin uyum indeksleri.....	119
Tablo 4.18. Kız ve erkek öğrenciler faktör ortalamaları karşılaştırmaları.....	120
Tablo 4.19. ÇABİYAT-R güvenirlik analizleri sonuçları.....	121
Tablo 4.20. ÇABİYAT-R korelasyon katsayı sonuçları	122
Tablo 4.21. ÇABİYAT-R'nin kodlayıcılar arası güvenirlik sonuçları.....	124



ŞEKİLLER DİZİNİ

Sayfa

Şekil 2.1. Zihin Yapı Kuramı Modeli	11
Şekil 2.2. Sistemler Modeli	13
Şekil 2.3. SDSS Modeli ve Temel Bileşenleri	17
Şekil 2.4. Bilimsel Keşfin İkili Arama Modeli Bileşenleri.....	18
Şekil 3.1. Terleme Maddesi Animasyon Örneği.....	65
Şekil 3.2. ÇABİYAT-R'nin Uygulama Düzeni.....	66
Şekil 3.3. ÇABİYAT-R'nin Teorik Alt Yapısı	71
Şekil 3.4. ÇABİYAT' ın Teorik Alt Yapısı	71
Şekil 3.5. Örnek Bir EGA Grafiği	86
Şekil 3.6. AYEM analizi için Karar Ağacı	91
Şekil 4.1. ÇABİYAT-R Testi EGA Modeli Grafiği	108
Şekil 4.2. BootEGA Analizi Deneysel Olarak Boyutlarda Yer Alan Maddelerin Kararlılık Katsayıları.....	111
Şekil 4.3. ÇABİYAT-R DFA Modeli Teorik Alt Yapısı	114
Şekil 4.4. ÇABİYAT-R AYEM Modeli Teorik Alt Yapısı	115

KISALTMALAR DİZİNİ

AFA	:	Açımlayıcı Faktör Analizi
AVE	:	Ortalama Açıklanan Varyans
AYEM	:	Açımlayıcı Yapısal Eşitlik Modellemesi
BootEGA	:	Bootstrap Açımlayıcı Grafik Analizi
BÜT	:	Bilimsel Üretkenlik Testi
CR	:	Bileşik Güvenirlik
CSCT	:	Kimya Alanında Geliştirilmiş Bilimsel Yaratıcılık Testi
ÇABİYAT	:	Çocuklar için Animasyonlu Bilimsel Yaratıcılık Testi
ÇABİYAT-R	:	Çocuklar için Animasyonlu Bilimsel Yaratıcılık Testi- Revize
DFA	:	Doğrulayıcı Faktör Analizi
DPAS	:	Çoğul Düşünmeye Dayalı Problem Çözme Yetenek Testi
DT	:	Deney Tasarlama
EGA	:	Açımlayıcı Grafik Analizi
FSCT	:	Okul Öncesi Öğrenciler İçin Görsel Bilimsel Yaratıcılık Testi
HG	:	Hipotez Geliştirme
KD	:	Kanıt Değerlendirme
MSV	:	En Yüksek Ortak Varyans
SAGES-2	:	Üstün Yetenekli İlkokul ve Ortaokul Öğrencileri için Tarama Değerlendirme Bataryasının İkinci Baskısı
SDDS	:	Bilimsel Yaratıcılığın İkili Arama Modeli
SSCM	:	Bilimsel Yaratıcılığın Yapı Modeli
TTCT	:	Torrance Yaratıcı Düşünme Testi
YBÇT	:	Yaratıcı Bilimsel Çağrışımlar Testi

1 GİRİŞ

1.1 Problem

Yaratıcılık alanının tartışılan konularının başında yaratıcılıkta alana özgü-genel yaratıcılık ayrımı (Baer, 1998; Kaufman ve Baer, 2004; Plucker, 1998) ve yaratıcılığın ölçülmesinde kriter belirleme probleminin geldiği görülmektedir (Paek ve Runco, 2017; Runco, 1999; Shapiro, 1970). Kimi araştırmacılar bilişsel özelliklere, kişilik özelliklerine, tutum ve davranışlara atıfta bulunarak yaratıcılığı genel teoriler ile açıklarken (Guilford, 1967; Rhodes, 1961); kimileri belirli bir disiplinde ürün, performans ya da fikir ortaya koyulan alana özgü teoriler ile açıklamaya çalışmışlardır (Amabile 1983b, 2011; Csikszentmihalyi, 2014).

Yaratıcılıkta hem alana özgü becerilere hem de genel yaratıcılık becerilerine gereksinim duyulan disiplinlerden biri olarak bilimsel yaratıcılık gösterilmektedir (Heller, 2007; Hu ve Adey, 2002). Çok sayıda fikir üretme, problemlere farklı çözüm önerileri getirme genel yaratıcılık becerileri ile (Guilford, 1950;1967; Torrance, 1972); bilimsel konularda hipotez kurma, deney tasarlama ve kanıt değerlendirme (Klahr, 2000; Klahr ve Dunbar,1988); verilerden kuramlar geliştirme, buluş ortaya koyma (Kulkarni & Simon, 1988) ise bilim alanına özgü beceriler ile ilişkilendirilmektedir.

Bilim insanlarının yaratma süreçleri, farklı metodolojik yöntemler ile ortaya konulmaya çalışılmıştır. Kimi araştırmacılar doğrudan bilim insanları ile aynı laboratuvarları paylaşarak bu süreci tanımlamaya çalışırken (Dunbar, 1993; Jang ve Ko, 2017; Klahr, 2000; Klahr vd., 1993; Kuhn, 1989; Kuhn vd., 1988); kimi araştırmacılar bilim insanlarının otobiyografilerini kullanarak bilimsel keşiflerin ortaya çıkış süreçlerini anlamaya çalışmışlardır (Rocke, 2010; Rothenberg, 1988). Bilimsel keşif süreci, bilimsel yaratıcılıkla ilişkilendirilerek özellikle 1950'lerden itibaren bilimsel yaratıcılık testleri geliştirilmeye başlanmıştır. İlk yaratıcılık testi örneklerinden birini geliştiren Majumdar (1975) yaratıcı yeteneği fizik, biyoloji ve matematikte geliştirdiği görevler ile ölçmeye çalışmıştır. Sinha ve Singh (1987) orta okul öğrencileri için İngilizce ve Hinduca geliştirdikleri testte esneklik, özgünlük, hayal gücü, gözlem yeteneği, analiz edebilme becerisi ve dönüştürebilme yeteneklerine yer vermişlerdir. Sukla (1980), bilimsel yaratıcılık testinde akıcılık, esneklik ve evrensel bilimsel yaratıcılığı ölçmeye çalışmaktadır. Friedlander (1983),

lise öğrencileri için veri toplama, problem çözme, hipotez kurma ve deney planlama becerilerine yönelik çoğul düşünme testi geliştirmiştir.

Belirtilen yaratıcılık testlerinin belirli bir kuramsal çerçeveye sahip olmamaları, geçerlik ve güvenirlik çalışmalarının yetersiz olmaları sebebiyle eleştirildiği görülmektedir (Hu ve Adey, 2002; Liang, 2002). Ayrıca testlerin genellikle büyük yaş gruplarına yönelik olarak geliştirilmesi bir diğer sınırlılık olarak ifade edilmektedir (Ateşgöz ve Sak, 2021). Söz konusu sınırlılıkları gidermek amacıyla son dönem bilimsel yaratıcılık testlerinin, yaratıcılık kuramlarına ve bilimsel düşünme becerilerinin mekanik işleyişine odaklandığı görülmektedir. Hu ve Adey (2002), genel yaratıcılık kuramı ile ilişkilendirilen Guilford'un (1967) Zihnin Yapı Kuramından beslenerek Bilimsel Yaratıcılığın Yapı Modeli (SSCM) ile kavramsal bir model ortaya koymuşlardır. SSCM modeli temel alınarak ön lisans (Ramly vd., 2022), lise (Hu ve Adey, 2002), beşinci sınıf (Siew, Chong & Chin, 2014) ve okul öncesi öğrenciler (Chin & Siew, 2015) için bilimsel yaratıcılık testleri geliştirilmiştir. Fen bilimleri alanında çağrışıma ve analogik düşünmeye yeterince önem verilmediği gerekçesiyle, Mednick'in (1962) Çağrışımsal Düşünme Kuramı temel alınarak Yaratıcı Bilimsel Çağrışımsal Testi (YBÇ-T) geliştirilmiştir (Kanlı, 2014).

Bilimsel keşfi, hipotez ve deney şeklinde iki temel alan ile açıklamaya çalışan kimi araştırmacılar ise bilimsel yaratıcılığı ampirik olarak test etmeye çalışmışlardır (Klahr, 2000; Klahr ve Dunbar, 1988; 1993). Bilimsel Yaratıcılığın İkili Arama Modeli adı verilen kuram (SDDS), hipotez ve deney alanı olmak üzere iki temel alandan; hipotez kurma, deney tasarlama ve kanıt değerlendirme olarak üç bileşenden oluşmaktadır. Kurama göre bilimsel keşif problem çözme ile başlar. Problemin çözülebilmesi için hipotezlerin oluşturulması, oluşturulan hipotezlerin deneyler tasarlanarak test edilmesi gerekir. Deneylerden elde edilen verilerle öne sürülen hipotezin karşılaştırılması için ise kanıt değerlendirme sürecinden yararlanılmaktadır. Araştırmacılara göre bilimsel keşif söz konusu 3 bileşenin birbirleri ile etkileşimleri sonucunda ortaya çıkmaktadır. SDDS modeli ile alana özgü yaratıcılığı bir araya getiren Ayas ve Sak (2014), Ayas (2017), Ateşgöz ve Sak (2021), Bermejo vd. (2016) bilimsel yaratıcılığı ilkökul, ortaokul ve okul öncesi gruplarında test etmişlerdir.

SDDS modeli (Klahr, 2000) ve çoğul düşünmeye dayalı çalışmalar (Guilford, 1950; Torrance, 1972) temel alınarak Ayas ve Sak (2014) 6-8. sınıfa yönelik kâğıt kalem

testi; Ayas (2017) 3-5. sınıflara yönelik kâğıt kalem testi; Ateşgöz ve Sak (2021) anasınıfı, birinci ve ikinci sınıflar için animasyonlu bilimsel yaratıcılık testini geliştirmişlerdir. Ateşgöz ve Sak (2021) tarafından geliştirilen Çocuklar İçin Animasyonlu Bilimsel Yaratıcılık Testi (ÇABİYAT), hipotez geliştirme ve deney tasarlama alt ölçeklerinden oluşan ve tablette animasyonlarla uygulanan bir testtir. Testte öğrencilerin her bir maddeye ilişkin akıcılık, esneklik, orijinallik ve yaratıcılık bölümü puanları hesaplanmaktadır. ÇABİYAT' ın test maddeleri senaryolaştırılmış ve test küçük yaş grubu öğrencilerinin ilgisini çekecek şekilde animasyon haline getirilerek uygulanmıştır. ÇABİYAT' ın geliştirilme sürecinde belirli bir kuramsal alt yapının benimsenmesi ve ölçek geliştirme basamaklarının alanyazına paralel bir şekilde uygulanması testi hem kavramsal hem de yöntemsel olarak güçlü kılmaktadır. Ayrıca ÇABİYAT' ın küçük yaş gruplarına yönelik animasyon tabanlı geliştirilmiş olması testin bir diğer güçlü yönü olarak yorumlanabilir. Fizik, kimya ve biyoloji alanlarında hipotez geliştirme ve deney tasarlamaya ilişkin oluşturulan maddelerin ve ölçeğin bütününe ilişkin geçerlik ve güvenirlik çalışmalarının ise tatmin edici seviyede olduğu rapor edilmiştir (Ateşgöz ve Sak, 2021).

SDDS modeli temel alınarak geliştirilmiş bilimsel yaratıcılık testlerinde kuramsal olarak öngörülen 3 boyutlu yapının tam anlamıyla ortaya çıkarılamadığı söylenebilir. Geliştirilen bilimsel yaratıcılık testlerinde tek boyutlu (Ayas ve Sak, 2014; Sak ve Ayas, 2013) ve iki boyutlu yapılar elde edilmiştir (Ateşgöz ve Sak, 2021; Ayas, 2017). Araştırmacılar kanıt değerlendirme bileşeninin AFA ve DFA ile yapı geçerliğinin sağlanamadığını ve modelde revizyona giderek kanıt değerlendirme becerisini kuramsal çerçeveden çıkardıklarını rapor etmişlerdir (Ateşgöz ve Sak, 2021; Ayas, 2017). Söz konusu araştırmalarda kanıt değerlendirme bileşeninin doğrulanamamasının nedenleri arasında küçük yaş grupları için kanıt değerlendirme becerisinin ölçümüne yönelik sınırlılıklar gösterilmektedir (Amsel ve Brock, 1996; Bullock, 1991; Kuhn, vd., 1988; Schauble, 1996; Sodian vd., 1991). Ayrıca kanıt değerlendirme bileşeninin hipotez kurma ile deney tasarlama arasında aracı bir bilimsel düşünme becerisi olması sebebiyle geliştirilen maddelerde binişiklik sorunu olduğu rapor edilmektedir (Ateşgöz ve Sak, 2021; Ayas, 2017); Dolayısıyla araştırmacılara göre kanıt değerlendirme becerisi, hipotez geliştirme ve deney tasarlama görevlerinde örtük olarak işlev görmektedir. Yine araştırmacılar kanıt değerlendirmenin dikotomik

yanıtlara dayanmasından dolayı çoğul düşünmeye dayalı bilimsel yaratıcılık testlerinde kullanılamayacağını önermişlerdir (Ateşgöz ve Sak, 2021; Ayas, 2017).

Esasen kanıt değerlendirme becerisinin küçük yaş gruplarında farklı metodolojik yöntemlerle test edilebildiği literatürde görülmektedir. Kanıt değerlendirme becerisi; kanıt örüntülerinin değerlendirilmesi, (Amsel ve Brock, 1996; Kuhn, 2005; 2011; Kuhn vd., 1988), verilerin analiz edilmesi ve yorumlanması (Masnick vd., 2007; Masnic ve Morris, 2008) ve açıklamaların oluşturulması (Kuhn, 1989, 2005) başlıkları altında toplanmaktadır (Klahr vd., 2019).

Kanıt örüntülerinin değerlendirilmesi kanıtların hipotezle mükemmel uyumu, kısmen uyumu ya da uyumsuzluğu üzerinden görseller kullanılarak değerlendirilmesine dayanmaktadır. Küçük yaş gruplarında özellikle mükemmel uyumun işe koşulduğu görevlerde öğrencilerin başarılı oldukları (Ruffman vd., 1993), kısmi uyum ile uyumsuzluğun olduğu durumlarda ise öğrencilerin görevi anlamada zorlandıkları ya da başarısız oldukları görülmüştür (Koerber vd., 2005; Zimmerman ve Croker, 2013). Ruffman vd. (1993), 4 ile 7 yaş arasındaki çocuklar ile tek bir potansiyel nedenin dahil edildiği basit bir görev sunarak kanıt değerlendirme becerisinin varlığını araştırmışlardır. Çalışmada 4 yaş düzeyindeki çocukların bile kendilerine sunulan kanıt örüntüleri arasında nedensel muhakeme kurabildikleri görülmüştür. Piekny vd. (2013), Koerber ve arkadaşlarının (2005) çalışmalarında kullandıkları görevi kendi araştırmalarında küçük değişiklikler yaparak tekrar kullanmışlardır. Görevde çocuklara kukla yolu ile *diş sağlığı* ve *sakız rengi* arasındaki nedensel ilişkiye dayalı inançları sorulmuştur. 138 okul öncesi çocukla yapılan çalışmada kanıt değerlendirme becerisinin 4 yaş döneminde başladığı, yaş ve görevlerdeki uyum düzeyi arttıkça puanlarda artış meydana geldiğini rapor edilmektedir.

Kanıt değerlendirmede kullanılan bir diğer görev tipi verilerin *analiz edilmesi ve yorumlanmasına* dayanmaktadır. Bu görev tipinde katılımcılara veri setleri sunularak olası neden ve sonuçların neler olabileceği sorulmaktadır (Masnic ve Morris, 2008). Söz konusu görev tipinde genellikle büyük yaş grupları ile çalışıldığı görülmüştür (Masnic vd., 2016; Penner ve Klahr, 1996). *Açıklamalar oluşturma* görev tipinde ise katılımcılardan hipotezle ya da modelle uyumlu deney sonuçlarına ilişkin açıklamalar oluşturmaları beklenmektedir. Araştırmalar küçük yaş gruplarındaki çocukların, neden-sonuç ilişkisine dair temel bir anlayışa sahip olduklarını

göstermektedir (Bullock ve Gelman, 1979; Piaget, 1929). Fakat bilimsel düşünme ile ilişkilendirilen çalışmalarda hem çocukların hem de yetişkinlerin mevcut öncül inanışlarını korudukları ve kanıta dayalı olarak düşüncelerini değiştirmedikleri görülmüştür (Chinn ve Brewer, 1998; Chinn ve Malhotra, 2002).

Kanıt değerlendirme görev tipleri genel olarak değerlendirildiğinde görevlerde yaş düzeyi arttıkça daha tutarlı sonuçlar alındığı, kanıt değerlendirme becerisinin küçük yaş gruplarında ancak mükemmel uyum içeren, neden sonuç ilişkileri bağlamında değerlendirilebileceği görülmektedir (Chen ve Klahr, 1999; Croker ve Buchanan 2011; Kuhn, 1989; Zimmerman, 2007). Verilerin analiz edilmesi, yorumlanması ve açıklamaların oluşturulmasına dayalı kanıt değerlendirme görev tiplerinin ise büyük yaş gruplarına daha uygun olduğu belirtilmektedir (Kuhn vd., 1988). Kanıt değerlendirme üzerine yapılan çalışmalarda görevlerin öne sürülen hipotezi onaylama ya da reddetme amacına dayalı olarak tasarlanması, katılımcıların görevleri tamamlama sürecinde şans faktörünün etkili olabileceğini düşündürmektedir. Ayrıca bazı araştırmalarda görevlerin kovaryasyon temelli basit neden-sonuç ilişkisine dayandırılması, bireylerin önceden edindikleri bilgi ve rasyonel olma düşüncelerini bastırdığı gerekçesiyle eleştirilmiştir (Koslowski, 1996). Yapılan detaylı okumalarda kanıt değerlendirmenin kanıt üretmek çok sayıda fikir üretilabileceği bir başka görev tipi ile gerçekleştirildiği görülmüştür (Kuhn vd., 1988). *Kanıt üretme* görev tipinde katılımcılara bir hipotez ve deney sonuçları sunulmaktadır. Katılımcılardan hipotezin kabulü ya da reddine ilişkin deney sonuçları ile uyumlu kanıtlar üretmeleri beklenmektedir. Kanıt üretme görev tipi temelde deney sonuçlarının karşılaştırılmasına dayandırılmaktadır. Fakat söz konusu görev tipi literatürde sıklıkla rastlanan deney tipi görevlerden ayrılmaktadır (Klahr vd., 1993; Tschirgi, 1980). Deney tipi görevlerde, potansiyel değişkenlerden hangisinin sonuç üzerinde nedensel bir etkiye sahip olduğunu ortaya çıkarmaya yönelik katılımcıdan basit bir deney düzeneği tasarlanması beklenmektedir. Kanıt üretme görev tipinde ise katılımcıya tamamlanmış deney düzenek sonuçları (deliller) sunulmaktadır. Katılımcıdan deney sonuçlarını değerlendirerek hipotezi doğrulayan ya da yanlışlayan kanıtlar üretmesi beklenmektedir. Kuhn ve arkadaşlarının (1988) yaptıkları bir dizi çalışmada kanıt üretilmesine dayalı görev tipi için her ne kadar evet/hayır ya da var/yok şeklinde ikili yanıt sisteminden uzaklaşsa da üretilebilecek yanıtların sınırlılığı göze çarpmaktadır.

Sonuç olarak bilimsel yaratıcılığın ölçülmesinde kullanılan testlerin; geçerlik ve güvenilirlik çalışmalarının zayıf olması (Hu ve Adey, 2002; Liang, 2002), çoğunlukla belli bir yaş düzeyine uygulanması (Kanlı, 2014; Liang, 2002; Siew vd., 2014), kültürel değişkenler barındırmaları (Majumdar, 1975; Sinha ve Singh, 1987; Sukla, 1980) ve kuramsal alt yapılarının zayıf olması nedenleriyle eleştirilmektedirler (Hu ve Adey, 2002; Sharma ve Mahrshi, 2017; Usta ve Akkanat, 2015). Son dönem geliştirilen bilimsel yaratıcılık testleri ise söz konusu sınırlılıkların bir ya da birkaçına çözüm üretmiştir (Ateşgöz ve Sak 2021; Ayas ve Sak, 2014, Ayas 2017; Chin ve Siew, 2015; Hu ve Adey, 2002; Kanlı, 2014; Siew vd., 2014). Fakat SDDS modeli temel alınarak geliştirilen testlerde modelin bütünüyle doğrulanamadığı ve özellikle kanıt değerlendirme becerisinin ampirik olarak test edilemediği görülmektedir (Ateşgöz ve Sak 2021; Ayas 2017). Kanıt değerlendirme becerisinin küçük yaş gruplarında uygulandığı çalışmalarda ise görevlerin yalnızca hipotezi kabul etmek ya da reddetmek üzerine kurulu olduğu söylenebilir (Koerber vd., 2015; Piekny ve Maehler, 2012; Piekny vd., 2013; Ruffman vd., 1993). Bu bağlamda bilimsel yaratıcılık ile ilişkilendirilen kanıt değerlendirme, hipotez geliştirme, deney tasarlama bilimsel süreç becerileri ve yaratıcılıkta çoğul düşünmeye dayalı becerilerin bir arada tam anlamıyla test edilemedikleri söylenebilir.

1.2 Araştırmanın Amacı

Mevcut araştırmada SDDS modeli ile uyumlu çoğul düşünmeye dayalı kanıt değerlendirme görevlerinin hazırlanması ve ÇABİYAT testinin revizyonu amaçlanmıştır. Bu amaç kapsamında revize edilen ÇABİYAT (ÇABİYAT-R) için aşağıdaki sorulara cevap aranmaktadır:

1. ÇABİYAT-R'nin geçerliği nasıldır?
 - a. ÇABİYAT-R'nin yapı geçerliği nasıldır?
 - b. ÇABİYAT-R'nin ölçüt geçerliği nasıldır?
 - c. ÇABİYAT-R testinin madde yanlılığı nasıldır?
2. ÇABİYAT'-R'nin güvenilirliği nasıldır?
 - a. ÇABİYAT-R'nin iç tutarlılığı nasıldır?
 - b. ÇABİYAT-R'nin kodlayıcılar arası güvenilirliği nasıldır?

1.3 Araştırmanın Önemi

ÇABİYAT-R, yaklaşık 30 yıldır bilimsel keşif sürecinin kavranmaya çalışıldığı ve farklı deneysel çalışmalar ile test edildiği SDDS modeli (Klahr, 2000; Klahr ve Dunbar, 1993) temel alınarak revize edilmiştir. Test, teorik olarak sağlam bir zemine oturtulmuş olmasının yanında okul öncesi gruba yönelik geliştirilen bir bilimsel yaratıcılık testi olma özelliği de taşımaktadır. Ayrıca testte okul öncesi dönemi çocuklarının günümüze özgü gelişimsel özellikleri ve öğrenme gereksinimleri göz önüne alınarak hareketli görsel öğelerden/animasyonlardan yararlanılmıştır. Ayrıca test maddelerine verilecek yanıtlar bağlamında herhangi bir süre kısıtlamasına gidilmemesi, okul öncesi grubun bireysel farklılıklarının ön planda tutulduğu şeklinde yorumlanabilir. Yapılan çalışmalarda kanıt değerlendirme becerisini kimi sebeplerden dolayı bilimsel yaratıcılık testlerine dahil edilemediği görülmüştür. Bu sınırlılıkların başında ise kanıt değerlendirme görevlerinin küçük yaş grubu öğrencilerine uygun olmadığı ve kanıt değerlendirme görevleri için çoğul düşünmeye dayalı görev tiplerinin oluşturulamaması gelmektedir. Mevcut çalışma, kanıt değerlendirme becerisinin çoğul düşünme tipi görevlere uygun olduğunu, ayrıca bu görev için akıcılık puanının hesaplanabileceğini göstermiştir. Maddelerde öğrenciler kendilerine animasyon yolu ile sunulan kanıt kümelerine ilişkin madde bağlamında maksimum 7-9 arasında değişen yanıtlar üretebilmektedirler. Bu bağlamda kanıt değerlendirme becerisi için akıcılık puanının elde edilmesine yönelik literatürde belirtilen sınırlılığın (Ateşgöz ve Sak, 2021; Ayas, 2017) önüne geçildiği söylenebilir. Mevcut çalışma kapsamında açıklayıcı grafik analizi, açıklayıcı yapısal eşitlik modellemesi ve cinsiyet bağlamında ölçme değişmezliği şeklinde farklı ve yeni analiz yöntemleri benimsenmiştir. Faktör belirleme ve yapının doğrulanmasına ilişkin modern yaklaşım ve tekniklerin benimsenmesinin ise alanyazına önemli katkılar sağlayacağı düşünülmektedir.

1.4 Varsayımlar

- Araştırmada katılımcıların potansiyelleri dahilinde gerçek performanslarını sergiledikleri varsayılmıştır.
- Madde geliştirme ve madde seçiminde yer alan uzmanların maddeleri nesnellığe yakın bir şekilde değerlendirdikleri varsayılmıştır.

- Kodlayıcılar arası güvenilirlik analizleri kapsamında görüş bildiren alan uzmanının puanlamayı nesnelliğe yakın bir şekilde yaptığı varsayılmıştır.

1.5 Sınırlılıklar

- Araştırma Eskişehir ilinde iki anaokulu ve iki ilkokul ile sınırlıdır.
- Katılımcıların bilimsel yaratıcılıklarının ölçümü ÇABİYAT-R testinde yer alan beceriler ile sınırlıdır.
- Araştırmada kanıt değerlendirme becerisi için üretilebilecek yanıt sayısı, hipotez geliştirme ve deney tasarlama becerilerinde üretilebilecek yanıt sayılarına oranla sınırlıdır.
- Kanıt değerlendirme görevi için belirli sayıda fikir üretilmesi, esneklik ve orijinallik puanlarının hesaplanmasında bir sınırlılık yaratabileceği öngörüldüğü için öğrencilerin bilimsel yaratıcılıkları akıcılık puanları ile sınırlandırılmıştır.

1.6 Tanımlar

Yaratıcılık: Belirli bir alanda yeni ve kullanışlı fikir ya da ürünler ortaya koyabilme yeteneğidir (Ayas ve Sak, 2014).

Bilimsel Yaratıcılık: Alana özgü spesifik ve genel alan bilgisi ile bilimsel süreç becerilerini de içeren problem çözme yoludur (Klahr, 2000).

Tekil Düşünme: Üretilen fikirler içerisinde en uygun olanı belirleme sürecidir (Guilford, 1967).

Çoğul Düşünme: Problemin çözümüne ilişkin çok sayıda ve çeşitli fikir üretme sürecidir (Guilford, 1967).

Kanıt Değerlendirme: Deney sonuçları kapsamında hipotezle uyumlu kanıt üretme sürecidir (Kuhn vd., 1988).

2 ALANYAZIN

2.1 Yaratıcılık

Temelde problem çözme süreçleri ile sıklıkla ilişkilendirilen yaratıcılığın (Mumford vd., 1991; Runco 1984; Torrance 1971; Wallas 1926), tekerliğin icadından günümüzde yapay zekaya sahip robotların geliştirilmesine kadar tarihsel bir süreçte hizmet ettiği söylenebilir. Fakat yaratıcılığın bilimsel bir kavram olarak ele alınması ve ampirik çalışmalara konu olması J.P. Guilford 'un (1950) yayımladığı “Yaratıcılık” adlı makalesine dayandırılmaktadır. Bu tarihten itibaren yaratıcılığın tek boyutlu ve kutsal yapısından sıyrılarak çok boyutlu ve bilimsel bir kavram olarak çalışıldığı görülmektedir. Guilford (1950) yaratıcılığı, bilişsel süreç becerileri ile ifade ederek çoğul düşünme ile ilişkilendirirken; Stein (1953) yaratıcılığı, bir grup tarafından benimsenen kullanışlı, yeni bir ürünle sonuçlanan süreç olarak ele almakta; Plucker vd. (2004) yaratıcılığı yetenek, süreç ve çevre etkileşimi sonucunda kullanışlı ve yeni bir ürününün ortaya konulması olarak ifade etmektedirler. Her ne kadar yaratıcılık bir fikrin, ürünün ya da icadın yeni, orijinal, özgün olması ile ilişkilendirilse de söz konusu kavramların, yaratıcılığın ortaya çıkmasında gerekli ama yeterli olmadığı düşünülmektedir. Bu kapsamda son dönem yaratıcılık tanım ve kuramları incelendiğinde bir fikrin, ürünün ya da icadın orijinal olmasının dışında kullanışlı, etkili ya da yararlı olması gerektiği hususunda görüş birliği sağlandığı görülmektedir (Amabile, 1983a; Baer, 1993; Cropley, 2000; Mayer, 1999; Runco ve Charles, 1993; Sternberg vd., 2001; Sternberg ve Lubart, 1999). Yaratıcılığın tek bir boyut olarak ele alınamaması sebebiyle, çok boyutlu pek çok yaratıcılık kuram ve model ortaya çıkmıştır (Amabile, 2011; Baer ve Kaufman, 2005; Csikszentmihalyi, 2014; Guilford, 1967; Sternberg ve Lubart, 1993).

Yaratıcılığın genel bir yetenek mi yoksa alana özgü becerileri mi kapsadığı, yaratıcılık alanın tartışılan konularının başında gelmektedir (Baer, 1998; Plucker, 1998; Sawyer, 2024). Genel yaratıcılığı savunan araştırmacılar zekâ gibi yaratıcılığın da genel bir yetenek olduğunu, bir alandaki yaratıcılığın bir başka alana da transfer edilebileceğini iddia etmektedirler (Chen vd., 2006; Hocevar, 1976; Plucker, 1998; Runco, 1987). Diğer taraftan yaratıcılığın alana özgü olduğunu savunan araştırmacılar ise yaratıcılığın bir alandan bir diğerine kolayca transfer

edilemeyeceğini, bir alanda yaratıcılık sergileyebilmek için belirli bir bilgi, birikim ve deneyime sahip olunması gerektiğini belirtmektedirler (Baer, 1991, 1998, 2015; Han, 2003; Kaufman ve Baer, 2004; Runco, 1989; Silvia vd., 2009). Genel ve alana özgü yaratıcılığın pek çok araştırmada deneysel olarak çalışıldığı görülmektedir. Genel ve alana özgü yaratıcılık tartışmasının alanda yaratıcılık modellerinin oluşumundan, yaratıcılığın ölçülmesine kadar derin bir etki yarattığı söylenebilir. Genel yaratıcılık kapsamında yaratıcı kişilik özellikleri, yaratıcı süreç becerileri, yaratıcı tutum ve davranışlar değerlendirilmeye tabii tutulurken; alana özgü yaratıcılıkta ise sıklıkla yaratıcı ürün ve yaratıcı performans değerlendirmeleri yapılmaktadır (An vd., 2016; Baer, 1999; Lubart ve Guignard, 2004; Plucker, 1998). Geliştirilen yaratıcılık modellerinden Zihin Yapı Kuramı (Guilford, 1967) için genel yaratıcılık yaklaşımı benimsenirken; Bileşensel Model (Amabile 1983b, 2011), Sistemler Modelleri (Csikszentmihalyi, 2014) için alana özgü yaratıcılık yaklaşımı benimsenmiştir.

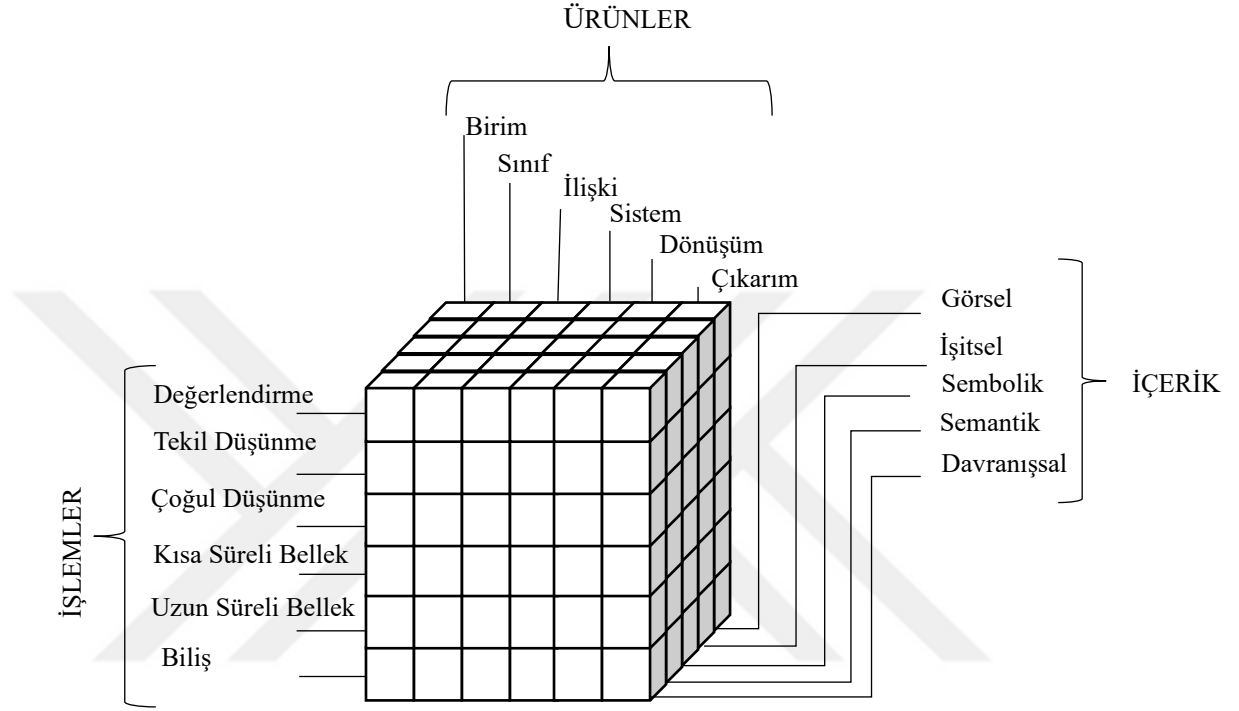
2.1.1 Zihin yapı kuramı modeli

En bilinen zekâ modellerinden biri olarak kabul edilen (Sternberg ve Grigorenko, 2001) ve Guilford (1967) tarafından ortaya atılan Zihin Yapı Modeli içerik, süreç ve ürün olmak üzere 3 boyuttan oluşmaktadır. Zihinsel süreç kapsamında işlenen bilgi türü içerik boyutu, bilginin nasıl işlendiği süreç boyutu ve zihinsel işlemler sonucunda elde edilen çıktılar ise ürün boyutu ile ilişkilendirilmektedir. Modelde içerik boyutu; şekilsel, görsel, işitsel, sembolik, anlamsal ve davranışsal bileşenlerden, süreç boyutu; biliş, hafızaya kaydetme, hafızada tutma, değerlendirme, tekil ve çoğul düşünme bileşenlerinden ve son olarak ürün boyutu ise; birim, sınıf, ilişki, sistem, dönüşüm ve uygulama bileşenlerinden oluşmaktadır. Modele 2.1’de yer ver verilmiştir. Modelin güncellenen hali ile boyutlar altında yer alan bileşenlerin farklı kombinasyonları oluşturulan 180 beceri bulunmaktadır. Guilford, modelinde yaratıcılığı zekanın bir parçası olarak ele almış ve ilişkilendirmiştir. Süreç boyutunda yer alan tekil ve çoğul düşünme bileşenleri ve ürün boyutunda yer alan dönüşümler bileşeni yaratıcılıkla ilişkilendirilen becerilerdir. Modele göre çoğul düşünme açık uçlu olarak tasarlanmış bir göreve ilişkin çok sayıda fikir üretmeyi, tekil düşünme ise üretilen fikirler arasından çözüm

için en uygununun seçilmesini ifade etmektedir. Özellikle çoğul düşünme testlerinin yaratıcılığın ölçülmesinde büyük bir özgül ağırlığa sahip olduğu söylenebilir.

Şekil 2.1

Zihin Yapı Kuramı modeli (Starko, 2017, s. 62)



Genel yaratıcılıkla ilişkilendirilen birçok yaratıcılık testi çoğul düşünme görevleri üzerinden yaratıcılığın değerlendirilmesinde rol almaktadır (Kaufman vd., 2008; Runco ve Acar, 2012). Söz konusu testlerin en bilineni Torrance (1972) tarafından geliştirilen Torrance Yaratıcı Düşünme Testi'dir (Torrance' s Test of Creative Thinking-TTCT). Çok sayıda dile çevrilmiş (Cramond vd., 1999) ve çok sayıda çalışmaya konu olmuş (Cramond vd., 2005; Humble vd., 2018; Kim vd., 2006a; 2006b; Krumm vd., 2014; Runco vd., 2010) TTCT testinde akıcılık, esneklik, orijinallik ve detaylandırma puanları hesaplanarak bireylerin yaratıcılıkları değerlendirilmektedir. Akıcılık üretilen fikir sayısı, esneklik kategori sayısı, orijinallik fikrin özgünlüğü ve detaylandırma ise fikirlerin ne kadar ayrıntıya sahip olduğu ile ilgilidir. (Torrance, 1972). Detaylandırma, her ne kadar yaratıcılık ölçümlerinde kullanılsa da deneysel ve psikolojik bir kavram olarak yaratıcılıkla ilişkilendirilmesinin sınırlı kaldığı söylenebilir (Atakaya vd., 2024; Dumas vd., 2021). Yaygın bir kullanım alanına sahip olmasına rağmen çoğul düşünme testlerinin

yordayıcı geçerlikleri (Kogan ve Pankove, 1974; Runco, 1999a), ayırt edici geçerlikleri (Zeng vd., 2011) ve yapı geçerlikleri konusunda (Baer, 1993) birtakım sınırlılıklara sahip olduğunu rapor eden çalışmalar görülmektedir. Ayrıca çoğul düşünmenin yaratıcılıkla bir ve benzer tutulmasının problemli olabileceği üzerine teorik tartışmalar yürütülmüştür (Runco, 2008).

2.1.2 Bileşensel yaratıcılık modeli

Yaratıcı bir ürün ortaya koyabilmek için bireyin psikolojik ve sosyal bileşenlerinin kapsamlı bir şekilde ele alınması gerektiğini savunan Amabile (1983b, 2011) bileşensel bir model ortaya koymuştur. Model, yaratıcılığın tarihsel olarak süreklilik arz etmesi ve belirli bir alanda dâhi yaratıcılık düzeylerinin farklılaşabileceği varsayımlarına dayanmaktadır. Model temel olarak alana özgü beceriler, yaratıcılığa özgü beceriler ve motivasyon bileşenlerinden oluşmaktadır. Araştırmacı sonraki çalışmalarında bileşenleri çevreleyen sosyal çevre adını verdiği dördüncü bileşeni eklemiştir. Alana özgü beceriler; zekâ, özel yetenek, teknik bilgi ve uzmanlıktan oluşurken; yaratıcılığa özgü beceriler farklı bakış açılarına sahip düşünme stilleri, bireysel özellikler ve çok sayıda fikir üretmeye yatkınlığı kapsamaktadır. Motivasyon ise içsel ve dışsal olarak ele alınmakta ve içsel motivasyonun yaratıcılığın oluşmasında ayrı bir öneme sahip olduğu vurgulanmaktadır. Ödül, para, rekabet gibi dışsal motivasyon kaynaklarından ziyade, göreve adanmışlık ve tutkuyu barındıran içsel motivasyon kaynaklarının yaratıcılığın ortaya çıkışında daha etkili oldukları ifade edilmektedir. Sosyal çevre bireyin içinde bulunduğu ortamı belirtmektedir. Sosyal çevre bireyin yaratıcılığını destekleyebildiği gibi yaratıcılığın oluşmasına ket de vurabilir.

Alana özgü yaratıcılığın benimsendiği bileşensel yaratıcılık modelinin yaratıcı ürüne odaklandığı söylenebilir. Bu bağlamda Amabile yaratıcılığın değerlendirilmesinde altın standart olarak görülen (Carson, 2006’ dan aktaran Baer, 2016) Konsensüs Değerlendirme Tekniği’ ni (KDT) önermiştir. Belirli bir alanda hazırlanan problemlerin çözümü için üretilen fikirlerin veya ürünlerin alan uzmanları tarafından bağımsız olarak puanlatılmasına dayanan KDT tekniği Amabile tarafından yaratıcılığın kalbine inen ölçüm olarak adlandırılmaktadır (Amabile, 1983a). Teknik, daha önce farklı araştırmalarda kullanılmasına rağmen Amabile ’ın değerlendirme sürecini sistematik bir hale getirdiği söylenebilir. KDT, farklı disiplinlerde çok sayıda

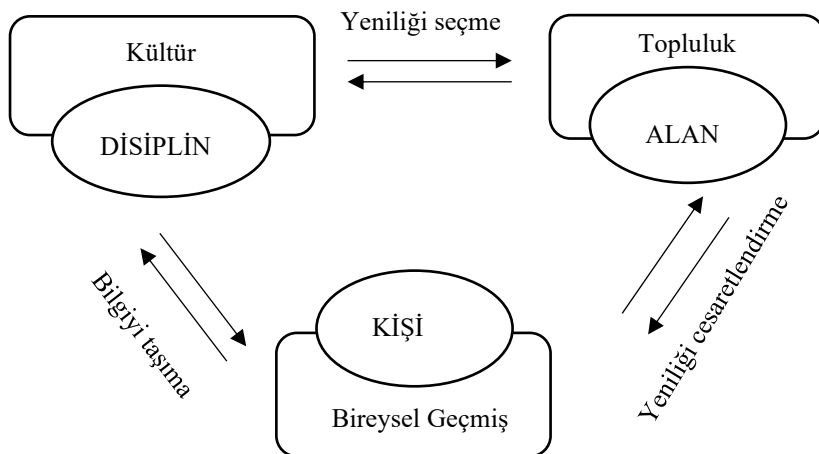
araştırmada uygulanmış (Arreola vd., 2016; Atakaya vd., 2024; Baer, 1994; Chen vd., 2006; Hickey, 2001; Long, 2014; Plucker vd., 2008; Reiter-Palmon ve Arreola, 2015) ve yaratıcılığın değerlendirilmesinde güvenilir sonuçlar alındığı raporlanmıştır. KDT tekniğinde alan uzmanlarının nasıl belirleneceği (Kaufman vd., 2008, 2013), yaratıcılık görevlerinin puanlanmasında doğabilecek zaman ve maliyet problemleri tekniğin sınırlı yönleri olarak kabul edilmektedir (Cseh ve Jeffries, 2019; Hennessey ve Amabile, 1999).

2.1.3 Sistemler modeli

Csikszentmihalyi (2014), yaratma eyleminin kişi (person) ile sınırlı kalamayacağını disiplin (domain) ve alan (field) bileşenlerinin bir sistem dahilinde birbirleri ile etkileşimleri sonucunda ortaya çıktığını savunmuştur. Modelde disiplin bileşenini kültürel öğeler ile (duygular, inançlar, öğrenilmiş kurallar) ilişkilendirmektedir. Her disiplinin kendine özgü birtakım kuralları ve uygulama adımları bulunmaktadır. Bir ürünün yaratıcı olarak kabul edilebilmesi için disiplinde yer alan kurallara uyması ya da disiplini yeni bir yöne çevirebilecek bir düşünce sistemi oluşturması gerekmektedir. Alan bileşeni ise disiplin içinde bulunan alan uzmanlarını kapsamaktadır. Bileşenlerin etkileşimlerine Şekil 2.2’de yer verilmiştir.

Şekil 2.2

Sistemler Modeli (Csikszentmihalyi, 1999, s. 315)



Uzmanlar bir nevi alanın bekçileri olarak görülmekte, yaratıcı biri ürünün kabulüne ya da reddine karar veren topluluk olarak ifade edilmektedir. Csikszentmihalyi' ye göre dışardan gelen mikroorganizmaların tehlikeli olup olmamaları durumunda karar mercii nasıl ki bağışıklık sistemidir, benzer şekilde bir ürünün yaratıcı olup olmadığına da alandaki uzmanlar karar vermektedir.

Modeldeki son bileşen kişide ise bireyin kalıtsal olarak getirdiği özellikleri ile yaşadığı çevre özellikleri yaratıcılıkla ilişkilendirilmektedir. Bireyin cinsiyeti, karakteristik özellikleri, aile ortamı, küçük yaşlardan itibaren zenginleştirilmiş bir çevrede büyümesi, yaratıcılığı besleyen diğer unsurlar olarak görülmektedir. Özetle, modelde yaratıcılık, bireyin genetik özellikleri ve yaşamını sürdürdüğü çevre ile etkileşimi sonucunda ilgisini çeken bir disiplinde yaratıcılık göstermesi ve gösterdiği yaratıcılığı alan uzmanları tarafından onaylanması ile ortaya çıkan bir süreci kapsamaktadır.

2.2 Bilimsel Yaratıcılık

Bilimsel düşünme becerilerinin ve bilimsel problem çözmenin yaratıcılıkla doğrudan ilişkili olduğu (Feist, 2011; Sternberg, 1999) hem yaratıcılık hem de bilimsel düşünme kavramlarının benzer keşif süreçlerini barındırdıkları (Hu ve Adey, 2002; Hu vd., 2010), yaratıcılığın bilim ve bilimsel düşünmenin ayrılmaz bir parçası olduğu ifade edilmektedir (Hodson ve Reid, 1988). Yaratıcı düşünme becerileri ile bilimsel süreç becerilerinin bir arada kullanıldığı (Ayas ve Sak, 2014; Ateşgöz ve Sak, 2021; Sak ve Ayas, 2013), bilimsel bilginin üretilmesinde yaratıcılığın önemi üzerinde durulduğu görülmektedir (Hadzigeorgiou vd., 2012; Kind ve Kind, 2007, Yıldız ve Yıldız, 2021).

Literatürde yaratıcılığın alana özgü olduğu ile ilgili yaklaşık 40 yıldır genel bir görüş birliğinin olduğu söylenebilir (Amabile, 1983b; Baer, 1998, 2016; Huang et al., 2017; Silvia vd., 2009). Feldhusen (2006), Weisberg (1999) ve Baer (2003) yaratıcı düşünme sürecinde alana özgü bilginin önemi üzerinde durmuşlardır. Benzer şekilde Alexander (1992) ve Amabile (1996) yaratıcı düşünme için belirli alan ya da disipline özgü bilgi ve becerilerin ihtiyacına değinmişlerdir. Alana özgü bilgi, beceri ve deneyimlerin kritik öneme sahip olduğu disiplinlerden birinin ise fen bilimleri olduğu iddia edilebilir. Bilimsel birikimin ötesine geçebilmek, yeni teknikler ortaya koyabilmek ya da bilimsel alanda keşifler sağlayabilmek için yaratıcılığa ihtiyaç

duyulmaktadır. Bu bağlamda farklı yaratıcılık türleri içerisinde bilimsel yaratıcılıktan da söz edilebilir (MacKinnon, 1975). Bilimsel yaratıcılık; bilimsel süreç becerileri, bilimsel bilgi ve bilimsel teknikler içermesi sebebiyle diğer yaratıcılık türlerinden ayrılmaktadır (Hu ve Adey, 2002; Lin vd., 2003). Hu ve Adey (2002) bilimsel yaratıcılığın hem statik hem de gelişimsel bir yapıya sahip olduğuna vurgu yapmaktadırlar. Getzels ve Csikszentmihalyi (1967), problem çözebilmenin yanında yeni sorunlar ortaya koyabilme yeteneğinin bilimsel yaratıcılıktaki önemine vurgu yapmışlardır. Einstein ve Infeld'a (1938) göre, yeni sorular ortaya koyabilmek, eski sorulara farklı bir açıdan bakabilmek yaratıcı hayal gücü gerektirmektedir. Moravesik (1981) bilimsel yaratıcılığı; yeni teoriler oluşturabilmek, yeni deneyler yapabilmek, var olan bilimsel bilgiye yeni kavramlar ekleyebilmek ve eklenen yeni fikir ve kavramları kavrayabilmek olarak tanımlanmaktadır.

Mansfield ve Busse (1981) bilim alanlarındaki yaratıcı süreci problemin dikkatli bir şekilde seçilmesi, problemi çözmeye yönelik yoğun çaba, problemin çözümünde kullanılacak yöntem ve tekniklerin seçilmesi, kararların hipotezler doğrultusunda değiştirilmesi, doğrulama ve ayrıntılandırma olmak üzere beş aşamaya ayırmışlardır. Simonton (2004) bilim tarihine adını yazdırmış ünlü bilim insanlarının hayatlarını detaylı bir şekilde incelemiş ve bilimsel yaratıcılığı; şans, mantık, dahilik ve zamanının ruhu (zeitgeist) faktörleri ile ilişkilendirmiştir. Yazara göre bilimsel yaratıcılığın ortaya çıkmasında; beklenmedik ya da kazara çıkan durumlar, mantıksal ve analitik düşünce, bireyin sahip olduğu zekâ, hayal gücü ve yaratıcı problem çözme yetenekleri, sosyal, kültürel ve bilimsel bağlam etkili olmaktadır. Hu ve Adey (2002) bilimsel yaratıcılığı, belirli bir amaç doğrultusunda bilgileri kullanarak özgün, bireysel ya da sosyal değeri olan, üretilmiş ya da üretilme potansiyeline sahip zihinsel özellik ya da yetenek; Heller (2007) ise karmaşık, bilimsel ve teknik konuların yenilikçi ve üretken bir şekilde çözümüne ilişkin bireysel ve toplumsal kapasite olarak tanımlanmaktadırlar. Bazı araştırmacılar bilimsel yaratıcılığı alana özgü spesifik ve genel alan bilgisi ile bilimsel süreç becerilerini de içeren problem çözme yolu olarak ifade etmektedirler (Klahr, 2000). Bilimsel yaratıcılık kapsamında yapılan tanımlamalar genel olarak değerlendirildiğinde bilimsel keşfin özgünlüğü ve birey-toplum nezdindeki yaygınlığının ön plana çıktığı söylenebilir. Söz konusu kavramların genel yaratıcılık becerileri bağlamında orijinallik ve uygunluk kriterleri örtüştüğü söylenebilir. Aynı zamanda bilimsel yaratıcılığın alana özgü birtakım

bilimsel süreç becerileri, bilimsel bilgi ve deneyimi de içerdği görülmektedir. Çalışma kapsamında bilimsel yaratıcılık tanımı “*bilim alanında orijinal ve kullanışlı fikir ya da ürün yaratabilme yeteneği*” olarak kabul edilmiştir (Sak ve Ayas, 2013; Ayas ve Sak, 2014).

Bilimsel keşfin doğasını anlamaya yönelik farklı türde yapılmış araştırmalar mevcuttur. Klahr ve Simon (1999) bilimsel keşfi anlamaya yönelik çalışmaları; bilimsel keşiflerin tarihsel anlatıları (Feist, 1994; Gruber, 1974), bilim insanı olmayan kişiler ile yapılan psikolojik deneylere dayalı çalışmalar (Klahr vd., 1993; Kuhn, 1989; Kuhn vd., 1988; Sodian vd., 1991), bilimsel laboratuvarlarda doğrudan gözleme dayalı çalışmalar (Dunbar, 1995; Dunbar ve Baker, 1994), simülasyona dayalı hesaplamalı model çalışmaları (Cheng ve Simon, 1992; Gooding, 1990) ve sosyolojik yaklaşımlara dayalı çalışmalar (Pickering, 1992; Shadish ve Fuller, 1994) olmak üzere beş başlık altında toplamışlardır. Bilimsel yaratıcılığı öncül değerlendirmelerden farklı olarak matematiksel anlamda açıklamaya çalışan araştırmacılara da rastlanmaktadır (Simonton, 2021). Laboratuvar çalışmaları genel olarak alana özgü ve genel alan bilgisi şeklinde iki boyutta sınıflandırılmaktadır. Alana özgü bilgi ve genel alan bilgisi, bilimsel düşünme sürecini doğrudan etkilemektedir. Bu çalışmalardan özellikle deneysel çalışmalara dayalı olan, bilimsel yaratıcılığın temellerini ve kuramsal yapısını ortaya çıkarmayı hedefleyen ve mevcut çalışmanın da teorik alt yapısını barındıran Bilimsel Yaratıcılığın İkili Arama Modeli’nin detaylı bir şekilde ele alınması faydalı olacaktır.

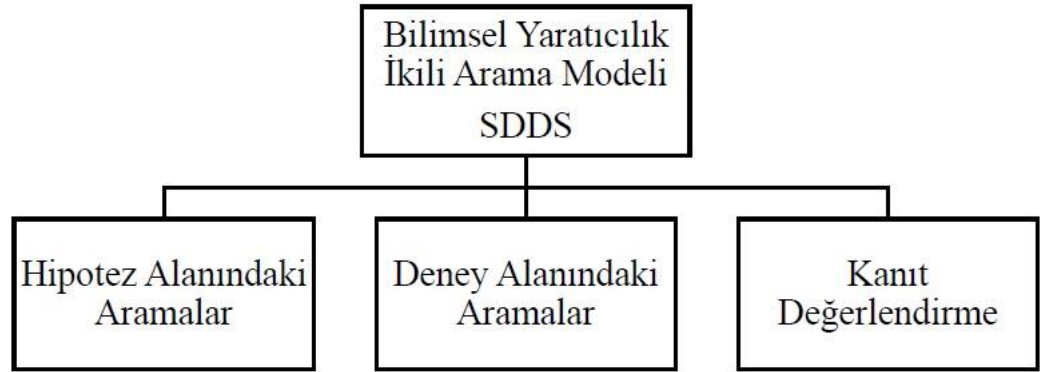
2.2.1 Bilimsel yaratıcılığın ikili arama modeli

Literatürde bilimsel düşünmeye ilişkin çalışmalar temel olarak birbirleri ile ilişkili iki ayrı yaklaşımla sürdürülmüştür. Bunlardan birincisi deney tasarımlarındaki belirsizlik ve kontrollü değişken arayışlarını barındıran bilimsel araştırmaların genel ilkelerinin incelenmesi (Kuhn vd., 1988; Somerville vd., 1979), diğeri ise disipline özgü bilimsel kavramların edinimi ve geliştirilmesine dayalı çalışmalardır (Carey, 1985; Smith vd., 1985). Bilimsel yaratıcılığın ikili arama modeli (SDDS) ise bilimsel düşünme çalışmalarında önemini koruyan bu iki yaklaşımı bir araya getirdiği söylenebilir (Klahr, 2000; Klahr ve Dunbar, 1988). SDDS modeli bilimsel yaratıcılığı hipotezlerin geliştirilmesi ve deneysel kanıtlara göre hipotezlerin revize edilmesi amacıyla hipotez ve deney alanlarında gerçekleşen problem çözme süreci ile

kavramsallaştırılmaktadır. SDDS, alana özgü bilgi ve genel alan deney yapma süreçlerinin birbirlerini nasıl ve neden etkilediğini açıklayarak bilimsel yaratıcılık sürecinde iki hattı birbirine bağlamayı hedeflemektedir. Model birbirleriyle ilişkili hipotez geliştirme, deney tasarlama ve kanıt değerlendirme olmak üzere üç ana bileşene dayanmaktadır. Modele göre var olan önceki bilgilere dayanılarak hipotezler oluşturulur. Daha sonra hipotezler test edilmek üzere deneyler tasarlanır ve deneyler sonucunda kanıtlar üretilir. Üretilen kanıtlar oluşturulan hipotezle karşılaştırılarak hipotezin reddine ya da kabulüne karar verilir. Hipotez reddedilirse revize edilerek test edilir ve yeni kanıtlar üretilir. Şekil 2.3'te modelin temel bileşenlerine yer verilmiştir.

Şekil 2.3

SDDS modeli ve temel bileşenleri (Klahr, 2000, s. 31)



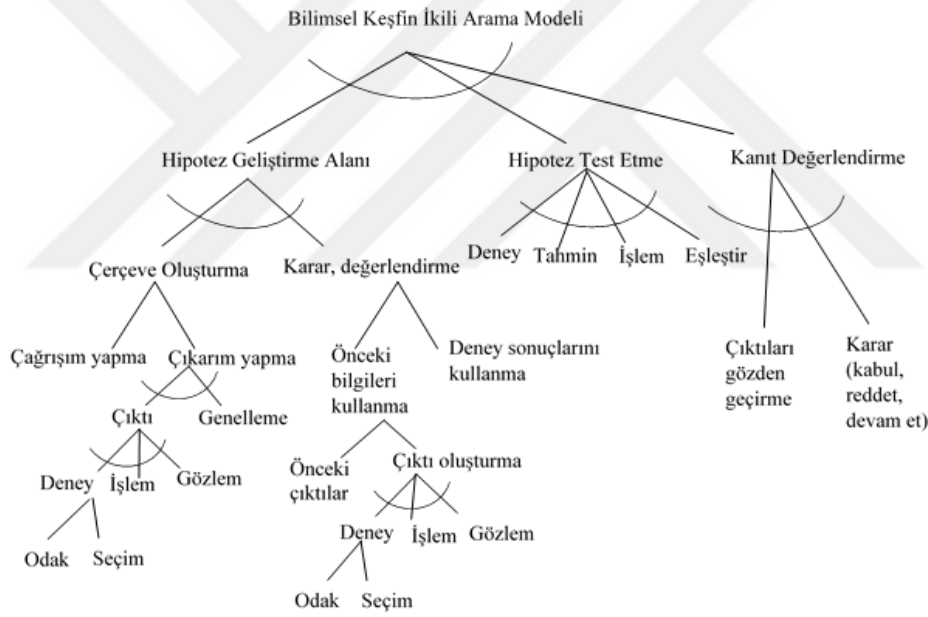
Hipotez geliştirme süreci bir tür problem çözmedir. Bu süreçte birey başlangıçta alanla ilgili var olan bilgilerini kullanarak problem durumunun bir kısmını ya da tamamını açıklayabilecek şekilde hipotezler oluşturmaya çalışır. Oluşturulan hipotezlerin makul olma düzeylerinin değerlendirilmesi ise bireylerin alandaki uzmanlıkları ile doğrudan ilişkilidir. Ayrıca bireyler yüksek ve düşük olasılıklı hipotezleri değerlendirirken farklı deneysel stratejilere başvurumaktadırlar.

Deney tasarlama bileşeninde hipotezi test etmek amacıyla yorumlanabilir sonuçlar üretmek adına deneyler gerçekleştirilmektedir. Birey bir deney alanı oluşturarak bir tahminde bulunur, deneyi gerçekleştirir ve sonucu tahminle eşleştirmektedir. Birey hipotez geliştirme ve deney tasarlama süreçleri sonrasında kanıt değerlendirme adımına yönlendirilmektedir. Bu aşamada hipotez ile kanıt arasındaki uyum

değerlendirilir. Eğer hipotez ile kanıtlar uyumlu ise süreç tamamlanır, değilse hipotez revize edilerek tekrar deney tasarlanır ve kanıtlar elde edilmeye çalışılır. Kanıt değerlendirme sürecinde bireylerin önsel bilgileri kanıtlar üzerinde baskın bir ön yargı oluşturabilmektedir. Kanıtlar hipotez ile uyumlu olmamasına rağmen katılımcı önceki bilgilerine sıkı sıkıya bağlı kalarak hipotezi reddetmeyebilir (Koerber vd., 2005; Piekny ve Maehler, 2012; Piekny vd., 2013). Modelin sonraki güncellemelerinde temel bileşenler daha da detaylandırılarak bileşenlere bilişsel süreçler, beceriler ve kimi yetenekleri içeren alt bileşenler eklenmiştir. Şekil 2.4’ de SDDS modelinin güncel haline yer verilmiştir.

Şekil 2.4

Bilimsel Keşfin İkili Arama Modeli bileşenleri (Klahr, 2000, s. 37)



Şekilde görüldüğü üzere hipotez geliştirme bileşeni, bir hipotezin kapsamlı bir şekilde ele alındığı *çerçeve oluşturma* ve hipotezin yeniden düzenlendiği, belirginleştirildiği, *karar, değerlendirme* alt bileşenlerinden oluşmaktadır. Modele göre çerçeve oluşturmaya; bellekten depolanmış bilgi, *çağrışım yapma* ve deneysel verilerden elde edilen bilgiler, *çıkarım yapma* kaynaklık etmektedir. Çıkarım yapmada ilk adım bir çıktının üretildiği *çıktı* ve çıktıların genelleştirildiği *sonuç genelleştirme* işlemlerinden oluşmaktadır. Karar, değerlendirme bileşeninde de benzer bir süreç işlediği söylenebilir. Bir hipotezin ayrıntılandırılması *önceki bilgileri kullanma* ve *deney sonuçlarını kullanma* süreçleri ile devam ettirilmektedir. Söz

konusu adımlar, çerçeve oluşturma bileşeninde yer alan çağrışım yapma ve çıkarım yapma adımları ile paralellik gösterse de burada yeni bir çerçeve oluşturmada ziyade var olan hipotezin ayrıntılandırması söz konusudur. Deney sonuçları kullanılırken eski deneylerden yararlanılabileceği gibi ampirik olarak türetilen yeni deney sonuçlarından da yararlanılabilir.

Şekilde görüldüğü üzere hipotez test etme *deneyin tasarlanması, tahmin, deneyin gerçekleştirilmesi ve eşleştirme* adımlarından oluşmaktadır. Bileşene göre hipoteze uygun bir deney tasarlanır, hipotez ve deney kapsamında tahminlerde bulunulur, ardından deney gerçekleştirilir ve tahmin ile gerçek sonuç arasındaki uyumun eşleştirilmesi sağlanır. Modelin üçüncü bileşeni kanıt değerlendirme sürecinde *çıktıların gözden geçirildiği* ve hipoteze ilişkin gerçekleştirilen deneyler sonucunda elde edilen kanıtların *hipotezi reddetmek ya da kabul etmek* için yeterli olup olmadığının belirlendiği adımdır. Kanıtların yeterli olmadığı ya da hipotezin reddine karar verildiği durumlarda hipotezin revize edilmesi ve yeniden test edilmesi sürecine dönülür. Kanıt değerlendirme sürecinde akla yatkınlık, işlevsellik ve yalınlık gibi ek faktörlerin varlığı tahmin edilse de söz konusu faktörlerin işleyişine yönelik net bir açıklama yapılamamaktadır. Mevcut çalışma kapsamında geliştirilen ÇABİYAT-R bilimsel yaratıcılık testi içerisinde hipotez geliştirme, deney tasarlama ve kanıt değerlendirme bilimsel süreç becerileri yer almaktadır. Bu kapsamda söz konusu becerilere yönelik literatürde gerçekleştirilmiş deneysel çalışmalara detaylı bir şekilde yer verilmiştir.

2.2.1.1 Hipotez geliştirme ile ilgili araştırmalar

Hipotez geliştirme, öncül bilgiler ya da kanıtlar ışığında problemin çözümüne yönelik mantıklı önermelerin sunulduğu (Klahr ve Dunbar, 1988, NRC, 2012), test edilecek değişkene rehberlik ederek, kanıt ve teoriyi birbirine bağlayan (Kuhn vd., 1988) bilimsel süreç becerisini ifade etmektedir. Klahr vd. (2019) yaptıkları sınıflandırmada hipotez geliştirme ve hipotezi yeniden düzenleyebilme becerisini; sorular sorabilme, analogiler kurabilme ve bilimsel modelleri kullanabilme ile ilişkilendirmişlerdir. Piekny ve Maehler (2012), 4- 13 arasında beş farklı yaş grubundan toplam 223 çocuk ile gerçekleştirdikleri çalışmalarında çocukların genel alan bilgileri üzerinden hipotez geliştirme, deney tasarlama ve kanıt değerlendirme becerilerini ölçmeyi amaçlamışlardır. Hipotez geliştirme becerisine yönelik görevler

araştırmacılar tarafından tasarlanmıştır. Tasarlanan maddelere göre çocukların kendilerini birer hayvan araştırmacısı olarak görmeleri gerekmektedir. Deneyi uygulayan uzman, öğrencilere bir örneğin de yer aldığı toplam 7 hayvanın farklı görsellerinin bulunduğu kartları göstermişlerdir. Kartlarda yer alan hayvanların belirli bir hayvan ailesine ait olup olmadıkları belirtildikten sonra, öğrencilerden bu aileye ait olmak için hayvanın hangi vücut parçasının önemli olup olmadığını bulmaları beklenmiştir. Araştırma sonuçlarına göre öğrencilerin aldıkları maksimum puanlar ve ortalamaları sırasıyla; küçük okul öncesi grup için 3 ve 0.79; büyük okul öncesi grup için 2 ve 0.87; ilkokul öğrencileri için 20 ve 3.08; ortaokul öncesi grup için 23 ve 3.00 ve son olarak ortaokul öğrencileri için 24 ve 9.74 olarak hesaplanmıştır. Yapılan ANOVA analiz sonuçlarına göre büyük okul öncesi grup ile ilkokul öğrencileri ve ortaokul öncesi grup ile ortaokul öğrencileri arasında anlamlı farklılıklar sağlanırken diğer gruplar arasında anlamlı bir farklılık elde edilememiştir. Araştırmacılara göre hipotez geliştirme becerisi okul öncesi grup için oldukça zorlayıcı ve bu yaş grubuna uygun değildir. Araştırmacılar hipotez geliştirme becerisinin sonraki yıllarda ortaya çıkabileceğini savunmuşlardır. Çalışmanın hipotez geliştirme görevleri incelendiğinde, söz konusu becerinin muhakeme becerileri ile ilişkili olduğu düşünülmektedir.

Ön test ve son test grupları ile gerçekleştirilen yarı deneysel bir çalışmada öğrencilere alan bilgisi sunulması durumunda hipotez geliştirme becerilerinin nasıl etkilendiği araştırılmıştır (Kuang vd., 2022). 69 ortaokul öğrencisi, rassal olarak kontrol ve deney gruplarına atanmış, deney grubuna fizik alanında temel alan bilgisi sunulurken kontrol grubuna herhangi bir müdahalede bulunulmamıştır. Araştırma sonuçlarına göre deney grubu öğrencilerinin hipotezlerinin daha test edilebilir oldukları ve hipotezlerinin değişkenler, koşullar ve ilişkiler bağlamında daha bilgilendirici oldukları görülmüştür. Benzer başka bir çalışmada da 118 ortaokul öğrencisi fizik alanından kuvvet ve hareket ile ilgili kontrol ve deney gruplarından oluşan 4 farklı gruba rastgele olarak atanmışlardır. Ön ve son test sonuçlarına göre öğrencilerin hipotez geliştirmeden önce değişkenlere ilişkin bilgi topladıkları, alan bilgisinin öğrencilerin hipotez geliştirme becerilerini geliştirdiğine yönelik bulgular raporlanmıştır (Kuang vd., 2024). Bilgisayar destekli probleme dayalı bir öğrenme ortamında 172 altıncı sınıf öğrencisinin hipotez geliştirme becerileri incelendiği bir çalışmada öğrencilerden petrol sızıntısı sonucunda zarar gören hayvanların nasıl

kurtulacağına yönelik hipotez geliřtirmeleri beklenmiřtir (Kim ve Pedersen, 2011). Arařtırmada deney grubu öğrencilerine üstbiliřsel becerilerini desteklemeye yönelik kontrol listeleri ve hipotez geliřtirme sürecinde bazı yönlendirici sorulara eriřim imkânı sunulmuřtur. Arařtırma bulgularına göre üstbiliřsel destek imkânı sunulan öğrencilerin anlamlı derecede daha iyi hipotezler geliřtirdiklerive hipotez geliřtirme performanslarının, problem geliřtirme performanslarını %15 düzeyinde açıkladıđı görölmüřtür.

Hipotez geliřtirme becerisinin temel alındıđı çalışmalar genel olarak deđerlendirildiđinde bu becerinin küçük yař gruplarına uygun olmadıđı (Piekný ve Maehler, 2012), çocukların mantıklı hipotezler geliřtirebildikleri ancak tek bir hipoteze odaklandıkları görölmektedir (Lazonder ve Harmsen, 2016), büyük yařtaki öğrencilerin ise genellikle daha az mantıklı hipotez geliřtirdikleri ancak birden fazla hipotezi düşünme olasılıklarının daha yüksek olduđu görölmektedir (Klahr vd., 1993). Diđer taraftan ergenler ve yetiřkinlerin bile sınırlı sayıda hipotez geliřtirdikleri ve hipotezlerin tahminler ile karıřtırıldıđı (Gijlers ve De Jong, 2005; Njoo ve De Jong, 1993), hipotez geliřtirmenin yař grupları arasında farklılařtıđı belirtilmektedir (Klahr vd., 1993). Ayrıca farklı çalışmalarda küçük yař gruplarında hipotez geliřtirme becerisinin uygulanabildiđi ve tatmin edici sonuçların alındıđı görölmektedir (Ayas ve Sak, 2014; Ayas, 2017; Sak ve Ayas, 2013; Ateřgöz ve Sak, 2021). Bu bağlamda hipotez geliřtirme becerisine iliřkin çeliřkili sonuçların yer aldıđı söylenebilir.

2.2.1.2 Deney tasarlama ile ilgili arařtırmalar

Deney tasarlama, gözlemler üzerinden verilerin oluřturduđu deseni ortaya çıkarmak için bir hipotezin türetilmesi ya da var olan bir hipotezin test edilmesi amacıyla gerçekteřtirilmektedir (Klahr ve Dunbar, 1988). Temel olarak deney yapma süreci, yorumlanabilir kanıtlar veya gözlemler üretmek, alternatif hipotezleri dışlayacak řekilde deđerřkenleri kontrol altına almaktır (Zimmerman, 2007). Klahr vd., (2019) deney süreci ile ilgili arařtırma becerilerini deneyin planlanması ve uygulanması olarak sınıflamıřlardır.

Deneylerin yürütöldüđu ve sonuçlarının deđerlendirildiđi çalışmalar Inhelder ve Piaget'in (1958) öncöl arařtırmalarına kadar dayanmaktadır. Çocukların ve

ergenlerin deney tasarlama ve deney sonucunda ortaya çıkan verileri değerlendirme becerilerinin incelendiği çalışmada bilimsel düşünme süreçlerinin ancak 11 yaş üst dönemde ortaya çıkabileceği rapor edilmiştir. Siegler ve Liebert (1975) Piaget'in çalışmasını temel alarak, 5.ve 8.sınıf öğrencilerinin deney tasarlama becerilerini elektrikli bir trenin doğru bir şekilde çalıştırılması üzerinden incelemişlerdir. 4 farklı açma kapama düğmesi olan elektrikli trenin hareket ettirilmesinde doğru seçimin yapılması 16 farklı kombinasyona kadar ertelenebilmektedir. Kendilerine herhangi bir eğitim verilmeyen ve sonuçların kaydını tutmayan öğrencilerin çok azı, beşinci sınıflar için %0 ve altıncı sınıf öğrencilerinin yalnızca %10'u, doğru anahtar kombinasyonlarının tamamını oluşturabildikleri görülmüştür. Diğer taraftan kendilerine eğitim verilen ve deney öncesinde pratik yaptırılan beşinci sınıf öğrencilerinin %70'i, sekizinci sınıf öğrencilerinin ise tamamı değişkenleri kontrol altına almayı ve treni hareket ettirebilmeyi başarmışlardır.

Deney tasarlama ve yürütme süreci üzerine gündelik yaşamdan problemlerin ele alındığı farklı çalışmalarda yapılmıştır (Chen ve Klahr, 1999; Siler ve Klahr, 2012; Zimmerman ve Croker, 2014). Tschirgi, (1980) 2, 4 ve 6. sınıf öğrencileri ve yetişkinler ile gerçekleştirdiği çalışmasında katılımcılara bir kekin tadının kötü olması ve kâğıttan yapılan bir uçağın iyi olması ile ilgili verdiği hipotezleri test etmelerini istemiştir. Katılımcılardan bir seferde tek bir değişkeni değiştirmeleri, tek bir değişkeni sabit bırakmaları ya da tüm değişkenleri değiştirebilecekleri 3 farklı hipotez test etme yönteminden birini seçmeleri istenmiştir. Araştırma sonuçlarına göre 2. ve 4. sınıf öğrencilerinin tüm değişkenleri değiştirmeye daha yatkın oldukları, katılımcıların tamamının olumlu sonuçları yeniden üretme ve olumsuz sonuçları dışlamalarının altında rasyonel nedenlerden ziyade pratik yaklaşımların yattığı görülmektedir. Sodian vd. (1991), 6-9 yaş arası öğrencilerin deney yapma becerilerini fare görevi üzerinden değerlendirmişlerdir. Öğrencilere evlerinde yaşayan farenin büyüklüğünü belirlemeye çalıştıkları hipotetik bir hikâye sunulmuştur. Hikâyeye göre büyük fare yalnızca büyük fare evine girebiliyorken; küçük fare hem büyük hem de küçük fare evine girebilmektedir. Öğrencilere, içerisinde yiyecek bulunan farklı büyüklükte fare evleri gösterilmiş, farenin boyutuna bakılmaksızın hangi fare evinin (besleme koşulu) ve farenin yiyeceğe ulaşip ulaşamama durumuna göre hangi kutunun seçilmesi (öğrenme koşulu) gerektiği sorulmuştur. Araştırma sonucunda birinci sınıf öğrencilerinin %90'ı ikinci sınıf

öğrencilerinin ise tamamı hangi kutunun seçilmesi gerektiğini doğru ayırt edebilirlerken; birinci sınıf öğrencilerinin %55'i ve ikinci sınıf öğrencilerinin %86'sı besleme koşuluna doğru cevap verebilmişlerdir. Araştırmacılara göre 6 yaş öğrencilerinde dâhi verilen bir hipotez üzerinden deney sürecinin test edilebildiği raporlanmıştır. Aynı görevin temel alındığı başka bir çalışmada ise öğrencilerin 7 yaşına kadar hipotezin kesin ya da kesin olmayan sonuçlarını test etmede başarısız oldukları görülmüştür (Piekný ve Maehler, 2012).

Zimmerman ve Glaser (2001), altıncı sınıf öğrencilerinin bir hikâyede kendilerine sunulan olumlu ve olumsuz görüşlere dayalı hipotezleri test etmek üzere bir deney tasarımları istenmiştir. Çalışma biyoloji alanında bitkilerin büyümesi ile ilgili bir görev olup hikâyede yer alan olumsuz hipotezin, *musluk suyunun bitkiler için kötü olduğu*, olumlu hipotezin ise *kahve telvesinin bitkiler için iyi olduğu* şeklinde belirtilmiştir. Öğrencilerin %79'unun olumsuz hipotezi test etmede başarı sağladıkları ve bağımsız değişken olan su türünü kontrol altına almayı önerdikleri görülmüştür. Öğrencilerin olumlu hipotezi yani kahve telvesinin bitkiler için iyi olduğunu test etmek yerine iddianın genellenebilir olmasına yönelik deney tasarımları önerdikleri görülmüştür. Öğrencilerden yalnızca dörtte birinin kahve telvesi değişkeni üzerinden deney tasarımı önerdikleri büyük çoğunluğunun odak değişken olarak farklı bitki türlerini ele alan deney tasarımları önerdikleri görülmüştür. Araştırmada öğrenciler düşük, orta ve yüksek yetenek gruplarına göre ayrılmış ve tüm yetenek gruplarının hikâye formatından etkilendikleri, en çok etkilenen grubun ise düşük yetenek grubu öğrencileri olduğu raporlanmıştır (Zimmerman ve Glaser, 2001).

Schauble vd. (1991) beşinci ve altıncı sınıf öğrencileri ile hidrodinamik üzerine geliştirilen kanal görevi ve hidrostatik üzerine geliştirilen yay görevi üzerine iki adet hikâye formatı ile bir çalışma gerçekleştirmişlerdir. Öğrencilerden birinci hikâye formatında bir bilim insanı olarak düşünmeleri, görevlerde hangi faktörlerin etkili olup olmadıklarını (örneğin kanal derinliği, botun büyüklüğü, bot şekli ve ağırlığı vb.) belirlemeleri istenmiştir. İkinci hikâye formatında ise öğrencilerden bir mühendis olarak düşünmeleri ve görevler bağlamında istenen etkinin en optimize edilmiş halini (örneğin kanal görevi için en hızlı botun nasıl tasarlanabileceği) ortaya koymaları beklenmiştir. Öğrencilerin bilim insanı bağlamında oldukça sistematik

çalıştıkları her değişkenin tek başına ve değişkenlerin birlikte olan etkilerini inceledikleri ve faktörleri ayırt etmede başarılı oldukları görülmüştür. Diğer taraftan öğrencilerin kendilerini bir mühendis olarak gördükleri hikâye formatında ise nedensel olduğu düşünülen oldukça zıt kombinasyonlara odaklanıldığı (örneğin botun çok ağır, büyük olması ve çok hafif, küçük olması) nedensel olmayan faktörlerin ise göz ardı edildiği görülmüştür. Bu bağlamda araştırmacıların teori odaklı olarak belirttikleri öğrencilerin kendilerinin bir bilim adamı görmeleri ile deneme ve görme odaklı olarak ifade ettikleri öğrencilerin kendilerini mühendis olarak gördükleri durumda öğrencilerin deney tasarlama süreçlerinin farklılaştığı söylenebilir.

Çalışmalar genel olarak değerlendirildiğinde deney tasarımı ile ilgili araştırmaların genellikle değişkenlerin kontrolü stratejisine dayandırıldığı (Schalk vd., 2019; Schauble vd., 1991; Tschirgi, 1980; Zimmerman ve Glaser, 2001), anaokulu düzeyindeki çocukların tek değişkenli deneyleri anlama ve test etmede başarılı olabildikleri (Cook vd., 2011; Croker ve Buchanan, 2011; Piekny ve Maehler, 2012; Siegler ve Chen, 1998; Van der Graaf vd., 2018) fakat değişken sayısı artırıldığında hipotezi test etmede ya da deneyi tasarlamakta zorlandıkları, bu durumun kimi zaman büyük yaş gruplarında da kendini gösterebildiği söylenebilir (Chen ve Klahr, 1999; Wilkening ve Huber, 2004). Katılımcılara önceden verilmiş bir eğitim ya da pratığın deneyin tasarlanması, neden-sonuç ilişkisinin açıklanması ve değişkenlerin kontrolü ya da manipülasyonu noktasında olumlu yönde katkı sağladığı görülmüştür (Chen ve Klahr, 1990; Schlatter vd., 2020; Siegler ve Libert, 1975).

2.2.1.3 Kanıt değerlendirme ile ilgili araştırmalar

Kanıt, hipotezin doğrulanması ya da reddedilmesine hizmet eden bilgi ya da gözlemler olarak ifade edilmektedir (Klahr ve Dunbar, 1988; Zimmerman, 2000). Kuhn vd. (1988) bilimsel düşünmenin doğasını açıklamaya çalıştıkları çerçeve modelde bilimsel düşünmeyi hipotez-kanıt arasındaki farkı ayırt ve koordine edebilme yeteneği olarak tanımlamaktadırlar. Klahr vd. (2019) kanıt değerlendirmeyi, kanıt ve hipotezler arasında nasıl bir ilişki olduğunu ortaya koymaya yönelik açıklamaların oluşturulduğu araştırma döngüsünün bir parçası olarak ele almaktadırlar. Kanıt değerlendirme becerisi; kanıt örüntülerinin

değerlendirilmesi, verilerin analiz edilmesi, yorumlanması ve kanıtlara yönelik açıklamalar oluşturma şeklinde sınıflandırılmaktadır (Klahr vd. 2019).

Kanıt değerlendirme becerisinin çalışmalarda sıklıkla *kanıt örüntülerine* dayandırıldığı, çocukların kendilerine sunulan kanıtlardan yararlanarak nedensel ilişkilere dayalı çıkarım yapmaları beklenildiği, kanıtların birbirleriyle gösterdikleri kovaryans düzeylerinin belirlenmeye çalışıldığı görülmektedir (Amsel ve Brock, 1996; Bullock ve Ziegler, 1999; Croker ve Buchanan, 2011; Koerber vd., 2005; Kuhn vd., 1988, 1995; Kuhn ve Pearsall, 2000; Piekny vd., 2013; Piekny ve Maehler, 2012; Ruffman vd., 1993). Söz konusu araştırmalarda katılımcılara görsel olarak sunulan bir dizi kanıt mükemmel uyum, kısmi uyum ya da tamamen uyumsuzluk barındırmakta, katılımcılara kanıtlar ile ilgili hangi sonuçların çıkarılabileceği sorulmaktadır. Amsel ve Brock (1996) çalışmalarında 77 ikinci ve üçüncü sınıf, 85 altıncı ve yedinci sınıf, 40 üniversite öğrencisi ve 36 üniversite eğitimi almamış yetişkinin kanıt değerlendirme becerilerindeki gelişimsel farklılıkları incelemiştir. Katılımcılar ile ilk olarak güneşin bitkilerin büyümesiyle nedensel bir ilişkiye sahip olduğuna ve bir tılsımın (dört yapraklı yonca) nedensel bir ilişkiye sahip olmadığına yönelik öncül inanışları konusunda mülakatlar yapılmıştır. Sonrasında katılımcılara bitkinin büyümesi ile ilgili olarak en az iki sağlıklı ve iki hasta bitki içeren dört farklı veri seti sunularak kanıtları değerlendirmeleri beklenmiştir. Veri setlerinde bitkinin güneş alıp almama durumu (öncül inanış bağlamında nedensel etki olduğu inanılan) ve tılsımın varlığı ya da yokluğuna (öncül inanış bağlamında nedensel bir etki olmadığına inanılan) ilişkin olarak bitkilerin büyümelerinde mükemmel korelasyonun sağlandığı ve korelasyonun olmadığı kanıtlar sunulmuştur. Araştırma bulgularına göre küçük yaş grubu öğrencilerinin kanıt değerlendirmede temel yeterliliklere sahip oldukları fakat yetişkinlere nazaran öncül inanışlarından ve mükemmel uyumun olmadığı eksik verilerden daha çok etkilendikleri rapor edilmiştir.

Ruffman vd. (1993), 4-7 yaş arasındaki çocukların gerçek ve sahte kanıtlardan yararlanarak farklı hipotezler oluşturup oluşturamayacaklarını belirlemeye çalışmışlardır. Çocuklara yeşil yiyecek türünün dişleri döktüğü, kırmızı yiyecek türünün ise dişleri sağlıklı yaptığına ilişkin mükemmel uyumun yer aldığı gerçek kanıtlar sunulmuştur. Ardından çocuklara hangi tür yiyeceğin dişleri döktüğü

sorularak çocukların doğru cevabı bulmaları istenmiştir. Sonrasında gerçek kanıtlar çocukların önünde tam tersi şeklinde tekrar düzenlenerek bu kez kırmızı yiyecek türünün dişleri döktüğü sahte bir kanıt deseni hazırlanmıştır. Hazırlanan kanıt deseninin gerçek olmadığı çocuklara hatırlatıldıktan sonra, orijinal kanıt deseni hakkında bilgi sahibi olmayan kurgusal bir karaktere sahte kanıtlar sunulmuştur. Çocuklardan sahte kanıtların hangi hipotezi desteklediğine ilişkin sorular sorulmuştur. Araştırma bulgularına göre 5 yaş grubu çocukların hesaplanan şans başarısının üstünde (0.25) bir performans sergiledikleri ve hipotez-kanıt arasında bir ayrıma gittikleri fakat yapılan ayırımın gerekçelendirmeleri istenildiğinde istenen başarıyı gösteremedikleri görülmüştür. Çeşitli kanıt değerlendirme görevlerine ilişkin yapılan ve okul öncesi grubun yer aldığı farklı çalışmalarda da çocukların mükemmel ve kısmi uyum içeren kanıt desenlerinde başarılı oldukları ancak öncül inanışların tersine kanıtlar devreye girdiğinde çocukların görevi anlamakta ve kanıtlarla başa çıkmakta zorlandıkları rapor edilmiştir (Koerber vd., 2005; Piekny ve Maehler, 2012; Piekny vd., 2013).

Kanıt değerlendirmenin küçük yaş grubu çocuklarına uygun olmadığı, çocukların öncül inanışlarından oldukça etkilendikleri, kendilerine sunulan kanıta rağmen bu inanışı sürdürdükleri ve hipotez-kanıt arasında ayırım yapabilecek beceriye sahip olmadıkları yönünde araştırma bulgularına da rastlanmaktadır (Bullock ve Ziegler, 1999; Kuhn vd., 1988;1995; 2000; Kuhn ve Pearsall, 2000). Kuhn vd. (1988), 3. 6. ve 9. sınıf öğrencileri ile kanıt değerlendirme üzerine farklı görevlerin tasarlandığı kapsamlı bir dizi araştırma gerçekleştirmişlerdir. Birbirleri ile birtakım kovaryasyona dayalı kanıtlar katılımcılara ardışık ve toplu olarak sunulmuş ve katılımcılardan kanıtları değerlendirmeleri beklenmiştir. Gelen yanıtlar araştırmacılar tarafından kanıta ya da teoriye (hipoteze) dayalı olarak kodlanmıştır (Zimmerman, 2007). Araştırma bulguları genel olarak incelendiğinde hipotez-kanıt ayırımını fark etme, koordine etme ya da öncül inanışları terk ederek kanıtları değerlendirebilme becerisinin orta çocukluk döneminden (3-6.sınıflar) ergenliğe doğru (9.sınıf) geliştiği görülmektedir. Katılımcıların hipotez ve kanıt arasında herhangi bir uyumsuzluk olması durumunda, her ikisini uyumlu hale getirmek için çeşitli stratejiler geliştirdikleri; hipoteze uydurmak için kanıtları görmezden geldikleri, kanıtları çarpıttıkları, ön yargılı bir şekilde kanıt seçtikleri, hipotezi kanıta uydurmaya çalıştıkları ya da hipoteze uygun kanıtlar ürettikleri rapor edilmiştir. Kanıt üretilen

görev tipinin kovaryansa dayalı görev tiplerinden ayrıldığı düşünülmektedir. Bu görev tipinde hipotez ve deney sonuçları katılımcılara verilmiş, katılımcılardan deney sonuçlarını değerlendirerek hipoteze uygun kanıtlar üretmeleri beklenmiştir (Kuhn vd., 1988). Bu bağlamda kanıt değerlendirmede hipotezin kabulü ya da reddine ilişkin ikili yanıt sisteminden uzaklaşıldığı söylenebilir.

Batan nesneler üzerine alana özgü bilgi ve bilimsel süreç becerilerinin etkileşimine yönelik 10, 12 ve 14 yaş grubu toplam 30 öğrenci üzerinden bir araştırma gerçekleştirilerek, öğrencilerin öncül inanışlarının kendi tasarladıkları deneyler aracılığı ile ortaya çıkarılması hedeflenmiştir (Penner ve Klahr, 1996). Araştırmada şekil (küp, küre), boyut (büyük, küçük) ve malzeme türü (çelik ve teflon) farklılık gösteren nesneler kullanılmıştır. Ardından katılımcılara nesnelerin neden ve nasıl battıklarına ilişkin bilgiler verilmiş, nesnelerin farklı kombinasyonları oluşturularak hangi nesnenin daha hızlı batacağına ilişkin katılımcılardan tahminde bulunmaları ve nesneleri en hızlı batandan en yavaş doğru sıralamaları istenmiştir. Sonrasında katılımcılardan nesnelerin batmaları üzerine hipotezler geliştirmeleri ve hipotezi test etmeleri beklenmiştir. Araştırma sonuçlarına göre tüm yaş gruplarında çocukların %86-88 oranında nesnenin batma süresini yalnızca ağırlığı ile ilişkilendirdikleri görülmüştür. Diğer taraftan büyük yaş grubu öğrencileri ağırlığın dışında farklı özellikleri de araştırmışlardır. Bu bağlamda ağır olan nesne hızlı batan öncül inanışın katılımcılar arasında yaygın bir görüş olduğu söylenebilir. Araştırma bulguları genel olarak incelendiğinde katılımcıların yaşla birlikte öncül inanışlarını terk etmeye eğilimli oldukları, büyük yaş grubu öğrencilerinin kurdukları hipotezi test etmede daha tutarlı oldukları ve daha sistematik bir yaklaşım belirledikleri raporlanmıştır (Penner ve Klahr, 1996).

Kanıt değerlendirme becerisinde kullanılan bir diğer yaklaşımın *verileri analiz etme ve yorumlama* olduğu belirtilmekte, sonuca ulaşabilmek için veri setinde gerekli ve gereksiz veri desenlerini birbirlerinden ayırt edebilme becerisi olarak tanımlanmaktadır (Klahr vd., 2019). Katılımcıların verileri yorumlama becerilerinin değerlendirildiği bir çalışmada 39 üçüncü sınıf, 44 altıncı sınıf ve 50 üniversite öğrencisi olmak üzere toplam 133 katılımcı üzerinden iki deneysel araştırma eş zamanlı ve art arda yürütülmüştür (Masnick ve Morris, 2008). Sunulan veri seti; örneklem büyüklüğü, grup içi ve gruplar arası değişkenlik bağlamında ele alınmıştır.

Araştırma sonuçlarına göre tüm katılımcıların veri setlerinde örneklem büyüklüğü farklarına duyarlı oldukları, altıncı sınıf ($F(3, 38) = 3.33, p = .030$, kısmi etki büyüklüğü $\eta^2 = .21$) ve üniversite öğrencileri için ($F(3, 44) = 50.59, p < .001$, kısmi etki büyüklüğü $\eta^2 = .78$) güven değerlendirme ölçeğinde veri seti büyüklüğüne bağlı olarak anlamlı artışlar elde edilirken; üçüncü sınıf öğrencileri için bir azalma eğilimi gösterdiği rapor edilmiştir ($F(3, 34) = 2.26, p = .099$, kısmi etki büyüklüğü $\eta^2 = .17$). Büyük yaştaki katılımcıların veri setlerini bir bütün olarak daha iyi yorumladıkları, daha iyi ayırım yapabildikleri ve veri özelliklerinin sayısında yaşa bağlı olarak önemli artışlar kaydedildiği raporlanmıştır.

Üniversite ve 9-11 yaş arasındaki öğrencilerden oluşan bir katılımcı grubu ile bir sarkacın salınımını etkileyebilecek faktörlerin belirlenmesi amacıyla katılımcıların deneyler yaptıkları, veri topladıkları, kaydettikleri ve yorumladıkları bir araştırma gerçekleştirilmiştir (Masnick vd., 2016). Araştırmada katılımcıların sarkacın salınımını etkileyen faktörlere ilişkin sahip oldukları öncül inanışları ile deneylerden elde ettikleri veriler sonrasında bu inanışları revize edebilme becerileri değerlendirilmiştir. Ön test araştırma bulgularına göre çocukların %85.7 ve yetişkinlerin % 73.1 salınımında yay uzunluğunun nedensel olduğuna ilişkin öncül bilgiye sahip oldukları ve deney sonrası düşüncelerini revize ederek hem çocuklarda hem de yetişkinler bu oranların %100'e çıktığı görülmüştür. Kavanoz ağırlığı ve kavanozun bırakılma açısı değişkenleri bağlamında yetişkinlerin elde edilen veriler sonrasında öncül inanışlarını revize etmede daha başarılı oldukları, yetişkinlerin yapılan deneylerden daha çok öğrendikleri ve daha iyi yorum yapabildikleri görülmüştür (Masnick vd., 2016).

Açıklamalar oluşturma kanıt değerlendirme becerisini ölçmede kullanılan bir diğer stratejidir. Kanıtlara yönelik açıklamalar oluşturmaya Kuhn'un (1988, 1989, 2005) çalışmalarında sıklıkla rastlanmaktadır. Kuhn'a göre kanıtlar ve kanıtlara ilişkin hipotezlerin ayırımının yapılabilmesinde bilişsel ve üstbilişsel becerilere ihtiyaç duyulmaktadır. Kuhn, kanıt ve hipotez arasındaki koordine durumun sağlanabilmesinde; kanıt ve hipotezlerin ayı ayrı kodlama ve temsil edilebilme yeteneği, hipotez ve açıklamaları bağımsız bir şekilde ele alabilme yeteneği, hipotezlerin ya da açıklamaların yanlış olduğu durumları fark edebilme yeteneği olmak üzere üç üstbilişsel becerinin etkili olduğunu savunmaktadır.

Kuhn ve Pearsall (2000) çalışmalarında 4 ve 6 yaş çocuklarına iki koşucunun yarıştığı bir dizi görsel sunmuşlardır. Resimlerde yarışmacılardan kimin kazanacağına ilişkin birtakım ipuçları yer almakta (şık koşucu ayakkabılarının olup olmaması) ve son resimde yarışmacılardan birinin elinde bir kupa tuttuğu ve gülümsediği bir sonuç bulunmaktadır. Çocuklara yarışmayı kimin kazandığı sorulmuş ve akabinde şu iki soruya cevap vermeleri beklenmiştir: “*Nasıl biliyorsun?*” ve “*Neden öyle düşünüyorsun?*”. Hipotez - kanıt ayırımını 4 yaşındaki çocukların ancak üçte birinin yapabildiği ve çocukların gerekçelendirme noktasında oldukça zayıf kaldıkları görülmüştür. Şöyle ki çocukların büyük bir bölümü bunu nasıl biliyorsun sorusuna verdikleri yanıt “*yarışmacı kupayı elinde tutuyor*” cevabından ziyade “*çünkü onun hızlı ayakkabıları var*” şeklindedir. Araştırmacılara göre hipotez kanıt arasındaki ayrımları fark etmede 4 yaş grubuna kıyasla 6 yaş grubu daha yüksek performans göstermişlerdir.

Literatürde hem çocukların hem de yetişkinlerin açıklama oluşturmaya meyilli oldukları hatta bebeklerin dâhi neden sonuç ilişkisi kurmaya yatkınlıklarının olduğu söylenebilir (Piaget, 1929). Koslowski’nin (1996) yaptığı çalışmalar bireylerin, kanıtlar arasında bağlantıyı açıklayacak nedensel bir mekanizmanın bulunması halinde nedensel çıkarımlar yapmaya eğilimli olduklarını göstermektedir. Bireylerin rasyonelliğe dayalı neden-sonuç mekanizmalarını kurmaları durumunda, var olan kanıtları tekrar değerlendirerek görüşlerini değiştirmek yerine öncül inanışlarını sürdürdükleri bu durumun yalnızca çocuklarda değil yetişkinlerde de kendini gösterdiği belirtilmektedir (Chinn ve Brewer, 1998; Chinn ve Malhotra, 2002; Koslowski vd., 2008).

Kanıt değerlendirmeye yönelik yapılan araştırmalar genel olarak değerlendirildiğinde, bu becerinin küçük yaş gruplarında mükemmel uyumunun varlığı ya da yokluğu temel alınarak uygulanabildiği ve başarılı sonuçlar alındığı görülmüştür. Diğer taraftan kanıtlara ilişkin kısmi uyum ya da uyumsuzluğun olduğu durumlarda kanıt değerlendirme becerisinin söz konusu yaş grubu için uygun olmadığına ilişkin bulgulara ulaşılmıştır. Verilerin analizi, yorumlanması ve kanıtlara yönelik açıklamalar oluşturma yaklaşımlarının kanıt değerlendirme görevi dışında farklı bilişsel ve üstbilişsel becerilere duyulan ihtiyaçtan dolayı, küçük yaş gruplarından ziyade büyük yaş grubu öğrencilerde ve yetişkinlerde

uygulanabilecekleri söylenebilir. (Klahr vd., 2019; Kuhn vd., 1988; Zimmerman, 2007). Kanıtlar arasındaki mükemmel uyuma dayalı kanıt değerlendirme görevleri için hipotezin kabulü ya da reddine ilişkin olarak, evet/hayır ve var/yoka dayalı dikotomik bir yapıya sahip olduğu bu yüzden küçük yaşlarda uygulanabildiği düşünülmektedir. Literatürde kanıt değerlendirmeye ilişkin çoğul düşünmeye dayalı yaklaşımın temel alındığı çalışmaların ise sınırlılığı göze çarpmaktadır. Kuhn vd. (1988) çalışmalarında kanıt üretmeye dayalı bir görev tipi geliştirmelerine rağmen üretilen kanıtların oldukça sınırlı sayıda kaldığı ve çok sayıda kanıt üretmeye uygun görevler olmadığı görülmektedir.

2.3 Bilimsel Yaratıcılığın Gelişimi

Yaratıcılık yalnızca bilişsel süreçleri içeren bir olgu değil sosyal-duygusal özelliklerin, aile yapısının, eğitimsel dinamiğin de yer aldığı çok boyutlu bir yapı olarak açıklanabilir. Bu bağlamda yaratıcılığın gelişiminde bilişsel, sosyal ve duygusal süreçlerin, aile ortamının, formal ve informal eğitsel hazırlığın, alan ve disiplin özelliklerinin, sosyo-kültürel bağlamın ve tarihsel olgu ve olayların etkili olduğu belirtilmektedir (Feldman, 1999). Benzer şekilde yaratıcılığın yaş (Simonton, 1988; 1997), bilgi edinimi (Sternberg ve Lubart, 1995; Weisberg, 1999) ve eğitim düzeyi (Simonton, 1984) gibi faktörlerden etkilendiği ifade edilmektedir (Sak ve Maker, 2006). Söz konusu faktörlerin ise yaratıcılığı doğrusal olmayan bir yörüngede etkilediği (Anderson, 1992; Sak ve Maker, 2006; Simonton, 1997), yaratıcılığın kimi zaman zirve yaparken kimi zaman ani düşüşler yaşadığı, hatta zaman zaman zirve ve düşüşlerin birden fazla şekilde kendini gösterdiği belirtilmektedir (Charles ve Runco, 2001).

2.3.1 Bilgi

Yaratıcılığın altında yatan düşünce mekanizmasını ortaya çıkarmak için psikometrik çalışmalardan (Guilford, 1950), Darvinci yaklaşımlara (Campell, 1960; Simonton, 1999), sosyal-psikolojik perspektiften (Amabile, 1983b), yatırım yaklaşımına (Sternberg ve Lubart, 1993) geniş yelpazede teoriler geliştirildiği görülmektedir. Weisberg (1999), her ne kadar geliştirilen teorilerin farklı bakış açılarına sahip oldukları düşünülse de teorilerin yaratıcılık ve bilgi arasında ortak bir varsayıma dayandıklarını savunmaktadır. Yazarın kendi deyimiyle yaratıcılık ve bilgi arasında bir gerilim (tension) olduğu, bu olumsuz kavramın son zamanlarda terkedilerek daha

olumlu bir yaklaşımla, yaratıcılığın bilgiye dayandırıldığı (Hayes, 1989; Kulkarni ve Simon, 1988) ve *temel görüşe* (foundation view) evrildiği belirtilmektedir (Weisberg, 1999). Çoğu araştırmacı temel görüşü destekleyici bir şekilde, yaratıcılıkta alana özgü bilgi ve becerinin kritik önemi üzerinde durmuşlardır (Alexander, 1992; Amabile, 1996; Feldhusen, 2006; Urban, 1991). Feldhusen (2002), yaratıcılığın boşlukta kendiliğinden gerçekleşemeyeceğini, geçmiş öğrenmelere ve bilgiye önemli ölçüde bağlı olduğunu ifade ederek yaratıcılık için bilginin ne kadar kritik bir role sahip olduğu konusunda görüş bildirmiştir. Bilim genel olarak, insan ve evren ile ilgili gözlem ve deneylerden elde edilen, çeşitli olgular, fenomenler, yasalar ve teorileri barındıran bilgi birikimi şeklinde tanımlanmakta (Chalmers, 1999; Popper, 1959) ve bu bağlamda bilimsel yaratıcılık için de bilgi birikiminin önemi üzerinde durulmaktadır (Hu ve Adey, 2002; Huang ve Wang, 2019).

Lise öğrencileri ile gerçekleştirilen bir çalışmada fen bilimleri alanında okul puanları yüksek olan öğrencilerin okul puanı düşük olan öğrencilere kıyasla bilimsel yaratıcılık puanlarının anlamlı düzeyde ($p<.01$) daha yüksek oldukları raporlanmıştır (Sharma, 2015). Mohamed (2006) öğrencilerin bilimsel yaratıcılıkları ve öğretmenlerin KDT tekniği ile değerlendirdikleri fen bilimleri içerik bilgisi arasında .42 düzeyinde korelasyon katsayısı bulmuştur. Bulgu öğrencilerin fen bilimleri alanında sahip oldukları bilgi, birikim ve becerilerin bilimsel yaratıcılıkları ile orta düzeyde bir ilişkiye işaret etmektedir. Türkiye’de gerçekleştirilen benzer bir çalışmada 145 ortaokul öğrencisinin bilimsel yaratıcılık puanları ile fen ve teknoloji dersi akademik başarı puanları arasındaki ilişki farklı değişkenler açısından irdelenmiştir (Ayverdi vd., 2012). Araştırmadan elde edilen bulgulara göre öğrencilerin bilimsel yaratıcılık puanları ile fen ve teknoloji dersi akademik puanları arasında .356 düzeyinde korelasyon katsayısı hesaplanmıştır. Bu bağlamda fen ve teknoloji dersinde yüksek puana sahip öğrencilerin bilimsel yaratıcılıklarının da görece yüksek olduğu söylenebilir.

Çoğul ve tekil düşünmenin yer aldığı bilimsel yaratıcılık testleri ile açık uçlu ve çoktan seçmeli sorulardan oluşan bilimsel araştırma becerilerinin ölçüldüğü bir başka çalışma 3. 4., 5. ve 6. sınıflardan toplam 321 öğrencinin katılımı ile Tayvan’da gerçekleştirilmiştir (Yang vd., 2016). Araştırmada Hu ve Adey’in (2002) çoğul düşünmeye dayalı SSCM testi, Sternberg ve Lubart’ın (1995) çalışmalarına dayalı

olarak tasarlanmış iki adet açık uçlu tekil düşünme envanteri, 7 maddeli açık uçlu (Cuevas vd., 2005) ve 24 maddeli çoktan seçmeli (Burns vd., 1985; Yang vd., 2016) bilimsel araştırma beceri testleri kullanılarak sınıf düzeyleri arasında bir farklılık olup olmadığı incelenmiştir. Bilimsel araştırma beceri testleri araştırma sorusu oluşturma, hipotez kurma, araştırma planlama ve sonuç çıkarma becerileri kapsamında fen bilimlerine yönelik alan bilgisini ölçmeyi hedeflemektedirler. Araştırma bulgularına göre, bilimsel araştırma beceri puanlarının bir üst sınıf lehine anlamlı şekilde daha yüksek oldukları, en düşük puan ortalamalarının 3.sınıflarda görüldüğü, çoğul ve tekil düşünmeyi kapsayan bilimsel yaratıcılık puanları ile bilimsel araştırma beceri puanları arasında .21 ile .49 arasında değişen korelasyon katsayıları rapor edilmiştir. Araştırmacılara göre Tayvan'da fen bilgisi eğitiminin 3.sınıfta başlamasından dolayı 3.sınıf öğrencilerinin, 4., 5. ve 6.sınıf öğrencilerine kıyasla alan bilgilerinin ve bilimsel araştırma yeterliliklerinin daha sınırlı oldukları düşünülmektedir.

Hu vd. (2013), Düşünmeyi Öğren (Learn to Think) adını verdikleri müdahale programının ortaokul öğrencilerinin bilimsel yaratıcılıkları üzerine etkisini incelemek için ön test ve son teste dayalı deneysel bir çalışma gerçekleştirmişlerdir. Müdahale programı genel olarak öğrencilerin bilimsel bilgi ve becerilerini arttırmaya, bilimsel problemlere yaratıcı çözüm önerileri getirmeye, fen bilimleri alanında temel bilgi ve yasaları öğrenmeye yönelik tasarlanmıştır. Bir buçuk yıllık süre sonunda müdahale programının uygulandığı deney grubunun bilimsel yaratıcılık puanlarının kontrol grubunda bulunan öğrencilerin bilimsel yaratıcılık puanlarından anlamlı derecede ($p<.01$) daha yüksek oldukları ve aradaki farkların .72 ile .84 arasında geniş etki büyüklüğüne sahip oldukları rapor edilmiştir (Hu vd., 2013). Literatürde incelenen çalışmalar genel olarak değerlendirildiğinde alan bilgisinin kaçınılmaz bir şekilde yaratıcılığı olumlu yönde etkilediği söylenebilir.

2.3.2 Yaş

Sanatçı ya da bir bilim insanının yaratma süreci ile bir çocuğun büyüüp olgunlaşma sürecindeki gelişimi arasında benzerlikler olduğu düşünülmektedir (Sawyer, 2012). Piaget (1932), çocukların farklı evrelerden geçerek düşünme stillerinde niteliksel bir değişim olduğunu iddia etmiştir. Runco (2007), Piaget'in adaptasyon teorisini yaratıcılıkla ilişkilendirerek ister çocuk ister yetişkin olsun adaptasyon sürecinde bir

problemin ihtiyacına işaret etmiştir (Runco,1999b). Çocukların okul öncesi dönemde yaratıcı yetenekleri daha kolay ortaya çıkabilir çünkü yetişkinlere kıyasla bu dönemde çocuklar daha samimi ve doğal olma eğilimindedirler (Moran vd., 1983). Benzer şekilde Rosenblatt ve Winner (1988) çocukların sanatsal yaratıcılıklarının estetiksel özellikler bağlamında okul öncesi dönemde yüksek seviyede olduğunu, orta çocukluk döneminde giderek azaldığını raporlayarak bu döneme *geleneksel* adını vermişlerdir. Araştırmacılara göre yalnızca görsel sanatlar alanında yeteneği olan öğrenciler ilkokul yıllarında öğrenilen kuralları ihlal etmeye ve *geleneksel sonrası* aşamaya geçebilmektedirler. U şeklinde gelişim modeli olarak tanımlanabilecek (Strauss ve Stavy, 1982) bu süreçte, okul öncesi öğrencilerinin geleneklerden yoksun olmaları, kendini alanında yetiştirmiş yetişkin sanatçıların ise gelenekleri reddedebilmelerinden dolayı iki grubun birbirleri ile benzerlik gösterdikleri savunulmaktadır (Rosenblatt ve Winner, 1988).

Simonton (1988), yaratıcı üretkenliğin yaşa bağlı değişimini ters J ile ifade etmiş, 30'lu yaşların sonu ve 40'lı yaşların başında yaratıcı üretkenliğin zirve yaptığını ifade etmiştir. Yazara göre yaratıcı üretkenlik sonraki dönemde düşüş eğilimi göstererek zirve yaptığı seviyesinden yaklaşık yarısı kadar azalmaktadır. Smith ve Carlsson (1983; 1985; 1990) yaptıkları bir dizi çalışmada çocukların okula başladıktan sonra 7-8 yaşlarında yaratıcılıkta ani bir düşüş olduğunu, 10-11 yaşlarında yaratıcılıkta zirvenin yakalandığını, 12 yaşından sonra 7-8 yaşları kadar olmasa da yaratıcılığın yine azaldığını, ergenlik dönemi sürecince yaratıcılıkta kademeli olarak bir artış sağlandığını ve 16 yaşında ikinci bir zirvenin yaşandığını bulmuşlardır. Çin'de yapılan bir çalışmada ise yaratıcılık puanlarının 10-16 yaşları arasında genel bir azalma eğiliminde olduğu, 12-14 yaşları arasında ise yaratıcılık puanlarındaki azalmanın belirgin bir şekilde arttığı raporlanmıştır (Yi vd., 2013). Söz konusu çalışmadan elde edilen sonuçların doğu ve batı kültürleri bağlamında farklılaştığı söylenebilir.

Bilimsel yaratıcılık alanında yapılan çalışmalarda da yaratıcılığın yaşa bağlı gelişiminde benzer sonuçların alındığı ve gelişimin doğrusallıktan ziyade dalgalanmalar gösterdiği ifade edilebilir. Hu vd. (2004) Çin'den 11-15 yaş arası 1087 ve İngiltere'den 12-18 yaş arası 1190 öğrencinin katılımı ile gerçekleştirdikleri çalışma bulgularına göre, bilimsel yaratıcılığın yaşa bağlı olarak artış gösterdiğini,

14 yaşından itibaren azalma eğilimine girdiğini, yaratıcılık puanlarının zirve yaptığı yaş aralıklarının ise 11-13 ve 14-16 arasında olduğunu raporlamışlardır. Hu ve Adey (2002) geliştirdikleri bilimsel yaratıcılık testinde 12-15 yaş arası 160 ortaokul öğrencisi ile gerçekleştirdikleri çalışmada, 12 yaşındaki öğrencilerin bilimsel yaratıcılık puanlarının 13 yaşındaki öğrencilerden anlamlı derecede daha düşük olduğunu ($p=0.01$), 15 yaş grubunun 13 yaşa göre puanlarının daha yüksek fakat $p=0.05$ düzeyinde anlamlılık sağlandığını raporlamışlardır. Araştırmacılara göre öğrencilerin bilimsel yaratıcılıkları yaşa bağlı artış göstermekle birlikte bu artış doğrusal olarak ilerlememektedir (Hu ve Adey, 2002).

2.3.3 Sınıf

Yaratıcılığın, yaş gibi sınıf düzeyi bağlamında da doğrusal bir yol izlemediği aksine bazı dönemlerde zirve ve ani düşüş yaşadığı konusunda araştırma bulgularına rastlanmaktadır. Torrance (1968) boylamsal çalışmasında, 4. sınıf sonrasında öğrencilerin akıcılık, esneklik, orijinallik ve detaylandırma puanlarında %45 ile %61 arasında keskin bir düşüş yaşandığını, öğrencilerin yalnızca %11 ile %38 arasında yaratıcılıkları kapsamında bir artış gözlemlendiğini rapor etmiştir. Torrance, *dördüncü sınıf düşüşü* olarak adlandırdığı bu fenomeni okullardaki katı ve disipline dayalı eğitim anlayışının öğrenci yaratıcılığını engellediğini savunmuştur. Söz konusu fenomen sonraki dönemlerde farklı araştırmacılar tarafından da test edilmiştir. Yapılan araştırmalarda kimileri dördüncü sınıf fenomenini doğrularken (Camp, 1994; Raina, 1980; Saggar vd., 2019; Smith ve Carlsson, 1983) kimileri dördüncü ve altıncı sınıf sonrasında bir düşüş yaşandığını (Alacapınar, 2013, Lau ve Cheung, 2010) kimileri ise fenomene yönelik tutarsız sonuçlar alındığını raporlamışlardır (Barazza vd., 2019; Charles ve Runco, 2001; Claxton vd., 2005; Kiehn, 2003; Sak ve Maker, 2006, Salı, 2019). Yapılan araştırmalarda belirli bir sınıf düzeyinden bağımsız olarak öğrencilerin yaratıcılıklarında bazı dalgalanmalar yaşandığı ancak genel olarak sınıf bazında yaratıcılığın artış eğiliminde olduğu ifade edilmektedir. Yaratıcılıkta dördüncü sınıf düşüşü üzerine yapılan bir meta analiz çalışmasında, 1-12. sınıflar arasında genel olarak yaratıcılığın artış gösterdiği, dördüncü sınıfta yaratıcılığın düşüşüne dair yeterli kanıt bulunamadığı, aksine düşüşün yedinci sınıftan sonra gerçekleştiği bulgularına ulaşılmıştır (Said-Metwaly vd., 2021). Runco'ya (1996) göre gelişim döneminde yaşanan dalgalanmaların bir düşüş ya da gerileme olarak

nitelendirilmesi hatalıdır. Torrance'ın (1968) ifade ettiđi gibi yaratıcılıktaki sert düşüşün sebebi yapılandırılmış okul sistemlerinden değil geleneksel sonrası döneme geçiş için bir gereklilik, bir olgunlaşma süreci olarak görölmektedir (Gardner, 1982; Winner, 1982).

Bilimsel yaratıcılık alanında da dördüncü sınıf düşüşünün (Ayas, 2017) ve sınıf düzeylerinde bazı dalgalanmaların yaşanmasına ilişkin çalışma bulgularına rastlanmıştır (Hu vd; 2010; Hu ve Han, 2006; Kanlı, 2014). Ayas (2017) BÜT 3-5 formunun psikometrik özelliklerini incelediđi çalışmada hipotez geliştirme ve testin tamamından elde edilen puanların üst sınıflar lehine anlamlı farklılık gösterdiğini ancak deney tasarlama alt testinde 3. ve 4. sınıf puanları arasında istatistiksel olarak anlamlı bir fark olmadığını raporlamıştır. Hu vd. (2010) öğrencilerin 5.sınıfta, bilimsel yaratıcılık puanlarında önemli bir sıçrama yakaladıklarını, 8.sınıfa kadar yaratıcılık puanlarını tutarlı bir şekilde arttırdıklarını ve 8.sınıfta zirveye ulaştıklarını, 9-11.sınıflar arasında ise yaratıcılık puanlarının sabit kaldığını rapor etmişlerdir. Yang vd. (2016), 3-6. sınıf öğrencilerinden oluşan 321 kişi ile gerçekleştirdikleri çalışmada, 3.sınıf öğrencilerinin bilimsel yaratıcılık puanlarının diğer sınıflardan anlamlı derecede daha düşük olduklarını bulmuşlardır. Aynı araştırmada 4.sınıf öğrencilerinin 5.sınıf öğrencilerine kıyasla daha yüksek puan aldıkları, 6.sınıf öğrencilerinin ise 5.sınıf öğrencilerinden kısmen daha yüksek bilimsel yaratıcılık puanına sahip oldukları rapor edilmiştir. Çeliker vd. (2015), 484 ortaokul öğrencisinin bilimsel yaratıcılık düzeyleri ve fen öğrenimine yönelik tutumlarını inceledikleri çalışmada beş ve altıncı sınıf öğrencilerinin, yedinci ve sekizinci sınıf öğrencilerinden daha yüksek fen motivasyonlarının olduğunu; beşinci sınıf öğrencilerinin, altıncı, yedinci ve sekizinci sınıf öğrencilerinden daha yüksek bilimsel yaratıcılık puanlarına sahip olduklarını rapor etmişlerdir. Ayverdi vd. (2012), 6-8.sınıf arası 145 öğrenci ile gerçekleştirdikleri çalışmalarında, 6 ve 7. sınıf öğrencilerinin bilimsel yaratıcılık puanlarının 8. sınıf öğrencilerinden anlamlı derecede daha yüksek olduğunu, 7. sınıf öğrencilerinin bilimsel yaratıcılık puanlarının 6. sınıflara kıyasla daha yüksek olmasına rağmen aradaki farkın istatistiksel olarak anlamlı olmadığını tespit etmişlerdir. Araştırmacılara göre 8. sınıf öğrencilerinin bilimsel yaratıcılıklarının düşük olmasının sebebi liselere giriş sınavı ve bu sınavın öğrencilerde yarattığı kaygı ve endişedir. Bilimsel yaratıcılık çalışmalarında incelenen sınıf değişkenine ilişkin sonuçlar genel olarak

değerlendirildiklerinde genel yaratıcılık çalışmalarına benzer biçimde bilimsel yaratıcılık ve sınıf düzeyinin çoğunlukla doğrusal olmayan bir çizgide, zirve ve düşüşler yaşayarak ilerledikleri söylenebilir.

2.4 Bilimsel Yaratıcılık ve Cinsiyet

Cinsiyet, bireysel yaratıcılıktaki farklılıklardan biri olarak kendini göstermiş (Nakano vd., 2021) ve bir asır süredir tartışılan bir konu haline gelmiştir (Özdemir ve Sak, 2013). Cinsiyet farklılıklarının nedenlerini açıklamaya yönelik farklı teoriler ortaya atılmış, konu hakkında çok sayıda derleme makaleler yazılmış (Abraham, 2016; Baer ve Kaufman, 2008, Kogan, 1974; Nakona vd., 2021; Taylor vd., 2024) ve meta analiz çalışmaları yapılmıştır (Alabbasi vd., 2025; Hora vd., 2021; Thompson, 2016). Yaratıcılıkta cinsiyete dönük farklılıklar sosyo-kültürel ve çevresel faktörler (Ai, 1999; Romo, 2018), toplumsal cinsiyet rolleri (Baer ve Kaufman, 2008), her bir cinsiyet tarafından algılanan çeşitli engeller (Morais ve Almeida, 2019), genetik faktörler (Vernon, 1989), nörobilimle ilişkili çalışmalar (Abraham, 2016) ve Erkeklerde Daha Büyük Değişkenlik Hipotezi (Greater Male Variability Hypothesis-GMVH) şeklinde daha bütünsel teoriler ile açıklanmaya çalışılmıştır (He ve Wong's, 2011). Yaratıcılık alanında yapılan araştırma bulguları; kadınların daha yüksek puan aldıkları çalışmalar (Dudek et al., 1993; Kaufman vd., 2010; Kim ve Michael, 1995; Krumm et al., 2014; Lin ve Wong, 2014), erkeklerin daha yüksek puan aldıkları çalışmalar (Gralewski ve Karwowski, 2013; Kiehn, 2003; Lau ve Cheung, 2015; Matud vd., 2007; Stoltzfus vd., 2011), cinsiyetler arasında bir farkın olmadığı (Beghetto vd., 2011; Chae, 2003; Chen vd., 2002; Dikici, 2014; Kousoulas, 2010; Leu ve Chiu, 2015; Xiong vd., 2015) çalışmalar ve değerlendirilen kritere özgü farklılaşmaların bulunduğu çalışmalar şeklinde özetlenebilir (Bart vd., 2015; Cheung ve Lau, 2013; He ve Wong, 2011; Kim ve Michael, 1995; Madjar vd., 2011; Runco vd., 2010; Sternberg, 2006). Çoğul düşünmeye dayalı yaratıcılık testlerinin faktör yapıları ve belirlenen değişkenler üzerinden ölçüm değişmezliklerinin test edildiği sınırlı sayıda çalışmaya rastlanmaktadır. Bu çalışmalardan bazılarının cinsiyet değişkeni bağlamında gruplar arasındaki ölçme değişmezliğini sağlamaya çalıştıkları görülmektedir (Kim vd., 2006; Krumm vd., 2014; 2016; Miroshnik vd., 2022; Purc ve Laguna, 2019; Said-Metwaly vd., 2021).

Bilimsel yaratıcılık alanında yapılan cinsiyet arařtırmalarının yaratıcılık alanında yapılan arařtırmalara kıyasla daha sınırlı kaldığı söylenebilir. Özdemir ve Sak (2013), üniversite tabanlı bir üstün yetenek programına başvuran 704 altıncı sınıf öğrencisi ile cinsiyet farklılıklarını arařtırmak üzere bir çalışma gerçekleřtirmişlerdir. Çalışmada kullanılan bilimsel yaratıcılık testi; hipotez geliştirme, deney tasarlama ve kanıt değerlendirme bilimsel süreç becerileri ile akıcılık, esneklik ve bileşik yaratıcılıktan oluşan genel yaratıcılık becerilerini kapsamaktadır. Arařtırma sonuçlarına göre hipotez geliştirme becerisinde $t(693.762) = -3.090$, $p < 0.008$, $\eta^2 = 0.013$; akıcılık puanlarında $t(681.937) = -3.214$, $p < .008$, $\eta^2 = 0.014$ ve bileşik yaratıcılık puanlarında $t(691.453) = -3.047$, $p < .008$, $\eta^2 = 0.013$ erkek öğrenciler lehine anlamlı fakat düşük etki düzeyinde bulgular rapor edilmiştir. Diğer taraftan deney tasarlama, kanıt değerlendirme bilimsel süreç becerileri ile esneklik puanlarında kızlar ve erkek öğrenciler arasında anlamlı bir fark bulunamamıştır (Özdemir ve Sak, 2013). İlkokul, ortaokul ve liseden 1367 öğrencinin katılımı ile gerçekleştirilen, açık ve kapalı uçlu sorular ile öğrencilerin bilimsel anlamda yaratıcılığının incelendiği bir başka çalışmada akıcılık, esneklik ve orijinallik puanlarının kız ve erkek öğrenciler arasında anlamlı bir fark yaratmadığı görülmüştür (Hu vd., 2010). Xu vd. (2024) liseli öğrencilerin bilimsel yaratıcılıklarını ölçmek amacıyla Kapsamlı Bilimsel Yaratıcılık Testi'ni (Xu, 2013, Comprehensive Scientific Creativity Assessment, C-SCA) kullanmışlardır. C-SCA, bilimsel bilgi, bilimsel yaratıcılık motivasyonu ve çoğul ile tekil düşünmeyi barındıran düşünme stilleri olmak üzere 3 bileşenden meydana gelmektedir. Yapılan arařtırmada öğrencilerin bilimsel yaratıcılık motivasyonu kapsamında kız ve erkek öğrenciler arasında anlamlı bir fark bulunamamıştır. İki farklı versiyonu ile çoğul ve tekil düşünme stillerinin test edildiği çalışma bulgularına göre çoğul düşünme stilinde anlamlı bir fark elde edilmezken versiyonlardan birinin tekil düşünme alt bileşeninde kızlar lehine anlamlı bir farklılık rapor edilmiştir ($t = -2.303$, $p < 0.05$).

Üniversite öğrencilerinin bilimsel yaratıcılıklarını incelemek üzere gerçekleştirilen bir çalışmada, 59 öğrenciye 20 farklı çalışma başlığı ile özeti verilmiştir. Öğrencilerden değerlendirmeleri çalışmaları değerlendirmeleri ve potansiyel olarak tasarlayabilecekleri bir çalışma hakkında 3 soruya cevap vermeleri beklenmiştir (Sternberg vd., 2020). Öğrencilerden gelen cevaplar 1-5 arasında arařtırmacılar tarafından puanlanmıştır. Yapılan t testi analizleri sonucuna göre erkek ve kadınlar

arasında anlamlı bir farklılık elde edilememiştir $t(56) = -0.72, p = 0.478$. Kanlı (2017) bilimsel yaratıcılık ve bilimsel tutumlar arasındaki ilişkiyi incelediği çalışmada, erkek ve kız öğrenciler arasında anlamlı bir fark ortaya koyamamasına rağmen $t(56) = .864, p > .05$; $t(56) = -.380, p > .05$, kadınların fen bilimleri, matematik ve teknoloji alanlarında temsil oranlarının düşük olduğunu ve araştırmadan elde edilen bulgunun gerçek yaşamla örtüşmediğini rapor etmiştir. Ateşgöz ve Sak (2021), ÇABİYAT testi ile gerçekleştirdikleri çalışmada deney tasarlama ve hipotez geliştirme bilimsel süreç becerilerden oluşan 6 maddenin tamamının, kız veya erkek öğrenciler lehine bir yanlılık barındırmadıklarını raporlamışlardır. Dördüncü sınıf öğrencileri ile gerçekleştirilen Hu ve Adey'in (2002) bilimsel yaratıcılık testinin kullanıldığı bir başka çalışmada da (Baysal vd., 2013) kız ve erkek öğrencilerin bilimsel yaratıcılık puanlarının birbirlerinden farklılaşmadığı $t(73) = -.308, p > .05$ bulgulanmıştır. Nakona vd.'nin (2021) derleme çalışmasında yaratıcı bilimsel problem çözme, mekanik ve bilimsel yaratıcılık alanlarında erkeklerin kadınlara göre daha iyi oldukları sonucuna ulaşılmıştır.

2.5 Bilimsel Yaratıcılığın Ölçülmesi

Yaratıcılık değerlendirmelerinin genel yaratıcılık anlayışından alana özgülüğe kayması (Amabile, 1983a; Baer, 1998; 2015; Kaufman ve Baer, 2008), bilimsel yaratıcılığın, yaratıcılığın bir türü sayılmakla birlikte edebiyat, sanat, müzik dallarındaki yaratıcılıklardan farklı olduğu görüşünün ön plana çıkması (Hu ve Adey, 2002; Lin vd., 2003; MacKinnon, 1975), bilimsel yaratıcılığın ölçülmesinde bilimsel bilgi, beceri ve deneyimin bilimsel keşif süreç ve sonucunun önemli birer parçaları olmaları (Heller, 2007; Klahr, 2000; Klahr ve Dunbar, 1988) bilimsel yaratıcılık testlerinin geliştirilmesinin nedenleri arasında sayılabilir. Bilimsel yaratıcılığın değerlendirilmesinde genel yaratıcılık ölçümünde kullanılan benzer yaklaşımların benimsendiği söylenebilir. Geliştirilen bilimsel yaratıcılık testleri incelendiğinde kimi araştırmacıların çoğul düşünmeye dayalı testler geliştirdikleri (Aschauer vd., 2022; Ayas ve Sak, 2014; Ateşgöz ve Sak, 2021; Hu ve Adey, 2002), kimi araştırmacıların çoğul ve tekil düşünmeyi birlikte kullandıkları (Lubart vd., 2011; Vries ve Lubart, 2019; Xu vd., 2024; Yang vd., 2016), kimilerinin öz değerlendirme ölçümleri ile bilimsel yaratıcılığı ölçmeye çalıştıkları (Carson vd., 2005; Liang, 2002; Zhu vd., 2019), bazı araştırmacıların proje temelli ölçümler ile (Eichenberger,

1978; Kimbell, 2012; Nilada vd., 2024) ve bazılarının da konsensüse dayalı değerlendirme tekniğini kullanarak (Atakaya vd., 2022; Jo, 2009; Mohamed, 2006) bilimsel yaratıcılığı ölçmeyi hedefledikleri görülmektedir.

Genel yaratıcılığın ölçülmesinde olduğu gibi bilimsel yaratıcılığın ölçülmesinde de birtakım sınırlılıklar yaşandığı söylenebilir. Öz değerlendirmeye dayalı ölçeklerin geçerlik ve güvenirlik çalışmalarının yeterli olmaması, ölçümde sübjektifliğin baskın olması (Kaufman vd., 2008; Lee vd., 2002) ve kültürel değerler barındırmaları (Hu ve Adey, 2002) söz konusu sınırlılıklar arasında sayılabilir. Proje temelli ölçümlerin uzun, zaman alıcı ve takibinin zor olması bu yaklaşımın sınırlılıkları olarak ifade edilebilir (Fairchild ve Hunter, 2014).

Bilimsel yaratıcılık testlerinin geliştirilme sürecinde Guilford (1950) ve Torrance'ın (1967) çoğul düşünmeye dayalı çalışmalarının büyük bir iz bıraktığı görülmektedir. TabloÇalışma kapsamında çoğul düşünmeye dayalı geliştirilmiş bir bilimsel yaratıcılık testinin revize edilmesi hedeflendiği için literatürde bu kapsamda var olan testler ayrıntılı bir şekilde ele alınmıştır. Çoğul düşünmeye dayalı geliştirilen bilimsel yaratıcılık testleri ilk dönem testleri ve kuramsal temellere dayandırılarak geliştirilen testler bağlamında değerlendirilebilir.

2.5.1 Öncül bilimsel yaratıcılık testleri

Bilimsel yaratıcılık testlerine ilişkin ilk örneklerle 1970'li yılların başından itibaren rastlanmaktadır. Majumdar'ın (1975) geliştirdiği bilimsel yaratıcılık testi doğrudan söz konusu yaratıcılığı ölçmeyi hedeflediği için alanın ilk örneklerden biri olduğu söylenebilir. Guilford'un SOI modeli temel alınarak, matematik, fizik, kimya ve biyoloji alanlarında Hintçe olarak geliştirilen test toplamda 29 madden oluşmakta ve çeşitli yaratıcı yanıtları barındırmaktadır. Test temel olarak dünyanın güncel ve gelecekteki olası sorunlarına yaratıcı çözümler bulmayı hedeflemiş ve ortaokul ile lise öğrencilerine uygulanmıştır. Majumdar'ın geliştirdiği teste benzer şekilde, Hindistan'da Sinha ve Singh (1987) tarafından Hintçe-İngilizce olarak toplam 84 maddeden oluşan bir başka bilimsel yaratıcılık testi geliştirilmiştir. Test ortaokul öğrencileri için geliştirilmiş olup ölçülmesi hedeflenen bileşenler arasında esneklik, yenilik, hayal gücü, gözlem yetenekleri, analiz ve dönüşüm becerileri yer almaktadır. Shukla ve Sharma (1987) ortaokul öğrencilerine yönelik dört alt test ve 12 maddeden,

Gupta (1988) 9. ve 10. sınıf öğrencilerine yönelik 30 maddeden ve Misra (1986), 11. ve 12. sınıf öğrencilerine yönelik 6 alt test ve 9 maddeden oluşan çoğul düşünme temelli bilimsel yaratıcılık testleri Hindistan’da geliştirilen diğer testlerdir (Mohamed, 2006; Mukhopadhyay, 2013). Bu dönemde Hindistan dışında Friedlander (1983), 143 lise öğrencisi ile veri toplama, problem çözme, hipotez geliştirme ve deney tasarlama becerilerinin yer aldığı bir başka bilimsel yaratıcılık testi geliştirmiştir. Araştırmacı geliştirilen testin bilim alanında yaratıcı öğrencileri belirlemede kullanılabileceğini iddia etmiştir.

İlk geliştirilen bilimsel yaratıcılık testleri genel olarak değerlendirildiğinde testlerin, geçerlik ile güvenirlik çalışmalarının yapılmadığı, elde edilen sonuçların tatmin edici olmadığı ve kültürel öğelere bağlı kaldıkları rapor edilmektedir (Hu ve Adey, 2002; Mohamed, 2006). Geliştirilen testlerin kuramsal ve teorik temellere sahip olmamaları, yaratıcılık kuram ve modelleri ile organik bir bağ kurulmadan bilimsel yaratıcılığın ölçülmeye çalışılması önemli sınırlılıklardan biri olarak görülebilir. Ayrıca testlerin hipotez geliştirme, deney tasarlama ve kanıt değerlendirmeyi içeren bilimsel süreç becerileri ile ilişkilendirilmediği ifade edilmektedir (Mohamed, 2006). Testlerin hedef gruplarının ortaokul ile lise öğrencileri olması ve küçük yaş grubu öğrencilerinin çalışmalara dahil edilmemeleri testlerin kapsayıcı olmadıkları şeklinde yorumlanabilir. 2000’li yıllardan itibaren geliştirilen bilimsel yaratıcılık testlerinin söz konusu sınırlılıkların bir bölümünü ya da tamamını ortadan kaldırmayı hedefledikleri görülmüştür.

2.5.2 Kuramsal temelli bilimsel yaratıcılık testleri

Çoğul düşünmeye dayalı son dönem bilimsel yaratıcılık testlerinin sıklıkla, Hu ve Adey’in (2002) SSCM-Bilimsel Yaratıcılık Yapı Modeli (Chin ve Siew, 2015; Siew vd., 2014; Ramly vd., 2022) ile Klahr ve Dunbar’ın (1988) SDDS-Bilimsel Yaratıcılık İkili Arama Modeli temel alınarak (Ayas, 2017; Ayas ve Sak, 2014; Ateşgöz ve Sak, 2021) geliştirildikleri söylenebilir. Ayrıca Mednick’ in (1962) Çağrışımsal Modeli temel alınarak geliştirilmiş bir yaratıcılık testi ile (Kanlı, 2014), doğrudan yaratıcılık çalışmaları temel alınarak geliştirilmiş (Aschauer vd., 2022) farklı bilimsel yaratıcılık testlerine de rastlanmaktadır. Tablo 2.1’ de kuramsal temelli bilimsel yaratıcılık testlerine ilişkin özet bilgilere yer verilmiştir.

Tablo 2.1*Bilimsel yaratıcılık testlerine ilişkin bilgiler*

Bilimsel Yaratıcılık Testi	Kuramsal yapı/Boyut	Genel Yaratıcılık Becerileri	Sınıf Düzeyi	Güçlü Yönleri	Zayıf Yönleri
Ortaokul için bilimsel yaratıcılık testi (Hu ve Adey, 2002)	SSCM Tek faktör: Bilimsel Yaratıcılık	Akıcılık, esneklik, orijinallik,	Ortaokul öğrencileri	Kuramsal çerçeveye sahip olması	DFA eksikliği, farklı puanlama tekniklerinin kullanılması, alan bilgisinin yetersizliği
Beşinci sınıflar için bilimsel yaratıcılık testi (Siew vd., 2014)	SSCM İki faktör: Bilimsel problem çözme ve bilimsel yaratıcılık	Akıcılık, esneklik, orijinallik	Beşinci sınıflar	Kuramsal çerçeveye sahip olması, küçük yaş grubunun dahil edilmesi	Zayıf ve negatif faktör yüklemeleri, DFA eksikliği, Alan bilgisinin yetersizliği
Okul öncesi öğrenciler için görsel bilimsel yaratıcılık testi -FSCT (Chin ve Siew, 2015)	SSCM Tek faktör: bilimsel yaratıcılık	Akıcılık, orijinallik, detaylandırma başlıkların soyutluğu, kapanmaya karşı direnç	Okul öncesi öğrencileri	Kuramsal çerçeveye sahip olması, okul öncesi grubun dahil edilmesi	Örneklem büyüklüğü yetersizliği, Geçerlik ve güvenilirliğe ilişkin sınırlılıklar
Kimya alanında geliştirilmiş bilimsel yaratıcılık testi- CSCT (Ramly vd., 2022)	-	Akıcılık, esneklik ve orijinallik, ürünün işlevselliği	Ön lisans öğrencileri	Kuramsal çerçeveye sahip olması, belirli bir disipline odaklanması	Örneklem büyüklüğü, yetersizliği Yapı geçerliğinin test edilmemesi
Yaratıcı bilimsel çağrışımlar testi (Y-BÇT, Kanlı, 2014)	Mednick'in Çağrışımsal Düşünme Teorisi Üç faktör: çağrışımlar, analogik muhakeme ve analogik problem çözme	Akıcılık	Ortaokul öğrencileri	Kuramsal çerçeveye sahip olması, geçerlik ve güvenilirlik çalışmalarının yapılması,	Madde havuzu geliştirmede yaşanan sınırlılıklar, Alt boyutlarda faktör yüklerinin zayıf olması
DPAS (Aschauer vd., 2022)	Yaratıcılık çalışmaları İki faktör: DIET ve DIST	Akıcılık, Esneklik, CQ	Ortaokul ve 10.sınıf öğrencileri	Kuramsal çerçeveye sahip olması, geçerlik ve güvenilirlik çalışmalarının tatmin edici olması	Büyük yaş gruplarına yönelik tasarlanmış olması, kullanılan puan türlerinin akıcılık ile yüksek korelasyonu

Tablo 2.1 (devam) *Bilimsel yaratıcılık testlerine ilişkin bilgiler*

Bilimsel üretkenlik testi (BÜT-6-8, Ayas ve Sak, 2014; Sak ve Ayas, 2013)	SDDS ve yaratıcılık çalışmaları Tek faktör: bilimsel yaratıcılık	Akıcılık, Esneklik, CQ	6-8 sınıf öğrencileri	Bilimsel süreç becerilerinin dahil edilmesi, geçerlik ve güvenilirlik çalışmalarına detaylı bir şekilde yer verilmesi	SDDS modelinin 3 boyutlu yapısının (hipotez geliştirme, deney tasarlama ve kanıt değerlendirme) doğrulanamaması Kullanılan puan türlerinin akıcılık ile yüksek korelasyona sahip olması
Bilimsel üretkenlik testi (BÜT 3-5, Ayas, 2017)	SDDS ve yaratıcılık çalışmaları İki faktör: Hipotez geliştirme ve deney tasarlama	Akıcılık, Esneklik, CQ	3-5 sınıf öğrencileri	Ölçek geliştirme aşamalarının titizlikle uygulanması, küçük yaş gruplarının dahil edilmesi, testin sağlam psikometrik özelliklerinin sağlam olması	Kanıt değerlendirme bileşeninin analiz dışı bırakılması ve SDDS'in 3 faktörlü yapısının doğrulanamaması
Çocuklar için animasyonlu bilimsel yaratıcılık testi-ÇABİYAT (Ateşgöz ve Sak, 2021)	SDDS ve yaratıcılık çalışmaları İki faktör: Hipotez geliştirme ve deney tasarlama	Akıcılık, Esneklik, Orijinallik, CQ	Anasınıfı, birinci ve ikinci sınıf öğrencileri	Ölçek geliştirme aşamalarının titizlikle uygulanması, okul öncesi öğrencilerinin dahil edilmesi, öğrencilerin kısa animasyonlar ile değerlendirilmesi, testin sağlam psikometrik özelliklerinin sağlam olması	Kanıt değerlendirme bileşeninin analiz dışı bırakılması ve SDDS'in 3 faktörlü yapısının doğrulanamaması, Kullanılan puan türlerinin akıcılık ile yüksek korelasyona sahip olması,

SSCM: Bilimsel Yaratıcılığın Yapı Modeli. SDDS: Bilimsel Yaratıcılığın İkili Arama Modeli. DPAS: Çoğul düşünmeye dayalı problem çözme yetenek testi. DIET: Bilimsel görevlerde fikir üretimi. DIST: Deneysel görevlerde fikir üretimi.

2.5.2.1 Ortaokullar için bilimsel yaratıcılık testi

Test, Hu ve Adey'in (2002) önerdikleri SSCM (The Scientific Structure Creativity Model) modeli temel alınarak yine araştırmacılar tarafından ortaokul öğrencilerine yönelik geliştirilmiş bilimsel yaratıcılık testidir. SSCM modeli akıcılık, esneklik ve orijinalliğin bulunduğu *yetenek* özellikleri; teknik ürün, bilimsel bilgi, bilimsel problem ve bilimsel kavramın yer aldığı *ürün*; düşünme ile hayal gücünün yer aldığı *süreç* bileşenlerinden oluşmaktadır. Model içinde yer alan 24 farklı hücre (3*4*2) ve

hücre başına iki madde olmak üzere toplamda geliştirilen 48 madde, yüzeysel geçerliğin (face validity) sağlanması amacıyla Çin Halk Cumhuriyeti'nde görev yapan 50 ortaokul fen bilgisi öğretmeni ve fen eğitimi alanında araştırmacıya gönderilmiştir. Uzmanlardan madde havuzunda yer alan 48 maddenin hangilerinin ortaokul öğrencileri için uygun olduklarını değerlendirmeleri beklenmiştir. Uzmanların değerlendirmeleri sonucunda araştırmacılar belirledikleri 9 madde ile Londra'da tamamı kız öğrenci olan 60 kişi ile bir ön değerlendirme yapmışlardır. Ön değerlendirme sonucunda iki maddenin öğrenciler için zor olduğunu ifade eden araştırmacılar uygulamalarını 7 madde üzerinden sürdürmüşlerdir. Test grup olarak uygulanmakta ve yaklaşık olarak 60 dakikalık sürede tamamlanmaktadır.

Test maddeleri incelendiğinde 6 maddenin fen bilimleri alanında olağan dışı kullanımlar ile ilişkilendirilirken bir maddenin fen alanında yaratıcı bir ürün tasarlanması ile ilgili olduğu görülmektedir. Maddeler akıcılık, esneklik, orijinallik, hayal gücü, teknik ürün ve problem çözme becerilerini kapsamaktadır. İlk dört maddenin puanlanma süreci akıcılık, esneklik, orijinallik puanlarının toplamı üzerinden ilerlemektedir. Akıcılık, yanıtların niteliğine bakılmaksızın öğrencilerden gelen fikirlerin sayılması ile elde edilirken esneklik puanı ise farklı kategorilerde üretilen yanıtların toplamı ile hesaplanmaktadır. Orijinallik puanının hesaplanmasında frekansa dayalı yaklaşım benimsenerek; üst %5'lik dilim için 2 puan, %5-10'luk dilim için 1 puan ve %10'luk dilimden fazla olan fikirlere 0 puan verilerek hesaplanmıştır. 5. madde için yalnızca orijinallik puanı (üst %5'lik dilim için 3 puan, %5-10'luk dilim için 2 puan ve %10'luk dilimden fazla olan fikirlere 1), 6. madde için ise esneklik (araç 3 puan, prensip 3 puan, prosedür 3 puan) ve orijinallik puanları (üst %5'lik dilim için 4 puan, %5-10'luk dilim için 2 puan ve %10'luk dilimden fazla olan fikirlere 0 puan) hesaplanmıştır. Son olarak 7.maddede tasarlanacak olan ürünün her bir özelliği için 3 puan verilerek akıcılık puanı hesaplanırken, ürünün orijinalliği 1-5 arasında puanlanır.

Testin asıl uygulaması ortaokul öğrencilerinden 90'ı kız, 70'i erkek olmak üzere toplamda 160 kişiye uygulanarak gerçekleştirilmiştir. Madde analizleri kapsamında %27'lik alt ve üst madde ayırt ediciliği ve iç tutarlılık için maddeler arası korelasyon katsayıları hesaplanmıştır. Madde ortalamalarının katılımcıların üst %27'lik kısmı ile alt %27'lik kısmı arasında $p=.001$ düzeyinde anlamlı olarak farklılaştığı rapor

edilmiştir. Maddelerin birbirleri ve madde toplam puanı ile aralarındaki korelasyon katsayılarının .345 ile .829 arasında değiştiği görülmektedir. Güvenirlik analizleri sonucunda elde edilen Cronbach Alpha katsayısı $\alpha=.893$ olarak hesaplanmıştır. Testin geçerlik analizleri bağlamında PCA analizi ve yüz görünüm geçerliklerine bakılmıştır. PCA analizi sonuçlarına göre tek faktörlü bir yapı elde edildiği ve yapının toplam varyansın %63'ünü açıkladığı rapor edilmiştir. Maddelerin ortaokul öğrencilerine uygunluğu yüz görünüm geçerliği kapsamında değerlendirilmiş ve 35 fen bilgisi öğretmeni ve fen bilimlerinde araştırmacının görüşlerine başvurulmuştur. Yüz görünüm geçerliği sonuçlarına göre maddelerin ortaokul öğrencileri için uygunluğu konusunda uzmanlar, %60 ile %88 arasında görüş birliğine varmışlardır.

Ortaokullar için geliştirilmiş bilimsel yaratıcılık testi değerlendirildiğinde, testin kuramsal bir yapı üzerine inşa edilmesi ve geliştirilme sürecinde geçerlik ile güvenirlik analizlerine yer verilmesi bakımından güçlü olduğu söylenebilir. Fakat literatürde önerilen ölçek geliştirme adımlarına (DeVellis ve Thorpe, 2022) çalışmada tam olarak yer verilmediği söylenebilir. Madde havuzunun oluşturulma ve maddelerin seçilmesi sürecinde uzmanların nasıl belirlendiği detaylı bir şekilde anlatılmamıştır. Ön değerlendirme sürecinin yalnızca kız öğrenciler ile yürütülmesi olası madde yanlılığı konusunu tartışmaya kapadığı söylenebilir. Geçerlik analizleri değerlendirildiğine yalnızca PCA analizi ile ölçeğe ilişkin açımlayıcı bir faktör analizi yürütüldüğü fakat yapının doğrulanmadığı görülmüştür. Ölçek geliştirme çalışmalarında açımlayıcı ve doğrulayıcı faktör tekniklerinin bir arada kullanılması önerilmektedir (Kline, 2011; Tabachnick ve Fidell, 2013). Yüz görünüm geçerliğinin son test üzerinde uygulanarak rapor edilmesinin metodolojik bir sorun olduğu düşünülmektedir. Yüz görünüm geçerlikleri ölçek geliştirme aşamasında kullanılan ilk adımlardan biri olduğu söylenebilir (Boeteng vd., 2021; Haynes vd., 1995). Testte yer alan maddelerin farklı puanlama tekniklerine sahip olması da diğer bir sınırlılık olarak görülmektedir. Farklı puanlama tekniklerine başvurulması, testin bir bütün olarak yorumlanması noktasında istatistiksel birtakım sorunlara yol açabileceği düşünülmektedir. Test fen bilimleri alanında çoğul düşünme odaklı maddeler içermesine rağmen maddelerin bilimsel yaratıcılıkla ilişkilendirilen bilimsel bilgi ve beceriler bağlamında yetersiz kaldığı söylenebilir. Örneğin maddelerden birinin *sıradan bir bisikleti daha ilginç, daha kullanışlı ve daha güzel hale getirmek için*

mümkün olduğunca iyileştirilmesine yönelik maddenin formülize ediliş şeklinin ve olası yanıtların genel yaratıcılıkla ilişkili olduğu düşünülmektedir.

2.5.2.2 Beşinci sınıflar için bilimsel yaratıcılık testi

SSCM modelinin temel alındığı bir diğer bilimsel yaratıcılık testi beşinci sınıflara yönelik Siew vd. (2014) tarafından geliştirilmiştir. Testte birbiri ile eşdeğer 4'er maddenin yer aldığı A ve B formları oluşturulmuştur. Hedef grubu olan beşinci sınıf öğrencilerin bilimsel yaratıcılıkları teknik ürün, bilimsel bilgi, bilimsel olgu ve bilimsel problem çözme bileşenleri ile değerlendirilmeleri hedeflenmiştir. İlk maddede bir teknik ürün oluşturma kapsamında; A formu için öğrencilerin bir kalem ve B formu için öğrencilerin tükenmez kalem ile ilgili bilim alanında çok sayıda fikir üretmeleri ya da çizim yaparak tasarımları beklenmektedir. İkinci madde bilimsel bilgi bileşeni kapsamında A formu için mıknatıs ve B formu için mikroorganizmalar konusunda öğrencilerin bilimsel anlamda çok sayıda fikir üretmeleri beklenmektedir. Üçüncü madde bilimsel olgu bileşeni kapsamında A formu için *güneşin verdiği ışığın azalması* ve B formu için *bitkilerin hayvanlar gibi hareket etmesi* fenomenleri üzerinden öğrencilerin bilimsel hikayeler üretmeleri beklenmektedir. Son madde bilimsel problem çözme bileşeni kapsamında A formu için *bir kareyi 4 eşit parçaya bölme* ve B formu için *kibrit çöpleri kullanarak sembol üretme* üzerine çok sayıda fikir üretilmesi ve çizilmesi beklenmektedir. Testte yer alan ilk 3 madde için Hu ve Adey'in kullandıkları hali ile akıcılık, orijinallik ve esneklik puanları hesaplanmıştır. Son maddede ise esneklik ve orijinallik puanları hesaplanarak toplamaları alınmıştır. Test beşinci sınıf öğrencilerinden 115'i kız ve 91'i erkek olmak üzere toplamda 206 öğrenciye uygulanmıştır. Test A formu uygulandıktan bir hafta sonra B formu uygulanarak süreç tamamlanmıştır.

Test, madde ayırt edicilik analizleri kapsamında %27'lik alt ve üst dilimlere bölünmüştür. Analiz sonuçlarına göre A ve B formlarında birer maddenin yeterince ayırt edici olmadığı ve maddelerin revize edilebilecekleri önerilmiştir. Testin güvenirlik analizleri kapsamında iç tutarlılık analizleri yapılmıştır. A ve B formu için Cronbach Alpa katsayı değerleri sırası ile .77 ve .68 olarak hesaplanmıştır. Testin geçerlik analizleri için yapı geçerliği ve yüz görünümü geçerlikleri incelenmiştir. Yapı geçerliği kapsamında PCA analizi yapılmış ve iki form için de iki faktörlü bir yapı elde edilmiştir. Yapı, toplam varyansın A formu için %61,9'unu B formu için

%65'ini açıklamaktadır. Yüz görünüm geçerliği kapsamında 8 fen bilgisi öğretmenin görüşlerine başvurulmuştur. Her bir maddenin 5. Sınıf öğrencilerine uygunluğunu değerlendiren öğretmenler arasında yüksek derecede uyumun sağlandığı belirtilmiştir.

Beşinci sınıflar için geliştirilen bilimsel yaratıcılık testi değerlendirildiğinde, ortaokullar için geliştirilen testten farklı bir yapı elde edildiği görülmektedir. Testin A ve B formları için iki faktörlü bir yapı elde edilmesine rağmen faktör yükleri bağlamında değerlendirildiğinde A ve B formları arasında da bir farklılık olduğu göze çarpmaktadır. Örneğin B formunda yer alan ikinci maddenin negatif yük değerine (-0.706) sahip olduğu görülmektedir. Ayrıca yine B formunda yer alan üçüncü maddenin faktör yük değerinin alan yazında yaygın olarak kabul edilen .40 değerinden küçük olduğu görülmektedir. Bu bağlamda A ve B formları iddia edildiği gibi eşdeğer formlar kapsamında değerlendirilmesi bir sınırlılık olarak kabul edilebilir. Testin geliştirilme aşamalarının yeterince detaylandırılmadığı, madde seçiminin yalnızca öğretmenlerin görüşleri kapsamında sağlandığı, ana uygulama öncesinde herhangi bir ön değerlendirme yapılmaması testin diğer sınırlılıkları olarak düşünülebilir. Diğer taraftan geliştirilen bir bilimsel yaratıcılık testinin küçük yaş gruplarına hitap etmesi ve kuramsal temellere sahip olması testin güçlü yönleri kapsamında değerlendirilebilir.

2.5.2.3 Okul öncesi öğrenciler için görsel bilimsel yaratıcılık testi -FSCT

FSCT, SSCM modeli temel alınarak geliştirilmiş bir başka yaratıcılık testidir (Chin ve Siew, 2015). Test, bu modele dayanarak geliştirilmiş diğer yaratıcılık testlerinden uygulandığı hedef grubun okul öncesi öğrencileri kapsamı ve maddelerin görsel ağırlıklı olması sebebiyle ayrılmaktadır. Test, 2 bilimsel bilgi, 2 bilimsel problem ve 2 bilimsel olgu bileşenleri kapsamında toplam 6 maddeden oluşmakta ve 5-10 kişilik öğrenci gruplarına bir arada uygulanabilmekte ve yaklaşık 60 dakika sürmektedir. Katılımcılar henüz okuma yazma bilmeyen öğrencilerden oluştuğu için her bir maddenin öğrenciler tarafından çizimler yaparak yanıtlamaları beklenmektedir. Maddeler temel müfredata uygun olarak geliştirilmiş ve bir uzman grubu tarafından kontrolleri sağlanmıştır. 1., 3., 4. ve 6. maddeler için akıcılık, orijinallik, detaylandırma, başlıkların soyutluğu ve erken kapamaya direnç puanları

hesaplanmıştır. 2. ve 5. maddeler ise akıcılık, orijinallik, detaylandırma ve başlıkların soyutluğu puanları hesaplanmıştır.

Uygulama altı yaş grubundan 30 öğrenci ile yürütülmüştür. Madde ayırt edicilik analizleri öğrenci puanlarının alt ve üst %27'lik dilimlere ayrılarak yürütülmüş ve maddelerin literatürde kabul gören .20'nin üzerinde olduğu belirtilmiştir. Güvenirlik analizleri sonucunda maddeler arası ve madde toplam korelasyon değerlerinin .209 ile .866 arasında değiştikleri, Cronbach Alpha güvenirlik katsayısı ise $\alpha=.806$ olarak rapor edilmiştir. Güvenirlik analizleri kapsamında ayrıca puanlayıcılar arası uyuma bakılmıştır. Öğrencilerin görevlere verdikleri yanıtlar iki uzman tarafından bağımsız olarak puanlanmış ve uyum değerlerinin .780 ile .933 arasında değiştikleri görülmüştür. Testin yapı geçerliğini sağlamak amacıyla faktör analizi tekniklerinden PCA yaklaşımı benimsenmiş ve tek faktörlü bir yapı elde edilmiştir. Faktör yük değerlerinin .542 ile .867 arasında değişmekte ve tek faktörlü yapı toplam varyansın %50.8'ini açıklamaktadır. İçerik geçerliği kapsamında dört alanın uzmanının görüşlerine başvurulmuş, maddeler ve maddelerin ait olduğu bileşenler değerlendirilmiştir. Son olarak maddelerin okul öncesi öğrencilere uygunluğunu değerlendirmek üzere 10 fen bilgisi ve okul öncesi öğretmenin görüşleri alınarak elde edilen sonuçlar yüz görünüm geçerliği kapsamında değerlendirilmiştir.

FSCT testi değerlendirildiğinde bilimsel yaratıcılık literatüründe okul öncesi gruba uygulanması ve teorik olarak belirli bir perspektifi benimsemesi bakımından önemli olduğu söylenebilir. Diğer taraftan testin örneklem grubunun küçük olması yapılan geçerlik ve güvenirlik analizlerini sorgulanır hale getirmektedir. Araştırmalarda 30 ve altı çalışma grubuna sahip çalışmaların parametrik olmayan istatistiksel analizler yapılması önerilmektedir. Faktör analizlerinin yapılabilmesi için gereken minimum örneklem sayısına ilişkin çok sayıda ve farklı kriterler bulunmaktadır. Söz konusu kriterler incelendiğinde FSCT' nin geliştirilme sürecindeki katılımcı sayısının oldukça sınırlı kaldığı söylenebilir.

2.5.2.4 Kimya alanında geliştirilmiş bilimsel yaratıcılık testi- CSCT

CSCT, son yıllarda SSCM modeli temel alınarak kimya alanında ön lisans öğrencilerinin bilimsel yaratıcılıklarını ölçmek için geliştirilmiş testlerden biridir (Ramly vd., 2022). Test kimya alanında uzman 3 kişi ile SSCM testinde lise

öğrencilerine yönelik geliştirilen 7 maddenin ön lisans grubuna ve kimya alanına uyarlama sürecini kapsamaktadır. SSCM’ de yer alan örnek bir madde “*Lütfen bir camın olabildiğince çok sayıda bilimsel kullanımını yazın*” şeklinde iken araştırma kapsamında bu madde “*Asitlerin günlük kullanımına uygun çok sayıda fikir üretin*” şeklinde kimya alanına uygun bir maddeye dönüştürülmüştür. Testte yer alan ilk dört madde için akıcılık, esneklik ve orijinallik puanları; beşinci ve altıncı madde için esneklik ve orijinallik puanları, son olarak yedinci madde için orijinallik ve ürünün işlevselliğine dayalı puanlama yapılmıştır.

Testin geçerlik çalışmaları bağlamında içerik ve yüz görünüm geçerliği incelenmiştir. İçerik geçerliği için 3 kişilik bir ekipten oluşan alan uzmanları CSCT testinin geliştirilme sürecinde yer alan 11 maddeyi 1 (hiç ilişkili değil) ile 5 (çok yüksek ilişkiye sahip) arasında puanlayarak görüş bildirmişlerdir. Hesaplanan Kappa katsayılarına göre maddelerden biri hariç diğerlerinde alan uzmanlarının yüksek düzeyde görüş birliğine sahip oldukları görülmüştür. Araştırma kapsamında 10 alan uzmanına daha gidilerek yüz görünüm geçerliği değerlendirilmiş ve uzmanlardan test maddelerinin ön lisans öğrencilerine uygun olup olmadığı yönünde görüş bildirmeleri istenmiştir. Uzman değerlendirmeleri sonucunda testin içerik ve yüz görünüm geçerliklerinin sağlandığı görülmüş ve güvenilirlik analizlerine geçilmiştir. Güvenirlik analizleri kapsamında test -tekrar test analizi yapılmıştır. İlk ölçümün ardından iki aylık bir süre sonunda tekrar edilen testin Pearson korelasyon katsayısı hesaplanmış ve $r=.83$ olarak bulunmuştur. Araştırmacılara göre iki ölçüm arasında mükemmel bir ilişkinin varlığı CSCT testinin güvenilir olduğuna işaret etmektedir. Ön lisans öğrencilerine yönelik kimya alanında geliştirilen bilimsel yaratıcılık testinin örneklem büyüklüğü, ölçek geliştirme adımlarının detaylı bir şekilde raporlanmaması ve en önemlisi yapı geçerliğinin test edilmemesi bakımından yetersiz olduğu düşünülmektedir.

SSCM modeli temel alınarak geliştirilen bilimsel yaratıcılık testleri genel olarak değerlendirildiğinde bu testlerin geliştirme gerekçelerinde önceki bilimsel yaratıcılık testlerinin geçerlik ve güvenilirlik çalışmalarındaki sınırlılıkların yer aldığı görülmektedir. Fakat söz konusu testlerde de kısmi olarak bu sınırlılıkların devam ettiği söylenebilir. Öncelikle incelenen testlerin kuramsal yapılarının açıklanmaya çalışıldığı fakat doğrulanmadıkları görülmektedir. Madde havuzu oluşturma

sürecinin nasıl ele alındığı, kaç maddenin geliştirilmesi hedeflendiği, hangi maddelerin hangi kriterler bağlamında nasıl ve neden seçildikleri gibi kritik öneme sahip noktalara çalışmalarda yer verilmediği görülmektedir. Ayrıca çalışmaların tamamında yüz görünüm geçerliğinin ölçek geliştirme aşamasında nihai bir karar mercii olarak kullanılması metodolojik birtakım sınırlılıkları doğurabilir. Geliştirilen testler aynı modeli temel almalarına rağmen tek ve iki boyutlu yapıların varlığına rastlanmaktadır. Elde edilen sonuçlar modelin tam anlamıyla test edilemediğini ve yapısal anlamda, çalışmalarda bir tutarlılık sağlanamadığı şeklinde yorumlanabilir. Bu bağlamda geliştirilen testlerin yaygın ve evrensel bağlamda kullanımlarından ziyade bölgesel uygulanabilecek farklı tipte testler olarak kabul edilmeleri doğru bir tercih olabilir. Geliştirilen testlerin maddeleri detaylı incelendiğinde, hipotez geliştirme, deney tasarlama ve kanıt değerlendirme gibi bilimsel süreç becerilerinin işe koşulmadığı, aksine maddelerin yalnızca bilimsel terim ve kavramların genel yaratıcılık tipi bir teknik olan *alışılmadık kullanımlarına* ilişkin fikir ve ürün tasarımları ile ilişkilendirildikleri söylenebilir.

2.5.2.5 Yaratıcı bilimsel çağrışımlar testi (Y-BÇT)

Y-BÇT' nin kuramsal temelleri Mednick'in Çağrışımsal Düşünme Teorisi ve Sternberg ve Dunbar'ın analogiler üzerine yaptıkları çalışmalara dayandırılmaktadır (Kanlı, 2014). Araştırmacı testi geliştirmeden önce Mednick'in teorisi üzerinden Yaratıcı Bilimsel Çağrışımlar Modeli' ni geliştirmiş, ardından modeli kullanarak 25 maddeden ve 3 alt testten oluşan testi geliştirmiştir. Test ortaokul kademesi için uygundur. Testin *Çağrışımlar* alt boyutunda fizik, kimya, biyoloji ve çevre alanlarından 10 soru bulunmakta, öğrencilerden kendilerine verilen kavramlar yolu ile fen bilimleri alanında sınırlı kalmak koşulu ile çok sayıda çağrışım, bağlantı, ilişki üretmeleri beklenmektedir. *Analojik Muhakeme* alt boyutunda fizik, biyoloji, çevre, kimya-çevre, biyoloji-fizik ve kimya-biyoloji alanlarından 10 soru yer almaktadır. Öğrencilerden ilişki kurmaları, kurdukları ilişkileri aktarmaları, benzerlik ve analogiler yolu ile çok sayıda farklı fikir üretmeleri beklenmektedir. *Analojik Problem Çözme* alt boyutunda fizik, kimya-fizik ve biyoloji-çevre alanlarından 5 soru yer almakta ve öğrencilerden benzeşimler bulma, ilişki kurma, zihinsel haritalama yapma, bilinenlerden bilinmeyenlere ulaşma becerilerini kullanarak çok sayıda fikir

üretmeleri beklenmektedir. Y-BÇT' nin tüm alt testleri için yalnızca üretilen fikirler bağlamında akıcılık puanı hesaplanmıştır.

Test maddesi oluşturma ve seçme süreci kapsamında öncelikle araştırmacı tarafından geliştirilen 50 maddelik havuz değerlendirilmek üzere 4 uzmana gönderilmiş, yapılan düzenlemeler ve eklemeler sonrasında 37 soruluk ikinci bir form oluşturularak 13 kişiden oluşan farklı bir uzman grubundan maddelerin anlaşılabilirlikleri, madde zorlukları, maddelerin fen bilimleri ile ilgili olma durumları kriterleri üzerinden tekrar görüşleri alınmıştır. Uzmanlardan gelen dönütler sonrasında 29 maddelik üçüncü formla her sınıf düzeyinden 3'er (5-6-7-8) toplamda 12 öğrenci ile nitel verilerin toplandığı pilot uygulama yapılmıştır. Araştırmacının birinci pilot uygulama adını verdiği bu adımda öğrencilerden gelen dönütler sonrasında maddelerde revizyona gidilerek ve yeni maddeler yazarak 27 maddelik dördüncü form oluşturulmuştur. Her sınıf seviyesinden 40'ar öğrenci olmak üzere dördüncü form ile 160 kişi ile yürütülen ikinci pilot uygulamasında araştırmacı madde çıkarımı, yeni maddeler eklenmesi, yamaç birikinti grafiğinden ziyade teorik alt yapının temel alınarak faktör sayısının belirlenmesi gibi bir dizi karar almıştır. Alınan kararlar neticesinde 18 maddelik, 3 faktörlü yapının kuramsal çerçeveye uygun olduğu, maddelerin Cronbach Alpha katsayı değerinin .91 olarak hesaplandığı rapor edilmiştir. Araştırmacı maddelerin fen bilimlerine dağılımlarını ön planda tutarak çağrışımlar boyutuna 1 ve analogik muhakeme boyutuna 6 olmak üzere toplam 7 yeni soru daha yazarak 25 maddelik beşinci forma ulaşmıştır. Yazılan yeni maddeler, ilk adımlarda madde seçim sürecinde uygulanan süreçlere benzer şekilde uzmanlara gönderilerek görüşleri alınmış, sonuç olarak 25 madde üzerinden Y-BÇT' nin psikometrik özellikleri incelenmiştir.

Testin yapı geçerliği AFA ve DFA ile incelenmiş, AFA sonuçlarına göre 3 boyutlu bir yapı elde edilerek yapının toplam varyansa katkısının %47,4 olduğu belirtilmiş ve DFA ile bu yapı doğrulanmıştır. Y-BÇT' nin güvenirlik analizleri kapsamında 25 maddenin Cronbach Alpha katsayısı .88, puanlayıcılar arası uyum değerleri .969 ile .984 ve test-tekrar test güvenirlik katsayıları .804 ile .935 aralığında hesaplanmıştır. Testin sınıf bağlamında ayırt edicilik analizleri incelendiğinde ise bazı boyutlarda sınıflar arasında anlamlı farklılıkların yakalanamadığı rapor edilmiştir. Y-BÇT' nin

normal ve üstün yetenekli öğrencileri alt test ve toplam test bağlamında istatistiksel olarak anlamlı bir şekilde ayırt ettiği ifade edilmektedir.

Y-BÇT testi incelendiğinde farklı bir kuramsal yapı ile bilimsel yaratıcılığın ölçülmeye çalışması literatüre katkı sağlamakla birlikte madde geliştirme ve madde seçim sürecinin birtakım sınırlılıklara sahip olduğu söylenebilir. Madde havuzunun yeterince geniş oluşturulmaması, madde seçim sürecinde sürekli yeni madde yazılması zorunluluğunu doğurmuştur. Alan yazında testin son halinin verileceği madde sayısı için en az 3-4 katı kadar madde havuzunun oluşturulması önerilmektedir (DeVellis ve Thorpe, 2022; Kaplan and Saccuzo, 2009). Çalışmada AFA sonrası maddelerde revizyona gidilerek yeniden yazılan maddeler ile DFA yapıldığı görülmektedir. Yapı geçerliği kapsamında üç bileşenli yapının öngörüldüğü fakat analiz sonuçlarının iki bileşenli bir yapı önerdiği görülmüş, araştırmacı ise teorik gerekçeler sunarak 3 boyutlu yapı üzerinden analizlerine devam etmiştir.

2.5.2.6 Çoğul düşünmeye dayalı problem çözme yetenek testi- DPAS

Çoğul düşünmeye dayalı problem çözme yetenek testi (divergent problem solving ability- DPAS), yaratıcılık çalışmalarına dayalı, ortaokul öğrencilerinin bilimsel yaratıcılıklarını değerlendirmek için geliştirilmiş bir testtir (Aschauer vd., 2022). DPAS, problem çözme yeteneği ve fikir üretimi, çoğul düşünmeye dayalı olarak akıcılık ve esneklik kriterlerini barındırması, fizik ve kimya alanına özgü olmasına yönelik 3 temel varsayım kapsamında geliştirilmiştir. Test, fikir üretim alt boyutunda yer alan (DIST) 6 madde ve deneye dayalı fikir üretim alt boyutunda yer alan (DIET) 7 madde olmak üzere iki alt boyut ve toplam 13 maddeden oluşmaktadır. DIST alt testinde öğrencilerin alana özgü bilgilerini ölçmeye yönelik sorular yer alırken; DIET alt testinde alana özgü bilgi ve becerilerin deneyler yolu ile ölçülmesi hedeflenmiştir.

Testin puanlanmasında öğrencilerden gelen yanıtların uygunluğu değerlendirilerek akıcılık, esneklik ve akıcılık ile esnekliğin logaritmik dönüşümü ile elde edilen bileşik yaratıcılık puanları (CQ) kullanılmıştır. Çalışma Avusturya’da öğrenim gören 7-10.sınıf öğrencilerinden toplam 1106 katılımcı ile ön test ve son teste dayalı tekrarlı ölçüm tasarımı ile gerçekleştirilmiştir. Araştırmacılar, zaman kısıtlaması nedeniyle ölçeğin tamamından ziyade ön testte 8 madde (4 DIST ile 4 DIET), son testte 10 madde (5 DIST ile 5 DIET) olacak şekilde her bir maddeye 3'er dakikalık zaman

sınırlaması getirerek uygulama sürecini gerçekleştirmişlerdir. Öğrencilerden gelen yanıtlar iki uzman tarafından birbirlerinden bağımsız bir şekilde değerlendirilmiştir.

Testin yapı geçerliği DFA ile gerçekleştirilmiş olup iki faktörlü yapısının doğrulandığı ve uyum indekslerinin kabul edilebilir düzeylerde olduğu raporlanmıştır. Araştırmanın ölçüt geçerliği analizleri kapsamında Runco (2020) tarafından geliştirilen rCAP çoğul düşünme testinin iki alt testi; mekanik ve fen bilimleri alanında geliştirilmiş öz değerlendirme bataryası (K-DOCS; Kaufman, 2012); ayrıca öğrencilerin dil, matematik ve fen bilimleri okul puanları kullanılmıştır. DPAS esneklik ve CQ puanlarının çoğul düşünmeye dayalı iki alt test puanları ile $r=.713$ ile $r=.882$ arasında değişen korelasyon katsayıları hesaplanmıştır. DPAS ile alana özgü yaratıcılık öz değerlendirme ölçeği arasında küçük ve anlamlı olmayan korelasyon katsayılarına ulaşılrken DPAS ile öğrencilerin matematik-fen puanları arasında orta düzeyde, istatistiksel olarak anlamlı fakat negatif korelasyon değerlerinin raporlandığı görülmektedir ($r=-.301$ ve $r=-.362$). DPAS testinin uyum geçerliği bağlamında AVE değerlerinin ön test ve son test için .352 ile .486 arasında; yapısal güvenirlik katsayılarının ise .692 ile .827 arasında değiştikleri raporlanmıştır. AVE ve yapısal güvenirlik bağlamında elde edilen sonuçların literatürde kabul gören değerlerin altında olduğu söylenebilir (Fornell ve Larcker, 1981). Faktörler arası korelasyon katsayıları ön test sonuçları için $r = .824$; $p < .001$; son test sonuçları için $r = .850$; $p < .001$ olarak hesaplanmıştır. İki faktör arasında oldukça yüksek korelasyon katsayısı elde edilmesine rağmen bu durum faktörlerin birbirlerine oldukça benzer maddeleri barındırdığı şeklinde yorumlanabilir.

DPAS bilimsel yaratıcılık testinin yapı geçerlik analizi sonuçlarının tatmin edici seviyelerde olduğu ve testin alana özgü çoğul düşünmeye dayalı test edilmiş farklı bir değerlendirme aracı kazandırılması bakımından önemli olduğu düşünülmektedir. Araştırmacıların belirttiği gibi CQ puanının yaratıcılık çalışmalarında kullanılması diğer puan türleri ile yüksek korelasyon göstermesi bakımından dikkat edilmesi gereken bir nokta olarak ön plana çıkmaktadır. DPAS' ın ortaokul ve 10. Sınıf öğrencilerine yönelik geliştirilmiş olması bilimsel yaratıcılık literatüründe çalışan araştırmacıların büyük sınıflar ya da büyük yaş grupları ile çalışma trendini devam ettirdiklerini ve küçük yaş gruplarını göz ardı ettikleri şeklinde yorumlanabilir.

2.5.2.7 Bilimsel üretkenlik testi (BÜT-6-8)

BÜT' ün teorik çerçevesi SDDS modeli (Klahr, 2000; Klahr ve Dunbar, 1988) ve yaratıcılık çalışmalarına dayandırılmaktadır (Guilford, 1950, 1956, 1967; Torrance, 1962, 1988). BÜT 5 maddeden oluşan, grup olarak uygulanan ve yaklaşık olarak 40 dakika süren bir kâğıt kalem testidir. Testi, 6-8. sınıf öğrencilerine yönelik geliştirilmesine rağmen Millî Eğitim Bakanlığı tarafından 4+4+4 sistemine geçilmesi ile 5. Sınıf öğrencilerinden başlayarak uygulanmaya başlamıştır. Test hipotez geliştirme, deney tasarlama ve kanıt değerlendirme bileşenlerinin yer aldığı *bilimsel süreç boyutu*; akıcılık, esneklik ve bileşik yaratıcılık becerilerine yer verilen *yaratıcılık boyutu*; fizik, kimya biyoloji ve ekoloji bileşenlerinin yer aldığı *fen bilimlerine özgü* 3 oluşmaktadır. Testin puanlanmasında üretilen fikir sayısı *akıcılık*, farklı kategorilerdeki fikir sayısı *esneklik* ve akıcılık ile esnekliğin logaritmik dönüşümü alınarak *bileşik yaratıcılık* kullanılmaktadır (Snyder vd., 2004).

BÜT maddeleri 2 hipotez geliştirme, 2 deney tasarlama ve 1 kanıt değerlendirme bilimsel süreç becerileni kapsayan, fizik, kimya, biyoloji, ekoloji ve disiplinler arası alanlardan oluşmaktadır. Sinek maddesinde, öğrencilerden biyoloji alanında kendilerine sunulan deneyin amaçlarına yönelik hipotez geliştirmeleri beklenmektedir. Etkileşim maddesinde, öğrencilerden kendilerine sunulan disiplinler arası alandaki grafiğe ilişkin değişken çiftlerini düşünmeleri beklenmektedir. Madde hipotez geliştirme bilimsel sürecini ölçmeyi hedeflemektedir. Şeker maddesinde öğrencilerden kendilerine kimya alanında sunulan deney ve hipoteze uygun olarak deney düzeneğinde değişiklikler yapması beklenmektedir. Madde deney tasarlama sürecine yönelik olarak geliştirilmiştir. Deney tasarlamaya yönelik olarak geliştirilmiş Yay maddesi fizik alanında öğrencilere sunulan görsel ile ilgili olarak deneyin hedefine uygun olacak şekilde öğrencilerden düzenlemeler yapmaları beklenmektedir. Son olarak Besin Zinciri maddesi ekoloji alanında kanıt değerlendirme becerisini ölçmek amacıyla öğrencilere bir grafik sunularak, grafikte meydana gelen değişimin nedenlerine yönelik fikirler üretmesi beklenmektedir.

Testin geliştirilme sürecinde geniş bir madde havuzu oluşturularak ön değerlendirme, pilot uygulamalar ve alan uzmanlarından elde edilen veriler ışığında maddelerin son hallerine karar verilmiştir. Sak ve Ayas (2013) BÜT' ün psikometrik özelliklerini inceledikleri replika çalışmalarında 288 altıncı sınıf öğrencisinin verileri üzerinden

yürüttükleri PCA analizleri doğrultusunda tek faktörlü bir yapı elde etmişlerdir. Tek faktörlü yapı toplam varyansın %34,45'ini açıklamaktadır. BÜT' ün uyum geçerliği matematik alanında geliştirmiş bir yetenek testi ile gerçekleştirilmiştir. Uyum geçerliği kapsamında elde edilen bulgulara göre matematik alanında yüksek yeteneğe sahip öğrenciler (M1), ortalama matematiksel yeteneğe sahip öğrencilere kıyasla (M2) BÜT testinde daha yüksek puan almışlar ve aradaki puan farkı istatistiksel olarak anlamlı bulunmuştur [$M1= 12.60$, $SD= 4.62$; $M2=8.16$, $SD=4.29$; $t(286)=5.23$, $p<.000$]. Güvenirlik analizleri sonucunda yaratıcılık becerileri içeren akıcılık, esneklik, bileşik yaratıcılık ve toplam test Cronbach Alpha katsayı değerlerinin .848 ile .96 arasında değiştikleri rapor edilmiştir.

Daha büyük bir örneklem grubu ile BÜT' ün tek faktörlü yapısının doğrulanmasına ilişkin bir başka replika çalışma daha gerçekleştirilmiştir (Ayas ve Sak, 2014). 326 kız ve 367 erkek öğrenci olmak üzere araştırmaya katılan toplam 693 altıncı sınıf öğrencisinin verileri için ilk olarak madde analizlerine yer verilmiştir. Maddelerin tamamının alt %27, orta %46 ve üst %27'lik dilimlerde anlamlı bir şekilde ayırt edildiği görülmüştür. Çalışmada gerçekleştirilen geçerlik analizleri kapsamında yapı geçerliğini sağlamak amacıyla DFA uygulanarak, BÜT' ün tek faktörlü bir yapıya işaret ettiği ve model uyum indekslerinin mükemmel uyum gösterdikleri rapor edilmiştir. Çalışmada BÜT' ün uyum geçerliği matematik yetenek test ile test edilerek iki ölçek arasında toplam akıcılık, toplam esneklik ve toplam bileşik yaratıcılık puanları arasında .31 ile .59 arasında değişen korelasyon katsayılarına ulaşılmıştır. Ek olarak sınıf ve yetenek bağlamında ayırt edicilik geçerliği incelenmiş 7. sınıf üstün yetenekli öğrencilerinin BÜT puanları (M1), 6. sınıf üstün yetenekli öğrencilerinin BÜT puanlarından (M2) anlamlı bir şekilde daha yüksek bulunmuştur [$M1 = 12.41$, $SD = 5.63$; $M2 = 7.38$, $SD = 4.43$; $t(286) = 5.33$, $p < .000$]. Araştırmacıların yaptıkları güvenirlilik analizleri sonuçlarına göre test güvenirlilik Cronbach Alpha katsayısı .87 bulunurken puanlayıcılar arası güvenirlilik katsayı değerlerinin .87 ile .96 arasında değiştikleri görülmüştür. Testin, farklı kültürlere uyarlama çalışmalarında da tek faktörlü yapının doğrulanarak Cronbach Alpha güvenirlilik katsayısı $a=.705$ olarak hesaplanmış ve puanlayıcılar arası uyum değerlerinin .80 ile .98 arasında değiştiği raporlanmıştır (Bermejo vd., 2016).

BÜT testi bilimsel yaratıcılık alanında kapsamlı bir şekilde ele alınan SDDS modeline dayandırılması, büyük örneklem grupları ile açımlayıcı ve doğrulayıcı yöntemler ile test edilmesi, tatmin edici alt test ve toplam test güvenilirlik değerlerine sahip olması bakımından bilimsel yaratıcılık testleri bağlamında testin itici bir role sahip olduğu söylenebilir. Ayrıca testin farklı kültürlerle uyarlama çalışmaları yapılarak testi geliştiren araştırmacılar ile benzer sonuçlar alınması testin tutarlı sonuçlar ortaya koyduğu şeklinde yorumlanabilir. Diğer taraftan testin teorik çerçevede iddia edildiği üzere 3 boyutlu yapısının ortaya çıkarılamadığı görülmektedir. Araştırmalarda akıcılığın diğer puanlar üzerindeki kirletici etkisini azaltmak adına akıcılık ve esnekliğin logaritmik bir dönüşümü ile elde edilen bileşik yaratıcılık puanı hesaplanmıştır. Bileşik yaratıcılık puanı ile diğer puan türleri arasında .90 üzerinde korelasyon katsayı değerleri hesaplanmıştır. BÜT' ün kullanıldığı farklı araştırmalarda da yüksek korelasyon katsayılarına raporlandığı görülmektedir (Aschauer vd., 2022; Bermejo vd., 2016; Huang ve Wang, 2019). Bu bağlamda türetilen puan türünün, akıcılığın kirletici etkisinin azaltılmasında kısmi bir farklılık yarattığı söylenebilir.

2.5.2.8 Bilimsel üretkenlik testi (BÜT 3-5)

SDDS modeli temel alınarak 3. 4. ve 5. sınıf öğrencilerine uygun olarak geliştirilmiş bir diğer yaratıcılık testi BÜT 3-5 formudur (Ayas, 2017). BÜT (3-5) formu, BÜT' ün teorik çerçevesine benzemekle birlikte araştırmacı geliştirilen test için modelde revizyona giderek kanıt değerlendirme becerisini dışarıda bırakmıştır. Kanıt değerlendirme becerisinin teorik çerçeveden çıkarılmasına gerekçe olarak söz konusu bileşenin küçük yaş gruplarında değerlendirilmesinin uygun olmadığı, kanıt değerlendirmenin hipotez geliştirme ve deney tasarlama görevlerinde aracı bir değişken olarak görev aldığı ve bu bağlamda bağımsız bir beceri olarak modelde yer alamayacağı ve kanıt değerlendirmeye dayalı geliştirilen maddelerin çoğul düşünme görevlerine uygun olmadıkları şeklinde sıralanmaktadır. BÜT 3-5, BÜT 6-8 testine benzer şekilde akıcılık, esneklik ve bileşik yaratıcılık puanları hesaplanarak değerlendirilmektedir. Test, hipotez geliştirme (2 adet) ve deney tasarlama (2 adet) becerileri kapsamında fizik, kimya ve biyoloji alanlarında toplam 4 açık uçlu maddeden oluşan ve grup olarak da uygulanabilen bir kâğıt kalem testidir. BÜT 3-5,

her madde için öğrencilere en az yaklaşık 8 dakikalık süre verilerek tamamlanmaktadır.

Madde havuzunun oluşturulması 10 fen bilgisi ve 2 sınıf öğretmeni olmak üzere 12 kişilik bir ekiple yürütülmüş olup, araştırmacının liderliğinde 10 oturumluk bir sürecin ardından hipotez geliştirme ve deney tasarlama becerilerine ilişkin toplamda 118 adet maddenin geliştirilmesi ile tamamlanmıştır. Maddeler araştırmacı tarafından benzerlikleri ve hedef görevle ilişki düzeyleri bağlamında değerlendirilerek önce 85'e indirilmiş, ardından 11 kişiden oluşan bir uzman grubunun görüşleri alınarak, her birinde 6 madde olmak üzere (3 adet hipotez geliştirme ve 3 adet deney tasarlama) toplamda 12 maddelik iki adet taslak form oluşturulmuştur. 18 dördüncü sınıf öğrencisi ile yapılan ön değerlendirme sonucunda madde sayısı bu kez toplamda 10 maddelik ve her birinde 2 hipotez geliştirme ve 3 deney tasarlama olmak üzere iki form ile pilot uygulama gerçekleştirilmiştir. Pilot uygulamada yapı geçerliğinin sağlanması amacıyla PCA analizi gerçekleştirilmiş ve araştırmacının öngördüğü şekli ile iki faktörlü bir yapı elde edilmiştir. Araştırmacı madde betimsel istatistik ve PCA bulgularını göz önünde bulundurarak birtakım kararlar almış ve pilot uygulama sonrası madde sayısını 7'ye indirmiştir. 7 madde üzerinden yapılan PCA analizi sonuçlarına göre iki faktörlü yapı toplam varyansın %54,22 açıklamış ve Cronbach Alpha katsayısı .826 olarak hesaplanmıştır.

BÜT 3-5 asıl uygulaması gerçekleştirildikten sonra yapı geçerliğini sağlamak amacıyla PCA analizi tekrar edilerek madde sayısı 7'den 4' e indirilmiş (2 adet hipotez geliştirme ve 2 adet deney tasarlama) ve iki bileşenli bir yapı elde edilmiştir. İki bileşenin varyansa toplam katkısı %78,99 olarak açıklanmıştır. PCA bulguları ile desteklenen iki bileşenli yapı 211 kişinin verileri üzerinden DFA analizi ile incelenmiştir. Dört madde üzerinden gerçekleştirilen DFA analizi sonucunda iki bileşenli yapı doğrulanmış ve model uyum indekslerinin mükemmel olduğu rapor edilmiştir. Ayırt edicilik geçerliği kapsamında öğrencilerin akıcılık, esneklik, bileşik yaratıcılık puanları alt ve üst %27'lik dilimlere ayrılarak, ortalamalar üzerinden madde ve toplam test bağlamında karşılaştırılmıştır. Karşılaştırma sonucunda üst grubun puan ortalamalarının alt grubun puan ortalamalarından hem maddeler hem de toplam testte anlamlı bir şekilde daha yüksek oldukları bulunmuştur. BÜT 3-5'in uyum geçerliği için öğrencilerin ders notları ile BYTÖ (Bilime Yönelik Tutum

Ölçeği) ve FABAÖ (Fen Bilimleri Akademik Benlik Algısı Ölçeği) ölçekleri arasındaki ilişkiye bakılmıştır. BÜT 3-5 puanlarının öğrencilerin fen notları ile aralarındaki ilişki .310 ile .405 arasında; matematik notları ile aralarındaki ilişki .332 ile .409 arasında değişkenlik gösterirken; BYTÖ ile .238 ile .293 ve FABAÖ ölçeği ile .286 ile .297 arasında değişkenlik göstermiştir. Elde edilen bulgular BÜT 3-5'in uyum geçerliğine kanıt sunmuştur. BÜT 3-5'in güvenirlik analizleri kapsamında toplam test Cronbach Alpha katsayıları akıcılık, esneklik ve bileşik yaratıcılık için sırasıyla $\alpha=.743$, $\alpha=.701$, $\alpha=.736$ olarak hesaplanırken; okuyucular arası güvenirlik uyum değeri $r_{ICC}=.994$ olarak hesaplanmıştır. Araştırmacı güvenirlik analizlerine ek kanıt sunmak amacıyla düzeltilmiş madde toplam korelasyon ve madde ortalama korelasyon değerlerini de hesaplamış ve BÜT 3-5'in genel olarak güvenirliğinin tatmin edici seviyelerde olduğunu rapor etmiştir.

BÜT 3-5. sınıflar bilimsel yaratıcılık testi incelendiğinde ölçek geliştirme adımlarının büyük bir titizlikle yürütüldüğü söylenebilir. Testin son halinde yer alan 4 madde için çok geniş bir madde havuzundan seçim yapılmıştır. Madde havuzunun geniş tutulması, ön değerlendirme ve pilot uygulamalar sonucunda nihai maddelere karar verilebilmesi açısından büyük öneme sahip olduğu söylenebilir. SDDS modeli temel alınarak bilimsel süreç becerileri ve yaratıcı becerilerin çoğul düşünmeye uygun bir şekilde tasarlanması, BÜT 3-5'in küçük yaş grup öğrencileri için geliştirilmesi testin güçlü yanları olarak kabul edilebilir.

Diğer taraftan testin kuramsal çerçevesinden kanıt değerlendirme becerisinin çıkarılarak iki bileşenli bir yapı elde edilmesi, SDDS modelinin bütünsel olarak test edilemediğini göstermektedir. Ayrıca BÜT 6-8 de olduğu gibi BÜT 3-5 testinin puanlanmasında da akıcılığın esneklik üzerindeki kirletici etkisini azaltmak amacıyla akıcılık ve esnekliğin logaritmik bir dönüşümü ile elde edilen bileşik yaratıcılık puanı hesaplanmıştır. Araştırmada sunulan Pearson Momentler çarpımı korelasyon katsayıları incelendiğinde akıcılık ve bileşik yaratıcılık arasında .971-.990 ve bileşik yaratıcılık ve esneklik arasında .919-.977 düzeyinde anlamlı korelasyonlar katsayılarına rastlanmaktadır. Elde edilen sonucun akıcılığın kirletici etkisinin ortadan kaldırılamadığı ve bileşik yaratıcılık puanının çalışmaya metodolojik bir esneklik getiremediği şeklinde yorumlamak mümkündür.

2.5.2.9 Çocuklar için animasyonlu bilimsel yaratıcılık testi- ÇABİYAT

Test, SDDS modeli ve çoğul düşünmeye dayalı çalışmalar temel alınarak anasınıfı, birinci ve ikinci sınıf öğrencilerinin bilimsel yaratıcılıklarını ölçmek amacıyla geliştirilmiştir. ÇABİYAT maddelerinin animasyonlardan oluşması onu diğer bütün yaratıcılık testlerinden farklı yapmaktadır. ÇABİYAT için BÜT 3-5 testinin geliştirilme sürecinde yapılan revizyona benzer şekilde ve aynı gerekçeler sunularak kanıt değerlendirme bileşeni teorik çerçevenin dışında bırakılmış, kuramsal revizyon sonucunda hipotez geliştirme ve deney tasarlama bileşenlerinin her birinden 3' er adet olmak üzere açık uçlu sorulardan oluşan 6 maddelik bir test geliştirilmiştir. Maddelerin tamamı için akıcılık, esneklik, bileşik yaratıcılık ve ek olarak orijinallik puanları hesaplanmaktadır. Puanların hesaplanmalarına ilişkin detaylı bilgiye veri toplama araçları bölümünde yer verilmiştir.

ÇABİYAT' ın madde geliştirme sürecinde iki anaokulu öğretmeni, iki ilkökul öğretmeni, dört fen bilgisi öğretmeni ve bir araştırmacıdan oluşan dokuz kişilik bir ekip oluşturulmuş, ekip üyeleri teorik ve pratik eğitim ile 116 maddeden oluşan bir havuz geliştirmişlerdir. Ekip üyeleri maddeleri teorik çerçeveye uygunluk, netlik ve yaş gruplarına uygunluk kriterlerine göre değerlendirerek madde sayısını 116'dan 40'a indirmişlerdir. Daha sonra iki uzman maddeleri teorik önem bağlamında tekrar değerlendirerek madde sayısını 19'a indirmişlerdir. Geriye kalan 19 madde için animasyon hikayeleri yazılmış, yazılan hikayeler altı kişiden oluşan bir uzman grubu tarafından içerik geçerliği kapsamında değerlendirilmiş ve uzmanlar arasında tüm maddeler için %90'nın üzerinden bir uyum sağlanmıştır. Devam eden süreçte animasyon hikayeleri 15 anasınıfı öğretmenine elektronik ortamda gönderilerek hikayelerin anlaşılabilirliği, hedef gruba uygunluğu ve öğrencilerin maddelere verecekleri olası doğru yanıt sayıları üzerinden değerlendirmeleri beklenmiştir. Öğretmenlerin görüşlerine dayanarak 7 madde testten çıkarılarak 12 maddenin iki boyutlu animasyonları uzman bir animatör tarafından tasarlanmıştır.

Ön değerlendirme anasınıfı, birinci ve ikinci sınıf öğrencilerinden oluşan 20 kişilik bir öğrenci grubu ile 12 madde üzerinden gerçekleştirilmiştir. Ön değerlendirme sonucunda araştırmacı öğrencilerin zorlandığı bir maddeyi çıkartarak, pilot uygulama adımıyla 11 madde üzerinden toplamda 121 öğrenciye ulaşılmıştır. Öncül analiz bulgularına göre, maddeler arası korelasyon katsayılarının orta ile yüksek

düzeyde ilişkiye ($r > .50$) sahip oldukları, Cronbach Alpha katsayısının .91 olduğu ve düzeltilmiş madde toplam korelasyonlarının .51 ile .72 arasında değiştikleri rapor edilmiştir. Madde seçim sürecinde açımlayıcı faktör analizi yapılarak iki faktörlü bir yapı elde edilmiş, üç maddenin faktör yüklerinin bulaşıcı etkilerinden dolayı testten çıkarılmalarına karar verilmiş, kalan maddelerin toplam varyansın %57'sini açıkladıkları raporlanmıştır.

Dört hipotez geliştirme ve dört deney tasarlama olmak üzere toplam 8 madde üzerinden 801 kişi ile ana uygulamaya geçilmiştir. Ana uygulamada yapı geçerliğini sağlamak adına ilk olarak açımlayıcı faktör analizi yöntemlerinden Maksimum Olabilirlik tekniği belirlenmiş ve analiz sonucunda iki faktörlü bir yapı elde edilmiştir. Sekiz maddenin toplam varyansa katkısı %65,387 iken maddelerden ikisinin çapraz yük değerlerine sahip olmaları sebebiyle analizden çıkarılmalarının ardından toplam varyans değerinin %70,68'e çıktığı görülmüştür. Açımlayıcı faktör analizinin ardından 6 madde üzerinden yapılan doğrulayıcı faktör analizi sonucunda iki faktörlü yapının doğrulandığı ve model uyum değerlerinin ($\chi^2 /sd = 1.067$; RMSEA = .013; RMR = .045; SRMR = .016; NFI = .993; CFI = 1; GFI = .992; AGFI = .980) mükemmel düzeyde oldukları rapor edilmiştir. ÇABİYAT' ın yakınsak ve ıraksak geçerlikleri DFA' da elde edilen faktör yükleri üzerinden hesaplanan AVE ve CR değerleri ile incelenmiştir. Faktör yükleri .749 ile .858 arasında değişirken; AVE değeri hipotez geliştirme bileşeni için .676, deney tasarlama bileşeni için .635 ve aynı sıra ile CR değerlerinin .862 ve .839 şeklinde hesaplandıkları görülmüştür. İki bileşen arasındaki korelasyon katsayısı $r = .561$ ve $p < .01$ düzeyinde anlamlı bulunmuş, AVE değerlerinin kareköklerinin (.822 ve .797) faktörler arasındaki korelasyondan büyük oldukları görülmüştür. Bu durum hipotez geliştirme ve deney tasarımı bileşenlerinin birbirleriyle ilişkili ancak farklı yapılar olduğunu gösterdikleri şeklinde yorumlanmaktadır. Sonuçların ÇABİYAT testinin yakınsak ve ıraksak geçerliğine kanıt sundukları söylenebilir.

ÇABİYAT' ın gelişimsel geçerliği için alt testler ve toplam test üzerinden sınıf seviyeleri üzerinden gerçekleştirilen analizler sonucunda sınıf seviyeleri yükseldikçe tüm puan türlerinde istatistiksel olarak anlamlı bir şekilde artış sağlandığı görülmüştür. ÇABİYAT'ın uyum geçerliği SAGES 2 (Screening Assessment for Gifted Elementary and Middle School Students-Second Edition-Üstün Yetenekli

İlkokul ve Ortaokul Öğrencileri için Tarama Değerlendirme Bataryasının İkinci Baskısı) bataryası ile incelenmiş, iki test arasındaki korelasyon katsayılarının .49 ile .72 arasında değiştiği ve bu sonuçların ÇABİYAT'ın uyum geçerliği için kanıt oluşturduğu belirtilmiştir. Testin geçerliği kapsamında madde yanlılık analizleri bağımsız örneklem t testi kullanılarak gerçekleştirilmiş ve hiçbir maddedenin kız ya da erkek öğrenciler lehine bir yanlılık oluşturmadığı görülmüştür. ÇABİYAT'ın güvenilirliği, alt test ve toplam test bağlamında iç tutarlılık ve puanlayıcılar arası güvenirlik analizleri ile sağlanmıştır. Tüm puan türleri üzerinden alt test ve toplam test Cronbach Alpha katsayılarının .79 ve .88 arasında değiştiği, puanlayıcılar arasındaki uyum değeri ortalama katsayısının ise .98 olduğu rapor edilmiştir. ÇABİYAT, son yıllarda geliştirilen testler bağlamında değerlendirildiğinde oldukça sağlam kuramsal temele sahip olması, bilimsel yaratıcılık alanında önemli bir sınırlılık olarak görülen küçük yaş grupları için geliştirilmiş bilimsel yaratıcılık testlerinin yetersizliğine çözüm sunması, bir kâğıt kalem testinden ziyade animasyon tabanlı bir değerlendirme ortamı sağlaması, diğer testler bağlamında güçlü yanları olarak kabul edilebilir. ÇABİYAT'ın psikometrik özellikleri incelendiğinde geçerlik ve güvenirlik sonuçlarının mükemmel düzeyde oldukları görülmektedir.

SDDS modeli temel alınarak geliştirilen BÜT 6-8, BÜT 3-5 ve ÇABİYAT testleri genel olarak değerlendirildiğinde bilimsel yaratıcılık literatüründe sıklıkla ifade edilen testlerin teorik çerçeveden yoksun olması veya teorik çerçevelerinin zayıf olması konusundaki sınırlılığa sahip olmadığı, hatta güçlü bir kuramsal yapı içerdiği görülmektedir. Ayrıca bilimsel yaratıcılık testlerinin genel olarak ortaokul ve lise öğrencilerine yönelik geliştirilmesi konusunda birtakım eleştiriler olduğu bilinmektedir. BÜT 6-8'in orta okul öğrencilerine, BÜT 3-5'in üçüncü, dördüncü ve beşinci sınıf öğrencilerine ve ÇABİYAT'ın anasınıfı, birinci ve ikinci sınıf öğrencilerine yönelik olarak geliştirilmesi literatürde bilimsel yaratıcılık testlerine getirilen eleştirilerin dikkate alındığını göstermektedir. SDDS temelli testlerin aynı teorik çerçeveden beslenmelerine rağmen DFA sonuçlarında farklılıklar barındırdıkları, SDDS'ın 3 bileşenli yapısının ise bütünsel olarak doğrulanamadığı görülmektedir. BÜT' te tek faktörlü bir yapı elde edilirken; BÜT 3-5 ve ÇABİYAT'ın teorik çerçevesinde revizyona gidilerek hipotez geliştirme ve deney tasarlama olmak üzere iki bileşenli yapının farklı yaş gruplarında test edildikleri görülmektedir. Kanıt değerlendirme becerisi araştırmacıların iddia ettiğinin aksine farklı tip görevler

ile küçük yaş gruplarında da değerlendirilebilmektedir (Amsel ve Brock, 1996; Bullock ve Ziegler, 1999; Croker ve Buchanan, 2011; Koerber vd., 2005; Kuhn vd., 1988, 1995; Kuhn ve Pearsall, 2000; Piekny vd., 2013; Piekny ve Maehler, 2012; Ruffman vd., 1993). Araştırmacıların söz konusu becerinin kuramsal çerçevenin dışında bırakılmasına gerekçe olarak gösterdiği kanıt değerlendirme becerisinin SDDS modelinde hipotez geliştirme ve deney tasarlama becerileri arasında aracı ve örtük bir değişken olarak ele alınmasının ise yanlış yorumlandığı düşünülmektedir. Modeli geliştiren araştırmacılar (Klahr, 2000; Klahr ve Dunbar, 1988) üç becerinin birbirleri ile etkileşimleri sonucunda bilimsel keşif sürecinin tamamlandığını ifade ettikleri görülmektedir.

Sonuç olarak; yaratıcılığın ölçülmesi genel yaratıcılığın değerlendirilmesinden sıyrılarak alana özgülüğe kaymış (Amabile, 1983a; Baer, 1998; 2015; Kaufman ve Baer, 2008), edebiyat, sanat ve matematik alanlarında olduğu gibi fen bilimleri alanında da bilimsel yaratıcılığın ölçülmesine (Hu ve Adey, 2002; Lin vd., 2003; MacKinnon, 1975) yönelik çalışmaların yapılmasına hız kazandırmıştır. İlk dönem bilimsel yaratıcılık testleri Hindistan’ da sıklıkla Guilford’ un Zihin Yapı Modeli ve Torrance’ın yaratıcılık çalışmaları temel alınarak farklı yaş grubu öğrencilerine Hinduca ve İngilizce olarak geliştirildiği görülmektedir (Gupta, 1988; Majumdar, 1975; Misra, 1986; Shukla ve Sharma, 1987; Sinha ve Singh, 1987). Testler her ne kadar ikinci bir dil olarak İngilizce geliştirilmiş olsa da bilimsel yaratıcılık literatüründe yaygın bir etki sağlayamamıştır. İlk dönem bilimsel yaratıcılık testleri kuramsal yapıların yetersizliği, kültüre özgü değerler barındırmaları, testlerde bilimsel süreç beceri, bilgi ve deneyime yer verilmemesi gibi nedenlerden ötürü eleştiriler almıştır (Hu ve Adey, 2002; Mohamed, 2006). Hu ve Adey’in (2002) geliştirdikleri bilimsel yaratıcılık testi söz konusu sınırlılıkların bir kısmını gidererek bilimsel yaratıcılık literatüründe görece yaygın bir etki bıraktığı söylenebilir. SSCM modeli temel alınarak lise, ortaokul, okul öncesi öğrencilerine yönelik testler geliştirilmiştir. Ayrıca bazı araştırmacıların tekil ve çoğul düşünme becerilerini birlikte kullanarak bilimsel problem çözmeye dayalı testler geliştirdikleri ve kimisinin Mednick’in Çağrışımsal Düşünme Teorisi’ne dayandırılarak bilimsel yaratıcılık testini ortaya koydukları görülmektedir. Klahr ve Dunbar’ın geliştirdikleri SDDS modeli bilimsel süreç becerilerinin işe koşulması, bilimsel yaratıcılığın kuramsal temelli ve ampirik olarak ele alınması bakımından ayrı bir öneme sahiptir.

SDDS modeli ve yaratıcılıkta çoğul düşünme çalışmaları temel alınarak ana sınıfı 1 ve 2.sınıflara, 3-5.sınıflara ve 6-8.sınıflara özgü geçerlik, güvenirlik çalışmaları yapılmış, deneysel olarak test edilmiş bilimsel yaratıcılık testleri geliştirilmiştir. Ortaya koyulan bilimsel yaratıcılık testlerinde kendinden öncekilerinin sınırlılıklarını ortadan kaldırmayı hedefleyerek testlerin hedef grup yaş düzeyi azaltılmış, çocukların hazırbulunuşluk düzeyine uygun şekilde animasyonlar oluşturulmuş, metodolojik ve istatistiksel olarak güncel analiz yöntemleri benimsenmiştir.



3 YÖNTEM

3.1 Araştırma Modeli

Mevcut araştırmada ÇABİYAT'ın kanıt değerlendirme bileşeni geliştirilmiştir. ÇABİYAT, Klahr ve Dunbar'ın (1988), Bilimsel Keşfin İkili Arama Modeli temel alınarak anasınıfı, birinci ve ikinci sınıf öğrencilerinin bilimsel yaratıcılıklarını ölçmek amacıyla geliştirilmiş bir testtir (Ateşgöz ve Sak, 2021). Çalışma kapsamında yenilenen ÇABİYAT'ın (ÇABİYAT-R) psikometrik özelliklerinin incelenmesinde nicel araştırma yöntemlerinden tarama modeli kullanılmıştır. Tarama modeli bir örneklemin genel eğilimlerini, tutumlarını ve davranışlarını tanımlamaya çalışan (Fraenkel vd., 2012), elde edilen veriler ile genellemelere ulaşmayı hedefleyen bir araştırma modelidir (Creswell, 2009). Araştırmada tarama modellerinden kesitsel tarama modeli kullanılmıştır. Kesitsel tarama modeli farklı gruplardan belirli zaman diliminde veri toplanmasına dayanmaktadır (Karasar, 2005). Diğer bir ifadeyle kesitsel tarama modeli, bir grup bireyin “anlık görüntüsü” olarak tanımlanmaktadır (Carlson ve Morrison, 2009).

3.2 Çalışma Grubu

Çalışmada katılımcıların belirlenmesinde seçkisiz olmayan örnekleme yöntemlerinden uygun örnekleme tekniği benimsenmiştir. Uygun örnekleme tekniği zaman, para ve işgücü gibi sınırlılıkların önüne geçerek (Büyüköztürk vd., 2017), katılımcıların belirlenmesinde erişim kolaylığı sağlayabilmesi adına tercih edilmektedir (Creswell, 2013). Çalışmanın uygulamaları Eskişehir Odunpazarı ve Tepebaşı merkez ilçelerinde bulunan birer ilkokul ve bunların anasınıflarında gerçekleştirilmiştir. Çalışmaya katılan öğrencilere ilişkin bilgilere Tablo 3.1’ de yer verilmiştir.

Tablo 3.1’de görüldüğü üzere anasınıfı öğrencilerinin 43’ü kız 41’i erkek; birinci sınıf öğrencilerinin 30’u kız, 57’si erkek ve ikinci sınıf öğrencilerinin 33’ü kız, 48’si erkek öğrenciden oluşmaktadır. Çalışmaya 106 kız ve 146 erkek öğrenci olmak üzere toplamda 252 öğrenci katılmıştır.

Tablo 3.1*Katılımcıların cinsiyet ve sınıf düzeylerine göre dağılımları*

Sınıf Düzeyi	Cinsiyet	Sıklık (f)	Yüzde %
Anasınıfı	K	43	51,2
	E	41	48,8
	Toplam	84	100
Birinci Sınıf	K	30	34,5
	E	57	66,5
	Toplam	87	100
İkinci Sınıf	K	33	40,7
	E	48	59,3
	Toplam	81	100
Toplam	K	106	42
	E	146	58
	Toplam	252	100

3.3 Veri Toplama Araçları

Araştırmada veri toplama araçları olarak yenilenen versiyonu ile Çocuklar İçin Animasyonlu Bilimsel Yaratıcılık Testi (ÇABİYAT-R) ve ölçüt geçerliği kıyaslaması için SAGES 2 (Screening Assessment for Gifted Elementary and Middle School Students Second Edition-Üstün Yetenekli İlkokul ve Ortaokul Öğrencileri için Tarama Değerlendirme Bataryasının İkinci Baskısı) testi (Johnsen ve Corn, 2001) kullanılmıştır.

3.3.1 ÇABİYAT-R

Çalışma kapsamında yenilenen ÇABİYAT-R; anasınıfı, birinci ve ikinci sınıfların bilimsel yaratıcılıklarını ölçmek için kullanılan bir testtir. Bu bölümde ÇABİYAT-R testinin içeriği, uygulama süreci, puanlama yöntemi ve akabinde testin revize edilme sürecine yer verilmiştir.

3.3.1.1 ÇABİYAT-R'nin içeriği

ÇABİYAT-R kanıt değerlendirme, deney tasarlama ve hipotez geliştirme alt ölçeklerinden oluşmaktadır. Çalışma kapsamında kanıt değerlendirme alt ölçeği teste sonradan eklenmiştir. Bu alt ölçeğin geliştirilmesi süreci bu kısmın sonlarında ele alınmıştır. Bilimsel süreç becerileri kapsamında kanıt değerlendirme görevinde, deney sonucunda öğrencilere sunulan kanıt ya da kanıt kümelerinin hipotezle karşılaştırılarak bir karara varılması sürecini ifade etmektedir (Klahr ve Dunbar,

1988; Klahr, 2000). Deney tasarlama görevinde, öğrencilerin kendilerine sunulan hipotezi test ederek deneyde değişiklikler yapabilmeyi ya da bir deney düzeni tasarlayabilmeyi hedeflemektedir. Hipotez geliştirme görevinde, öğrencilere sunulan deneyin sonuçlarına ve deneyin yapılma nedenlerine ilişkin hipotez üretmeleri beklenmektedir (Klahr, 2000).

Test; kanıt değerlendirme, deney tasarlama ve hipotez geliştirme alt ölçeklerine ait 3'er madde olmak üzere toplam 9 maddeden oluşmaktadır. Maddelerin tamamı hedef grubun özellikleri dikkate alınarak kısa animasyonlar ile oluşturulmuştur. Animasyonlarda öğrencilere sunulan deneyler, gündelik yaşamdan kesitler barındırmaktadır. Tüm maddelerde öğrencilerden çok sayıda fikir üretmeleri beklenmektedir. Öğrencilerin verdikleri yanıtlar bağlamında her bir maddeye ilişkin akıcılık puanları hesaplanmıştır. Maddelerin hizmet ettikleri bilimsel süreç becerilerine ve dahil oldukları disiplinlere ilişkin bilgilere Tablo 3.2'de yer verilmiştir.

Tablo 3.2

ÇABIYAT-R maddelerine ilişkin bilgiler

Madde	Bilimsel Süreç Becerisi	Disiplin
Şeker	Kanıt Değerlendirme	Kimya
Terleme	Kanıt Değerlendirme	Biyoloji
Balon	Kanıt Değerlendirme	Fizik
Hamster	Deney Tasarlama	Biyoloji
Kum havuzu	Deney Tasarlama	Fizik
Tünel	Deney Tasarlama	Fizik
Sinek	Hipotez Geliştirme	Biyoloji
Suluk	Hipotez Geliştirme	Kimya
Gemi	Hipotez Geliştirme	Fizik/Kimya

Şeker maddesi kimya alanına ait kanıt değerlendirme becerisini ölçmektedir. Senaryoda çocuk kahvaltı sırasında şekerin çay içindeki değişimini fark etmektedir. Anne çocuğa bir hipotez vererek farklı deney düzenekleri hazırlar. Senaryoda kısa bir süre sonra deney düzeneklerinde değişimler meydana gelmektedir. Çocuktan, kendisine sunulan deney sonuçlarını değerlendirerek hipotezle uyumlu çok sayıda kanıt üretmesi beklenmektedir. Madde için yalnızca akıcılık puanı hesaplanmaktadır.

Terleme maddesi biyoloji alanında kanıt değerlendirme becerisini ölçmektedir. Senaryoda bir dede ve çocuk ormanda gezmektedirler. Dede çocuğa bitkilerin terlemesi ile ilgili bir hipotez vermektedir. Bitkilerde meydana gelen değişimlerin

fotoğrafını çeken dede, çektiği tüm fotoğrafları çocuğa göstermektedir. Çocuktan, kendisine sunulan fotoğrafların sonuçlarını değerlendirerek hipotezle uyumlu çok sayıda kanıt üretmesi beklenmektedir. Madde için yalnızca akıcılık puanı hesaplanmaktadır.

Balon maddesi fizik alanında kanıt değerlendirme becerisini ölçmektedir. Senaryoda balon oynayan bir çocuğun balonu sönmektedir. Annesi çocuğa, balonu üflemeden farklı şekillerde de şişirilebileceğine yönelik bir hipotez sunmaktadır. Sonrasında anne, balonla ilgili bazı deney düzenekleri hazırlar. Senaryoda kısa bir süre sonra deney düzeneklerinde değişimler meydana gelmektedir. Çocuktan, kendisine sunulan deneylerin sonuçlarını değerlendirerek hipotezle uyumlu çok sayıda kanıt üretmesi beklenmektedir. Madde için yalnızca akıcılık puanı hesaplanmaktadır.

Hamster maddesi biyoloji alanında deney tasarlama becerisini ölçmektedir. Senaryoda bir çocuk ve babası hamsterlar için en uygun yaşam alanını hazırlamaktadırlar. Baba çocuğa bir hipotez sunmaktadır. Çocuktan, hipotezin gerçekleşebilmesi için hazırlanan yaşam alanında çok sayıda değişiklik yapması beklenmektedir. Madde için yalnızca akıcılık puanı hesaplanmaktadır.

Kum Havuzu maddesi fizik alanında deney tasarlama becerisini ölçmektedir. Senaryoda çocuk parkta oynarken topunu bir kum havuzuna düşürmektedir. Sonrasında çocuk, kum havuzunda meydana gelen çukuru fark eder. Senaryoda anlatıcı bir hipotez sunmaktadır. Çocuktan, hipotezin gerçekleşebilmesi için çok sayıda değişiklik yapması beklenmektedir. Madde için yalnızca akıcılık puanı hesaplanmaktadır.

Oyuncak Tünel maddesi fizik alanında deney tasarlama becerisini ölçmektedir. Senaryoda bir çocuk ve halası oyuncak tünel ve oyuncak arabadan oluşan bir oyun düzeneği hazırlamışlardır. Senaryoda halası çocuğa bir hipotez sunmaktadır. Çocuktan, hipotezin gerçekleşebilmesi için çok sayıda değişiklik yapması beklenmektedir. Madde için yalnızca akıcılık puanı hesaplanmaktadır.

Sinek maddesi biyoloji alanında hipotez geliştirme becerisini ölçmektedir. Senaryoda bir çocuk bataklığa gitmekte ve bataklıktaki sinekler ile ilgili aklına bir soru gelmektedir. Öğrencilerden sorunun cevabına yönelik çok sayıda hipotez üretmeleri beklenmektedir. Madde için yalnızca akıcılık puanı hesaplanmaktadır.

Suluk maddesi kimya alanında hipotez geliştirme becerisini ölçmektedir. Senaryoda iki çocuk birlikte oyun oynarlarken susadıklarını fark etmektedirler. Fakat sularını içerlerken bir durum ile karşılaşmaktadırlar. Öğrencilerden söz konusu durumun nedenlerine yönelik çok sayıda fikir üretmeleri beklenmektedir. Madde için yalnızca akıcılık puanı hesaplanmaktadır.

Gemi maddesi fizik ve kimya alanında hipotez geliştirme becerisini ölçmektedir. Senaryoda bir anne, kızı için bir oyun hazırlamaktadır. Anne, kızından oyunun amacı doğrultusunda değişiklik yapmasını ister. Öğrencilerden söz konusu değişikliğe yönelik çok sayıda fikir üretmeleri beklenmektedir. Madde için yalnızca akıcılık puanı hesaplanmaktadır.

3.3.1.2 ÇABİYAT-R'nin uygulama süreci

ÇABİYAT-R; anasınıfı, birinci ve ikinci sınıf öğrencilerine bireysel olarak uygulanabilen bir testtir. Testin her maddesinin bir animasyonu vardır (Şekil 3.1) ve animasyon sonunda bir soru sorulmaktadır. Animasyon tablet bilgisayarda çocuğa izlettirildikten sonra sorulan soruya ilişkin çocuğun verdiği yanıtlar uygulayıcı tarafından kayıt formuna kaydedilir.

Şekil 3.1

Terleme maddesi animasyonu



Kayıt formunda her bir madde için oluşturulmuş doğru yanıt havuzu yer almaktadır. Öğrenciden gelen yanıtlar uygulayıcı tarafından anlık olarak değerlendirilerek kayıt altına alınmıştır. Öğrencinin doğru yanıtları görme olasılığını engellemek için uygulayıcı ve öğrenci arasında karton bir paravan kullanılmıştır.

Testin uygulanması sırasında oturma düzeni Şekil 3.2’de görselleştirilmiştir.

Şekil 3.2

ÇABİYAT-R'nin uygulama düzeni



Öğrencinin önünde animasyonları izleyeceği bir tablet, öğrenci ile uygulayıcı arasında bir paravan ve uygulayıcı önünde doğru yanıt havuzunun yer aldığı kayıt formu bulunmaktadır. Testin uygulama süresi öğrencilerden gelen yanıtların sayısına ve öğrenci sınıf düzeyine bağlı olarak değişkenlik gösterebilmektedir. Test, ortalama 20-25 dakika sürmektedir. Maddelerin tamamının toplam animasyon süresi 11 dakikadır. Test süresince öğrenciler üzerinde bir baskı yaratmaması amacıyla yanıt üretme sürecinde süre kısıtlaması uygulanmamıştır. Tablo 3.3'te her bir maddenin animasyon gösterim süresine yer verilmiştir.

Tablo 3.3

Maddelerin animasyon süreleri

Madde	Animasyon Süresi (sn)
Şeker	KD1 84
Terleme	KD2 97
Balon	KD3 92
Hamster	DT1 92
Kum havuzu	DT2 50
Tünel	DT3 56
Sinek	HG1 55
Suluklar	HG2 53
Gemi	HG3 81
Toplam	660 saniye

KD: Kanıt Değerlendirme. DT: Deney tasarlama. HG: Hipotez geliştirme

ÇABİYAT-R uygulaması öğrenci ile kısa bir tanışma ile başlamaktadır. Hedef grubun yaş düzeyi küçük olması sebebiyle tanışma süresi gerekirse uzatılabilir. Tanışmanın ardından test hakkında genel bir bilgilendirme yapılmaktadır: “*Bugün seninle çizgi filmler izleyeceğiz. İzlediğin çizgi filmlerin sonunda bazı sorular olacak. Bu soruların çok sayıda doğru cevabı bulunmaktadır. Bu yüzden senden*

olabildiğince çok sayıda doğru cevaplar bulmanı istiyorum. Hazır mısın?” Öğrenci hazır olduğunu belirttikten sonra ilk animasyon öğrenciye izlettirilir. Öğrenci animasyonu izledikten sonra uygulayıcı: “Çizgi filmin sonundaki soruyu anladın mı? şeklinde bir soru yönlendirmektedir. Öğrenci soruyu anladığını ifade ederse: “Cevabın nedir? şeklinde ikinci soru yöneltilir. Eğer öğrenci soruyu anlamadıysa animasyon tekrar izlettirilir ve madde ile ilgili kısa bir açıklama yapılır. Öğrenci maddeye yanıt/yanıtlar ürettikten sonra cevabın doğru ya da yanlışlığı ile ilgili bir açıklama yapılmaz. Öğrencinin daha fazla yanıt üretmesini sağlamak için “Çok güzel cevap verdin, peki başka ne olabilir? şeklinde yönlendirmeler yapılmaktadır. Öğrenci eğer maddeye başka bir yanıt üretemiyor ya da üretemeyeceğini ifade ediyorsa o madde sonlandırılır. Sonra “Şimdi diğer çizgi filmimize geçiyoruz” ifadesi ile yeni bir animasyona geçileceği belirtilmektedir. Testin uygulanması tüm maddeler için aynı süreci kapsamaktadır.

3.3.1.3 ÇABİYAT-R puanlama yöntemi

ÇABİYAT (Ateşgöz ve Sak, 2021), çoğul düşünmeye dayalı açık uçlu maddelerden oluşan bir yaratıcılık testidir. Çoğul düşünmeye dayalı yaratıcılık testlerinde en sık olarak akıcılık, esneklik ve orijinallik puanları kullanılır. Akıcılık, madde için üretilen toplam yanıt sayısını; esneklik, üretilen yanıtların kategori sayısını (Guilford, 1968); orijinallik puanı ise üretilen yanıtların özgünlük veya sıradışılık derecesini gösterir (Runco, 1999). ÇABİYAT testi için söz konusu puanlara ek olarak yaratıcılık bölümü olarak adlandırılan (Creativity Quationt-CQ) bir başka puanlama daha yapılmaktadır. CQ, akıcılık ve esneklik puanlarının logaritmik dönüşümü ile elde edilmektedir (Snyder vd., 2004): $CQ = \log_2 (1 + u_1) + \log_2 (1 + u_2) + \dots \log_2 (1 + u_c)$. Formülde “u” üretilen fikir sayısını, “1, 2.. c” sayıları ise kategoriyi temsil etmektedir.

Çalışma kapsamında yenilenen ÇABİYAT-R için öğrencilerin madde bazında ve toplam akıcılık puanları hesaplanmış olup diğer puanlama yöntemleri araştırmaya dahil edilmemiştir. Çalışmada puanlama yöntemi olarak yalnızca akıcılığın benimsenmesi ve diğer puanların dışlanması ilişkin gerekçelere aşağıda yer verilmiştir.

- *Akıcılığın diğer puanlar üzerindeki etkisi*

Yaratıcılığın farklı kuram ve modeller ile açıklanması (Csikszentmihalyi, 2014; Guilford, 1950; Kozbelt vd., 2010; Sternberg ve Lubart, 1993), farklı ölçüm yöntemlerini de beraberinde getirmiştir (Kaufman vd., 2008; Park vd., 2016). Kimi araştırmacılar ölçümlerinde akıcılığı kullanırken (Storme vd., 2017; Zenasni ve Lubart, 2011), kimileri akıcılık, esneklik, orijinallik ve detaylandırma puanlarının bir bileşimini (Albert ve Kormos, 2004; Goff ve Torrance, 2002; Sak ve Maker, 2006), bazıları uygunluk ve uygunluktan türetilen puanları (Atakaya vd., 2024; Diedrich, vd., 2015; Runco ve Charles, 1993), kimileri de sürpriz puanlarını kullanmıştır (Acar vd., 2017). Fakat araştırmalar incelendiğinde genel olarak akıcılık, esneklik ve orijinallik puanlarının daha sık kullanıldıkları söylenebilir. Yaratıcılık ölçümlerinde benimsenen puanlama kriterinin dışında, araştırmacıların dikkat etmesi gereken asıl önemli unsurun ise akıcılık puanlarının diğer puanlar ile yüksek korelasyonlar göstermesidir. Yapılan araştırmalarda akıcılığın diğer puan türleri üzerindeki kirletici etkisine rastlanmaktadır (Hocevar, 1979; Runco vd., 2008; Silvia, 2008). Elde edilen diğer puan türlerinin, akıcılıktan üretilmesi söz konusu kirletici etkinin temel nedeni olarak gösterilebilir. Araştırmalarda akıcılığın kirletici etkisinin önüne geçebilmek amacıyla farklı puanlama yöntemleri kullanılmıştır. Kimi araştırmacılar orijinalligi puanlamada farklı teknikler kullanırlarken (Beketayevab ve Runco, 2016; Plucker vd., 2014; Silvia vd., 2008); kimileri akıcılık puanının kirletici etkisini azaltabilmek adına logaritmik dönüşümlerden yararlanmışlardır (Ateşgöz ve Sak, 2021; Bermejo vd., 2016; Sak ve Ayas, 2013; Snyder vd., 2004). Fakat akıcılık ve esnekliğin logaritmik dönüşümü alınarak gerçekleştirilen çalışmalarda da akıcılık ile diğer puan türleri arasında $r=.90$ 'nın üzerinde yüksek korelasyon değerlerine rastlanmıştır (Bermejo vd., 2016; Sak ve Ayas, 2013). Bu nedenle bazı araştırmacıların açımlayıcı ve doğrulayıcı faktör analizlerini yalnızca akıcılık puanları üzerinden gerçekleştirdikleri görülmektedir (Ateşgöz ve Sak, 2021; Ayas, 2017). Bu çalışmada da benzer gerekçelerle, akıcılık puanı dışındaki puan türlerinin (örneğin esneklik, orijinallik) ölçüme dahil edilmemesi tercih edilmiştir. Özellikle orijinallik ve esnekliğin akıcılıkla yüksek korelasyon değerlerine sahip olması bu puan türlerinin bağımsız birer değişken olarak yorumlanmasını zorlaştırmakta ve geçerliklerini sınırlamaktadır.

- *İstatistiksel sınırlılıklar*

Çalışma kapsamında yalnızca akıcılık puanının kullanılması yerine farklı beceriler için farklı puanlama yöntemlerinin neden kullanılmadığı sorusu akıllara gelebilir. Örneğin kanıt değerlendirme becerisi için yalnızca akıcılığın; hipotez geliştirme ve deneye tasarlama becerileri için akıcılığa ek olarak diğer puanlama türlerinin kullanılması önerilebilir. Fakat bu tarz bir yaklaşımın benimsenmesinin istatistiksel olarak bazı sınırlılıklara yol açabileceği düşünülmüştür. Çalışmada testin psikometrik özelliklerini incelemek amacıyla Açıklayıcı Yapısal Eşitlik Modellemesi'nden (AYEM) yararlanılmıştır. Kanıt değerlendirme, deney tasarlama ve hipotez geliştirme görevlerine ilişkin farklı puanlama yöntemlerinin benimsenmesi ölçüm hatalarına (Nunnally ve Bernstein, 1994), ölçüm heterojenliğine (DeVellis ve Thorpe, 2022) ve ölçüm eşdeğerliği sorunlarına (Meredith, 1993) yol açabilir. Söz konusu durumun, faktörler arasında sistematik bir hata ile model uyum indekslerinin bozulmasına, yanlış sonuçlar elde edilmesine, yapı geçerliği ve güvenirlik analizlerinin yanlış yorumlanmasına neden olacağı düşünülmüştür. Faktörler arasında eşdeğer bir puanlama kriterinin olmaması, faktörlerin birbirleriyle kıyaslanmalarını zorlaştırabileceği gibi, testin bir bütün olarak değerlendirilmesinde de sınırlılık yaratabilir. Belirtilen gerekçeler kapsamında çalışmada ÇABİYAT-R için yalnızca akıcılık puanının kullanılmasına karar verilmiştir. Analizler, öğrencilerin her bir madde için ürettikleri akıcılık puanı ile tüm testten elde edilen toplam akıcılık puanları üzerinden yürütülmüştür.

3.3.1.4 ÇABİYAT-R'nin revize edilme süreci

Revize sürecinde DeVellis ve Thorpe (2022) tarafından önerilen ölçek geliştirme adımları takip edilmiştir:

- Teorik altyapının şekillendirilmesi
- Test madde havuzunun oluşturulması
- Oluşturulan madde havuzundan maddelerin seçilmesi için alan uzman görüşünün alınması
- Animasyonların oluşturulması
- Ön uygulama: Maddelere ilişkin nitel verinin toplanması
- Ana uygulama
- Testin psikometrik özelliklerinin incelenmesi

Yukarıda belirtilen basamaklardan ana uygulamanın bulgularına ve testin psikometrik özelliklerinin incelenmesi adımlarına bulgular kısmında yer verilmiştir. Basamakların diğer adımları bu bölümde detaylı bir şekilde açıklanmıştır.

3.3.1.4.1 ÇABİYAT-R'nin teorik altyapısı

ÇABİYAT-R testinin kuramsal alt yapısı Bilimsel Keşfin İkili Arama Modeline (Klahr ve Dunbar, 1988; Klahr, 2000) ve yaratıcılık ölçümlerinde sıklıkla kullanılan çoğul düşünmenin temel alındığı araştırmalara dayanmaktadır (Guilford, 1950, 1967; Torrance, 1988). Benimsenen kuramsal çerçeve BÜT 6-8 (Bilimsel Üretkenlik Testi- Ayas ve Sak, 2014; Sak ve Ayas, 2013), BÜT 3-5 (3-5. sınıflar için Bilimsel Üretkenlik Testi- Ayas, 2017) ve ÇABİYAT (Çocuklar İçin Animasyonlu Bilimsel Yaratıcılık Testi- Ateşgöz ve Sak, 2021) testlerinde de kullanılmıştır.

Bilimsel Keşfin İkili Arama modeline göre bilimsel bir keşif genel ve alana özgü bilgi ve bilimsel süreç becerileri olmak üzere iki boyut altında toplanmaktadır (Klahr, 2000). Bilimsel süreç becerileri ise hipotez ve deney olmak üzere iki temel alandan oluşmaktadır (Klahr ve Dunbar, 1988). Bu alanlar hipotez geliştirme, deney tasarlama (hipotezi test etme) ve kanıt değerlendirme şeklinde birbirlerinden bağımsız 3 temel bileşene sahiptir (Klahr vd., 1993). Hipotez geliştirme, bireylerin sahip oldukları bilgilerinin geri çağrıldığı hipotez üretme safhasıdır. Deney tasarlamada birey kendisine verilen hipoteze uygun deney tasarlar, deneye ilişkin tahminlerde bulunur, deneyi yapar ve deneye ilişkin yaptığı tahminle deney sonucunu eşleştirir. Kanıt değerlendirmede ise birey, kendisine sunulan hipotez ve deney sonuçlarına ilişkin kanıt kümelerini değerlendirerek hipotezin reddine ya da kabulüne karar verir. Diğer bir ifadeyle birey kanıt değerlendirmede deney sonuçlarının hipotezle uygunluğunu kontrol eder. Modele göre bilimsel bir keşif, üç temel bileşenin etkileşimi ile ortaya çıkmaktadır (Klahr, 2000).

ÇABİYAT-R'nin teorik alt yapısı bilimsel süreç becerilerini ve genel yaratıcılık becerileri ile ilişkilendirilen akıcı fikir üretme becerisini kapsamaktadır. ÇABİYAT testinden farklı olarak, ÇABİYAT-R'ye bilimsel süreç becerileri kapsamında kanıt değerlendirme becerisi eklenmiştir. Ayrıca akıcılığın dışındaki genel yaratıcılıkla ilişkilendirilen esneklik, orijinallik ve akıcılık ile esneklikten türetilen yaratıcılık bölümü (CQ) becerileri testin dışında bırakılmıştır. Genel yaratıcılık becerilerinin dışarıda bırakılma gerekçelerine ÇABİYAT-R'nin puanlama yöntemi kısmında

detaylı bir şekilde yer verilmiştir. Revize edilen ÇABİYAT-R ve ÇABİYAT' ın teorik alt yapılarına Şekil 3.3 ve Şekil 3.4'te yer verilmiştir.

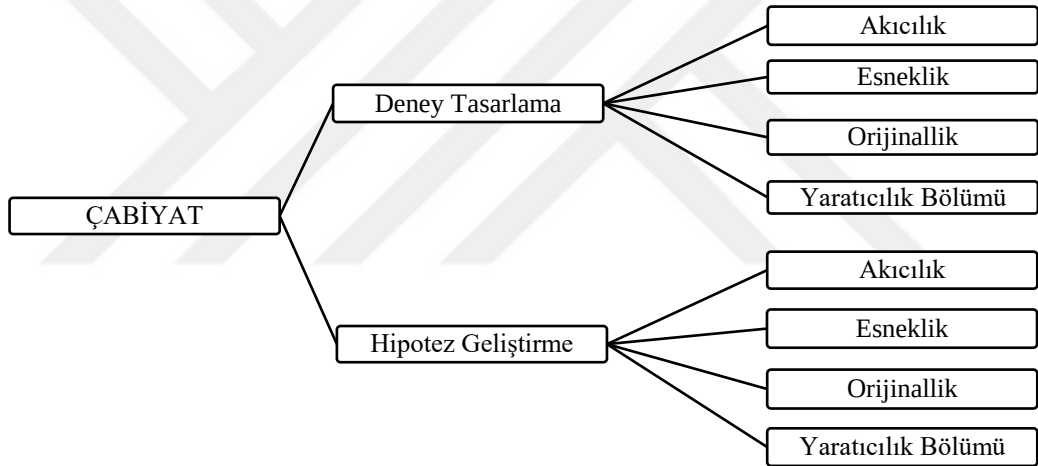
Şekil 3.3

ÇABİYAT-R'nin teorik alt yapısı



Şekil 3.4

ÇABİYAT'ın teorik alt yapısı



Kanıt değerlendirme becerisine ilişkin görevlerin; kanıt örüntülerinin değerlendirilmesi (Amsel ve Brock, 1996; Kuhn vd., 1988), verilerin analiz edilmesi ve yorumlanması (Masnick vd., 2016; Masnick & Morris, 2008) ve kanıta ilişkin açıklamalar oluşturma (Gelmen ve Klemmer, 1991; Kelemen vd., 2011) olarak üç başlık altında toplandığı görülmektedir (Zimmerman ve Klahr, 2018). Kuhn ve arkadaşları (1988), kanıt değerlendirme üzerine yaptıkları bir dizi çalışmada kanıt üretmenin, kanıt değerlendirme beceresine hizmet eden bir başka görev tipi olduğunu savunmuşlardır. Bu görevde öğrencilerden, kendilerine verilen hipotezin kabul ya da reddine ilişkin deney sonuçları ile uyumlu kanıtlar üretmeleri beklenmektedir. Yapılan okumalarda kanıt üretmenin çoğul düşünme çalışmaları ile ilişkilendirilebileceği görülmüştür. ÇABİYAT-R'nin teorik çerçevesinin bir diğer

ayağının çoğul düşünme çalışmalarına dayandırılması sebebiyle kanıt değerlendirme becerisinin, kanıt üretme görev tipi ile yürütülmesine karar verilmiştir. Bu görev tipinin benimsenmesindeki bir diğer neden, hedef kitlenin bilişsel özelliklerinin gözetilmesidir. Kuhn ve arkadaşları (1988), verilerin yorumlanması ve açıklanmasına dayalı kanıt değerlendirme görev tiplerinin küçük yaş düzeyi çocuklar için uygun olmadığını belirtmişlerdir. ÇABİYAT-R testi için kanıt üretme görev tipi, çoğul düşünmeye uygun olarak yeniden düzenlenmiştir. Çalışmada kanıt değerlendirme görevi için öğrenciye öncelikle hipotez verilmiştir. Akabinde hipotezle uyumlu her bir madde için ortalama 8 deney düzeneği bir bütün halinde öğrencilere sunulmuştur. Her düzenek gizil olarak yerleştirilmiş bir değişkene sahiptir. Deney düzeneklerinde yalnızca bulunması istenen değişkene yönelik farklılık yer almakta ve diğer değişkenler sabit tutulmaktadır. Öğrenci kendisine sunulan deney düzeneklerini birbirleriyle kıyaslayarak değişkenleri yani kanıtları bulmaya çalışmaktadır. Kanıt değerlendirme beceresini ölçmeye yönelik geliştirilen maddelerde, özellikle deney düzeneklerinin doğru bir şekilde sunulması oldukça önemlidir.

3.3.1.4.2 Test madde havuzunun oluşturulması

Ölçek geliştirme çalışmalarında kuramsal yapının oluşturulmasından sonraki önemli adımlardan biri de madde havuzunun oluşturulmasıdır. Bu süreçte öncelikli olarak madde havuzunda yer alacak madde sayısının belirlenmesi gerekmektedir. Madde havuzunda kaç maddenin yer alacağı konusunda literatürde net bir görüş birliği sağlanamasa da ilk havuzun kapsamlı olması gerektiği savunulmaktadır (Clark ve Watson, 2019; Loevinger, 1957). Kimi araştırmacılar testin planlanan son halinde yer alacak madde sayısının 2 katı (Kline 1993; Boteng vd., 2018, Hinkin, 1998) kimileri 3-4 katı (DeVellis ve Thorpe, 2022; Kaplan and Saccuzo, 2009), kimi araştırmacılar ise 5 katı (Weiner vd., 2012) kadar maddenin geliştirilmesi gerektiğini belirtmektedirler. ÇABİYAT testinin hipotez geliştirme ve deney tasarlama alt ölçeklerinde 3'er madde bulunmaktadır. ÇABİYAT-R'ye eklenecek olan kanıt değerlendirme alt ölçeği için de 3 maddenin teste dahil edilmesi planlanmıştır. Bu kapsamda kanıt değerlendirme alt ölçeği için 9-12 arası maddenin ana çalışmada test edilmesi hedeflenmiştir. Söz konusu hedeflenen maddelerin seçilebilmesi için ise madde havuzunun daha geniş tutulması planlanmıştır.

Madde havuzunun başarılı bir şekilde oluşturulması, doğrudan kapsam geçerliği ile ilgilidir (Clark ve Watson, 2019; Hinkin, 1998). Maddelerin ölçtüğü varsayılan yapıyı ölçebilmesi için kapsam geçerliği hayati önem taşımaktadır (DeVellis ve Thorpe, 2022). Kapsam geçerliği, değerlendirme aracının tüm unsurlarının geliştirilmesi hedeflenen yapıya uygunluğunu ve temsil edebilme derecesini ifade etmektedir (Haynes vd., 1995). Ölçek geliştirme sürecinde madde havuzu oluştururken tündengelim (mantıksal bölümlleme) ve tümevarım (gruplama) olmak üzere iki yaklaşım benimsenmektedir (Boeteng vd., 2018; Hinkin, 1998). Tündengelim yaklaşımında araştırmacı literatürdeki mevcut kuram ve önceki araştırmalara dayanarak madde geliştirmektedir. Tümevarım yaklaşımında ise belirli bir gruptan çoğunlukla nitel veriler toplanarak madde geliştirilmektedir. Çalışma kapsamında geliştirilen testin net bir teorik çerçeveye sahip olması sebebiyle madde geliştirmede tündengelim yaklaşımı benimsenmiştir. Madde yazım sürecinde kapsam geçerliğinin sağlanması adına alan uzmanlarından destek alınmıştır (Haynes vd., 1995).

Madde geliştirme sürecinde yer alan ekip üyeleri fen bilgisi, sınıf öğretmeni ve anaokulu öğretmenlerinden en az 10 yıllık deneyim sahibi olmaları gözetilerek Eskişehir ve Ankara sınırları içinde uygun örneklem tekniği ile belirlenmiştir. Ekip üyeleri 5 fen bilgisi, 1 sınıf ve 1 anaokulu öğretmeni olmak üzere 3'ü erkek, 4'ü kadın toplam 7 kişiden oluşmaktadır. Öğretmenlerin mesleki deneyim yıl ortalamaları $\bar{x}=14,1$ yıldır. Madde geliştirme süreci, haftada bir olmak üzere toplam 6 haftalık periyotta, 2023-2024 eğitim öğretim yılının Mayıs-Haziran ayları arasında tamamlanmıştır. Haftalık toplantılar araştırmacı moderatörlüğünde ortalama 60 dakikalık zaman dilimlerinde online olarak gerçekleştirilmiştir. Madde geliştirme süreci sonunda MEB'in belirlemiş olduğu ek ders ücreti haftalık toplantı bazında ele alınarak beş saat üzerinden hesaplanmış ve öğretmenlere ücret ödemeleri gerçekleştirilmiştir. Toplantılar ekip üyelerinin izinleri dahilinde online olarak kayıt altına alınmıştır. Toplantı tarihi, süresi ve toplantıların içeriklerine ilişkin bilgilere Tablo 3.4' te yer verilmiştir.

Tablo 3.4' te görüldüğü üzere öğretmenlere ilk toplantıda bilimsel süreç becerileri ve becerilerin ölçülmesine yönelik teorik eğitim verilmiştir. Eğitimde hipotez geliştirme

ve deney tasarlama becerilerine kısaca değinilerek kanıt değerlendirme becerisinin üzerinde durulmuştur.

Tablo 3.4

ÇABİYAT-R madde havuzu oluşturulma süreci

	Toplanma Nu./Tarih	Süre	İçerik
1.	04.05.2023	58 dk.	Bilimsel süreç becerileri ve kanıt değerlendirme becerisinin ölçülmesine yönelik eğitim
2.	11.05.2023	59 dk.	Kanıt değerlendirme için geliştirilen maddelerin değerlendirilmesi
3.	18.05.2023	62 dk.	Kanıt değerlendirme için geliştirilen maddelerin değerlendirilmesi
4.	25.05.2023	65 dk.	Kanıt değerlendirme için geliştirilen maddelerin değerlendirilmesi
5.	01.06.2023	57 dk.	Kanıt değerlendirme için geliştirilen maddelerin değerlendirilmesi
6.	08.06.2023	62 dk.	Kanıt değerlendirme için geliştirilen maddelerin değerlendirilmesi

Eğitimde hipotez geliştirme ve deney tasarlama becerilerine kısaca değinilerek kanıt değerlendirme becerisinin üzerinde durulmuştur. Literatürde yer alan kanıt değerlendirme görev tipleri hakkında bilgi verilerek, araştırmacı tarafından uygun görülen kanıt üretme görev tipinin içeriği ve değerlendirilmesi anlatılmıştır. Öğretmenler üzerinde kısıtlayıcı bir ön yargı oluşturmaması için örnek maddeler üzerinden bir anlatı yolu tercih edilmemiştir. Bu hali ile öğretmenlerin geliştirecekleri maddelere ilişkin zihinsel engeller oluşturulmaması hedeflenmiştir. Öğretmenlerden, ikinci haftadan itibaren kanıt değerlendirme becerisine ilişkin en az iki madde geliştirmeleri istenmiştir. Öğretmenler, toplantının yapılacağı tarihten iki gün önce araştırmacıya maddeleri elektronik ortamda iletmışlerdir. Öğretmenlerin madde geliştirirken takip edecekleri taslak forma Ek 1’ de yer verilmiştir. Öğretmenlerden gelen her bir maddeye araştırmacı tarafından ön bir değerlendirme yapılmıştır. Burada amaç maddelerin kabul ya da reddine ilişkin bir değerlendirme yapılmadan, toplantı sürecine hazırlıklarını kapsamaktadır. Maddeler, birer birer toplantı anında öğretmenler ile değerlendirilmiştir. Araştırmacının buradaki rolü öğretmenlerin geliştirdikleri maddeler ile mevcut teori arasında köprü oluşturmak ve maddelere ilişkin dönütler vermektir. Geliştirilen maddeler belirli kriterler bağlamında değerlendirilmiştir.

- *Hedef gruba uygunluk:* Maddeler anasınıfı, birinci ve ikinci sınıf öğrencilerinin hazırbulunuşluk ve bilgi düzeylerine uygun olmalıdır.
- *Açık uçluluk:* Maddeler için çok sayıda doğru fikir üretilebilmelidir.

- *Beceriye uygunluk:* Kanıt değerlendirme becerisi için öncelikli olarak bir hipotez verilmeli, hipoteze uygun deney sonuçları paylaşılmalı ve deney sonuçlarından yola çıkarak kanıtlar üretilebilmelidir.
- *Anlaşılabilirlik:* Maddeler dil bilgisi bakımından açık ve net olmalıdır. Aynı zamanda hedef grubun özellikleri dikkate alınarak karmaşık terim ve kavramlardan kaçınılmalı, dil sadeleştirilmelidir.
- *Bilimsellik:* Maddelerin içeriği bilimsel gerçekliğe dayanmalıdır.

Özellikle ilk haftalarda geliştirilen maddelerin beceriye uygunluk kriterini karşılamadığı görülmüştür. Maddelerde hipoteze yer verilmemesi, hipotezi destekleyen deney sonucunun yetersizliği ya da yanlışlığı, hipotez-deney sonuçlarının birbirleri ile uyumsuzluğu söz konusu eksikliklere örnek olarak gösterilebilir. Araştırmacı, her bir maddeye ilişkin testin teorik çerçeve bütünlüğünü sağlamak adına geri dönütler vermiştir. Üçüncü haftadan itibaren öğretmenlerin, beceriyi ölçmede daha uygun maddeler geliştirdikleri fark edilmiştir. Sürecin başında öğretmenlerden geliştirecekleri maddeleri görselleştirmeleri beklenmiştir. Fakat öğretmenler, görsel bulma ve görselleri beceriye uygun bir şekilde sunulmasının çok fazla vakit aldığını, görsel bulma ve oluşturma sürecine odaklanıldığında ise madde geliştirmenin zor olduğunu ifade etmişlerdir. Öğretmenlerin yaşadıkları probleme çözüm getirmek adına, görsel bulma ve oluşturma yerine deney sonuçlarını detaylı bir şekilde betimlemelerine karar verilmiştir. Benzer süreçler izlenerek toplantılar tamamlanmıştır. Toplantılar sonunda 16 fizik, 8 kimya ve 12 biyoloji alanında olmak üzere kanıt üretmeye yönelik toplam 36 adet madde geliştirilmiştir. Geliştirilen maddelerin birbirlerine benzerlikleri, hedef görevden uzaklaşmaları, görev bağlamında çok sayıda yanıt üretilememesi gibi nedenlerden dolayı madde sayısı araştırmacı tarafından 27'ye indirilmiştir. Maddelerden 10'u fizik, 7'si kimya ve 10 tanesi biyoloji alanından oluşmaktadır. Geliştirilen madde sayısının, literatürde madde havuzunda yer alması için önerilen sayının oldukça üstünde olduğu söylenebilir (DeVellis ve Thorpe, 2022; Kaplan and Saccuzo, 2009).

3.3.1.4.3 Havuzdan madde seçimi

Havuzdan madde seçimi üç aşamalı olarak gerçekleştirilmiştir. İlk aşama yaratıcılık alanında uzman bir akademisyen ve araştırmacı ile yürütülmüştür. Akademisyen doçentlik unvanına sahip, yaratıcılık ve bilimsel yaratıcılık alanlarında çalışmaları

olan bir alan uzmanıdır. Bu süreçte madde geliştirme kriterleri ve testin kuramsal alt yapısı göz önünde bulundurularak madde sayısı 18'e düşürülmüştür. İkinci aşamada bilimsel yaratıcılık alanında uzman farklı iki kişinin görüşlerine başvurulmuştur. Alan uzmanları yaratıcılık, bilimsel yaratıcılık alanlarında doktora ve yüksek lisans derecesine sahip akademisyenlerden oluşmaktadır. Maddenin beceriye, hedef gruba, bilimsel gerçekliğe uygunluğu, alan bilgisi yeterliliği ve maddenin anlaşılır olup olmaması kriterleri bağlamında, iki alan uzmanı birbirlerinden bağımsız olarak görüş bildirmişlerdir (Ek 2). Belirlenen kriterler kapsamında iki alan uzmanı arasında tam uyumun sağlanamadığı 5 madde çıkarılarak madde sayısı 13'e indirilmiştir.

Üçüncü ve son aşamada ise 13 maddenin temsili resimleri araştırmacı tarafından görselleştirilmiş, 1 fen bilgisi ve 3 ana okul öğretmenine görüşlerinin alınması amacıyla elektronik ortamda gönderilmiştir. Fen bilgisi öğretmeni madde havuzu oluşturulma sürecinde katkı sunan uzmanlardan biridir. Üç anaokulu öğretmeni ise madde oluşturulma sürecinde yer almamışlardır. Fen bilgisi öğretmeni maddeleri bilimsel yaratıcılık alanında uzman kişilere sunulan kriterler bağlamında değerlendirmiştir. Anaokulu öğretmenleri ise maddeleri zorluk kriteri bağlamında değerlendirmişlerdir (Tablo 3.5).

Tablo 3.5

Anaokulu öğretmenleri madde değerlendirme kriteri

Sınıfınızdaki öğrencileri düşündüğünüzde; öğrencileriniz ne düzeyde bu maddeye en az 1 yanıt üretir? Belirtiniz.				
Hiçbiri	Bir kısmı	Neredeyse yarısı	Çoğunluğu	Hepsi
☐	☐	☐	☐	☐
Önerileriniz				

Öğretmenlerden alınan görüşler kapsamında “hiçbiri” seçeneğinin 2 ana okul öğretmeni tarafından aynı maddelerde işaretlendiği görülmüştür. Fen bilgisi öğretmeni de söz konusu 3 maddeye yönelik kriterlerin yeterince kapsanmadığına ilişkin değerlendirmelerde bulunmuştur. Ayrıca anaokulu öğretmenlerinden ikisi bir maddenin hedef gruba uygun olmadığını belirtmişlerdir. Bu kapsamda madde sayısı 9'a düşürülmüştür. Madde seçimine ilişkin genel bir özete Tablo 3.6' da yer verilmiştir.

Tablo 3.6*Madde seçim süreci ve disiplinlere göre dağılımları*

Aşamalar	Madde Sayısı	Fizik	Kimya	Biyoloji
Havuzdaki toplam madde sayısı	27	10	7	10
Birinci Aşama	18	9	3	6
İkinci Aşama	13	6	3	4
Üçüncü Aşama	9	5	2	2

3.3.1.4.4 Animasyonların oluşturulması

Animasyonların oluşturulmasında Ateşgöz' ün (2020), ÇABİYAT' ı geliştirme sürecinde uyguladığı adımlar temel alınmış ve küçük değişimler yapılarak mevcut araştırmaya uygun hale getirilmiştir. Animasyonların oluşturulması beş adımda gerçekleştirilmiştir. İlk adımda maddeler hikayelere dönüştürülmüş, ikinci adımda hikayelerin kapsam geçerliğini sağlamak adına uzman görüşü alınmış, üçüncü adımda animasyon oluşturmaya yönelik hazırlıklar yapılmış, dördüncü adımda animasyonlar tasarlanmış ve son adımda oluşturulan animasyonlar uzmanlara gönderilmiştir.

Maddelerin hikayelere dönüştürülmesi

Madde havuzu sonucunda elde edilen 9 madde araştırmacı tarafından hikayelere dönüştürülmüştür. Hikayeler, günlük hayattan ve somut örneklerle dayandırılarak yazılmıştır. Hedef grubun somut işlemler döneminde olması sebebiyle öğrencilerin gündelik yaşamdan deneyimleyebileceği hikayelerin olması bu bakımdan önem arz etmektedir. Hikayelerin yazımında, ÇABİYAT' ın hipotez geliştirme ve deney tasarlama alt ölçeklerinde yer alan maddelerin hikayeleri ile tutarlılık göstermeleri hedeflenmiştir. Hikayelerin oluşturulması yaklaşık bir aylık süreci kapsamıştır. Her bir maddeye ilişkin hikâye yazımı tamamlandıktan sonra, hikâyenin dil bilgisi ve anlam bütünlüğünü değerlendirmek amacıyla bir Türkçe öğretmeninden destek alınmıştır. Uzmanın verdiği dönütler düzeltilerek bir sonraki maddenin hikayesi yazılmıştır. 10 madde için benzer süreç takip edilerek hikâye yazımı tamamlanmıştır.

Hikâyelerin kapsam geçerliği

Yazılan hikâyelerin animasyonlar oluşturulmadan önce uzmanlara gönderilmesi önemli bir adımdır. Bu adımdan sonra animasyonlar oluşturularak ön değerlendirme

ve ana uygulamaya geçilmiştir. Hikâyelerin kapsam geçerliği iki ayrı uzman grubunun yer aldığı iki ayrı aşamada tamamlanmıştır. İlk aşamada bir fen bilgisi öğretmeni ve bir akademisyenin görüşlerine başvurulmuştur. İki alan uzmanı da havuzdan madde seçimi sürecinde görev almışlardır. Bu aşamada uzmanlar hikâyenin, kanıt değerlendirme becerisine uygunluğunu ve anlaşılabilirliğini değerlendirmişlerdir. Uzmanlar, hikayelerin kanıt değerlendirme becerisine ve hedef gruba uygun olduğu konusunda görüş bildirmişlerdir. İlk aşama tamamlandıktan sonra ikinci aşamaya geçilmiştir. İkinci aşamada ise hikâyenin anasınıfına uygunluğu ve anlaşılabilirliğinin değerlendirilmesi amacıyla 3 anaokulu öğretmenin görüşleri alınmıştır. Alan uzmanlarından gelen dönütler sonrasında 9 madde ile animasyon oluşturulma süreci devam ettirilmiştir.

Animasyon oluşturma sürecine yönelik hazırlıklar

Animasyon oluşturma süreci öncesinde resimli taslakların (storyboard) çizilmesi önerilmektedir (Winder ve Dowlatabadi, 2013). Resimli taslaklar, hikayelerin bir düzen ve plan dahilinde sunulmasıdır (Hart, 2013). Araştırmacı, animasyon oluşturma adımı bir grafik tasarımcısından profesyonel destek almıştır. Araştırmacının resimli taslak oluşturmaya yönelik bir uzmanlığının olmaması sebebiyle, maddelerin animasyona dönüştürülme süreci detaylı betimlemelerin yer aldığı metinler yolu ile animatöre sunulmuştur. Animasyonun nasıl ve nerede başlayacağı, animasyonda yer alacak karakterlerin nasıl seçileceği, olay örgüsünün nasıl ilerleyeceği, karakterlerin nerede ve nasıl tepki verecekleri, deney düzeneklerinin nasıl hazırlanacağı ve animasyonların nasıl sonlandırılması gerektiği ile ilgili detaylı anlatımlar animatöre aktarılmıştır. Seslendirme için bir kadın ve bir erkek olmak üzere seslendirme alanında uzman iki kişiden yardım alınmıştır. Animatöre ve seslendirme ekibine animasyonların ve hikayelerin kelime sayıları üzerinden ücret ödemeleri gerçekleştirilmiştir.

Animasyonların oluşturulması

Hazırlıklar adımı sonrasında animatör, animasyonları oluşturmaya başlamıştır. Bu süreçte araştırmacı süreci yakından takip ederek animatörle sürekli iletişim halinde kalmıştır. Kanıt değerlendirme beceresini ölçmeye yönelik geliştirilen maddelerde özellikle deney düzeneklerinin doğru bir şekilde sunulması oldukça önemlidir. Bu yüzden deney düzeneklerinin animasyona dönüştürülmesi büyük bir titizlikle

ilerletilmiştir. Animasyonlar; Adobe Character Animator, Adobe Illustrator ve Adobe After Effect yazılımları kullanılarak oluşturulmuştur. Animatör, maddelerden biri olan *Mıknatıs* animasyonunu tamamladıktan sonra araştırmacıya elektronik ortamda göndermiştir. Araştırmacı tarafından madde üzerinde düzeltmeler yapıldıktan sonra animatör maddeye son halini vermiştir. Yaratıcılık alanında uzman bir kişi tarafından *mıknatıs* maddesinin ilk kontrolü sağlanmıştır. Yapılan değerlendirme sonucunda alan uzmanı madde üzerinde bazı değişiklikler önermiş ve gerekli düzenlemeler yapılmıştır. Süreç sonunda; Çalar Saat, Bıçak, Mıknatıs, Şeker, Yuvarlanma, Terleme, Yağmur, Balon ve Gölge adlı toplam 9 animasyon oluşturulmuştur. Animasyonların süresi 77 sn ile 98 sn arasında değişkenlik göstermekte, toplam süre ise yaklaşık 12 dakikadır. Animasyonların oluşturulması yaklaşık iki aylık bir süreçte tamamlanmıştır.

Animasyonların uzmanlarca değerlendirilmesi

Animasyonlar oluşturulduktan sonra özel yetenekliler ve yaratıcılık alanlarında iki ve animasyon alanında bir uzmanın görüşlerine başvurulmuştur. Uzmanlar, animasyonlarda kullanılan görsellerin ve hikayelerin anlaşılabilirliği, seslendirmelerin ve arka planda yer alan çocuk şarkılarının nitelikleri ve olay örgülerinin bütünsellikleri bağlamında görüş bildirmişlerdir. Uzmanlardan gelen dönütler sonrasında animatör, animasyonlarda düzeltmeler yapmış ve araştırmacı ön değerlendirme aşamasına geçmiştir.

3.3.1.4.5 Ön değerlendirmenin uygulanması

Ölçek geliştirme adımlarında bilişsel görüşme (cognitive interviewing) olarak da adlandırılan bu aşamada, katılımcıların maddeleri nasıl anladıklarını ve yorumladıklarını belirlemek adına nitel veri toplanmaktadır (Boeteng vd., 2018; Devellis ve Thorpe, 2022). Amaç, maddeler bağlamında araştırmacı ve katılımcılar arasında ortak bir dil oluşturmaktır. Ayrıca bu adımda araştırmacı, ölçeğin uygulama süresi hakkında bilgi sahibi olur. Willis, (2005) ve Tourangeau, (2003) bilişsel görüşmenin avantajlarını şu şekilde özetlemektedirler:

- Maddelerin hedefe uygunluğu
- Katılımcılar için anlaşılabilirliği güç olan maddelerin belirlenmesi ve maddelerde netlik sağlanması

- Cevaplanması zor soruların belirlenmesi
- Yanıtların netleştirilmesi
- Katılımcılardan maddelerdeki alan bilgileri hakkında bilgi alınması
- Maddelerin zorluk seviyelerine göre sıralanışları hakkında bilgi alınması

Ön değerlendirmede, geliştirilen 9 animasyon Eskişehir il merkezinde yer alan bir okulun anasınıfı, birinci ve ikinci sınıflarına uygulanmıştır. Uygulamaya toplam 54 öğrenci katılmıştır. Katılımcıların 16'sı anasınıfından 20'si birinci sınıftan ve 18'i ikinci sınıf öğrencilerinden oluşmaktadır. Ön değerlendirme sürecinde katılımcılar, seçkisiz olmayan örnekleme yöntemlerinden uygun örnekleme tekniği ile belirlenmiştir (Büyüköztürk vd., 2017). Katılımcıların belirlenmesinde rastgele bir seçimden ziyade öğretmen görüşüne başvurulmuştur. Sınıf öğretmenlerinin görüşlerine dayanarak sınıflarda bulunan yüksek, orta ve düşük öğrenme (kaynaştırma öğrencileri gruba dahil edilmemiştir) profiline sahip öğrenciler ön değerlendirme için araştırmaya dahil edilmişlerdir. Böylelikle maddelerin yanıtlanmasında yanlı bir eğilimin önüne geçilmesi ve ölçek maddelerinin farklı düzeylerde öğrenme profiline sahip öğrencilere ulaşması hedeflenmiştir. Ön değerlendirme 2024 yılının Ocak-Şubat ayları arasında gerçekleştirilmiştir.

Ön değerlendirmede 9 animasyon her öğrenciye bireysel olarak tek oturumda uygulanmış ve değerlendirme ortalama 20-25 dakika sürmüştür. Test sırasında bazı öğrencilerin hazırlanan deney düzeneklerinde, araştırmacının ön göremediği birtakım noktalara dikkat ettikleri görülmüştür. Örneğin deney düzeneklerinde kullanılan şeker renginin ve terlemenin görselleştirildiği su buharının eksik ya da yanlış anlaşıldığı görülmüştür. Yuvarlanma deney düzeneklerinde kullanılan kumun bazı öğrenciler tarafından net anlaşılamadığı, çalar saat deney düzeneklerinde kanıt için sunulan saat ses düzeylerinin birbirlerine çok yakın olması sebebiyle öğrenciler tarafından ayırt edilememesi fark edilen diğer noktalardır. Ön değerlendirme sırasında maddelere ilişkin detaylı notlar alınmıştır. Alınan notlar, düzeltme yapılması amacıyla gün sonunda animatöre iletilmiştir. Düzenlemeler sonrasında yapılan değerlendirmelerde, maddelere ilişkin yanlış ya da eksik yorumlamaların önüne geçildiği görülmüştür. Gölge ve Yağmur maddelerinde ise özellikle anasınıfı öğrencilerinin yeterince yanıt üretemedikleri fark edilmiştir. Mıknatıs, Bıçak ve

Çalar Saat, tüm sınıf düzeyleri kapsamında en çok yanıtın üretildiği maddeler olmuştur.

Ön değerlendirmede birinci ve ikinci sınıf öğrencileri daha konforlu davranışlar sergilerken, bazı anasınıfı öğrencilerinin daha çekingen oldukları gözlemlenmiştir. Birkaç öğrenci dışında neredeyse tüm öğrenciler, çizgi filmleri ve etkinlikleri çok sevdiklerini ifade etmişlerdir. Ön değerlendirme sürecinde ölçekten herhangi bir madde atılmamıştır. Kanıt değerlendirme becerisine yönelik 9 maddenin tamamı ana uygulamada kullanılmıştır.

3.3.2 SAGES 2 (Screening Assessment for Gifted Elementary and Middle School Students-Second Edition-Üstün Yetenekli İlkokul ve Ortaokul Öğrencileri için Tarama Değerlendirme Bataryasının İkinci Baskısı)

Mevcut çalışmada ÇABİYAT-R'nin uyum geçerliğini sağlamak için SAGES 2'nin fen bilimleri alt testi uygulanmıştır (Johnsen ve Corn, 2001). Üstün yetenekli öğrencileri anaokulundan sekizinci sınıfa kadar belirlemede kullanılan SAGES 2, öğrencileri akademik başarı ve yetenek temel alanlarında üç alt test ile değerlendirmektedir. Alt testlerden birisi matematik/fen bilimleri diğeri dil bilimleri/sosyal bilimlerden oluşmaktadır. Bu iki alt test öğrencilere çoktan seçmeli şeklinde sunulmaktadır. Söz konusu iki alt test öğrencinin sahip olduğu bilgi birikimine ve akademik başarısına dayanmaktadır. Muhakemenin ölçüldüğü üçüncü alt test ise öğrencilerin yetenek ve problem çözme becerilerine odaklanmaktadır. Muhakeme alt testinde öğrencilerden resim ve şekiller arasındaki ilişkileri keşfetmeye dayalı analogik problemleri çözmeleri beklenmektedir. Bataryanın uygulanması yaklaşık 20 dakika sürmektedir. Batarya puanları, bir bütün halinde yorumlanabileceği gibi alt testler tek tek de yorumlanabilmektedir. Bataryanın tüm alt ölçekleri sınıf düzeyine göre; anasınıfı-üçüncü sınıf ve dördüncü sınıf-sekizinci şeklinde iki ayrı form olarak sunulmaktadır: Çalışma kapsamında, SAGES 2'nin anasınıfı-üçüncü sınıf fen bilimlerine yönelik geliştirilen alt ölçeği kullanılmıştır. SAGES 2'nin, 1998 ve 1999 yıllarında sırasıyla 3023 ve 2290 öğrencinin katılımı ile norm çalışmaları yapılmıştır. Pratik bir uygulamaya sahip olan testin geçerli ve güvenilir olduğu belirtilmektedir. Yapılan analizler sonucunda test maddelerinin güvenirlik katsayıları .77 ile .95 arasında raporlanmıştır (Johnsen ve Corn, 2001).

SAGES 2'nin fen bilimleri alt ölçeğinin uyarlama süreci Ateşgöz (2020) tarafından gerçekleştirilmiştir. Uyarlama süreci çeviri, pilot çalışma ve geçerlik güvenirlik çalışmaları olmak üzere üç adımda yürütülmüştür. İlk adımda, İngilizce dil alanında uzman dört kişilik bir ekip fen bilimleri alt testinde yer alan maddeleri Türkçe'ye çevirmiştir. Ardından yaratıcılık, fen bilimleri ve okul öncesi öğretmeninden oluşan dört kişilik bir başka ekip, maddelerin ve görsellerin anlaşılabilirliği, hedef gruba ve kültüre uygunlukları bağlamında görüş bildirmişlerdir. Çeviri aşamasından sonra 5 yaş grubundan 15 kişilik küçük bir gruba pilot çalışması yapılmıştır. Pilot uygulamadan gelen veriler doğrultusunda test maddelerinde küçük düzeltmeler yapılarak geçerlik ve güvenirlik analizlerinin yapılması için büyük bir örneklem grubundan veri toplanmıştır. 801 öğrencinin katıldığı uygulamanın analizleri sonucunda madde güçlük değerlerinin .29 ile .83 arasında değiştiği; testin toplam güçlük indeksinin .58 olduğu; maddelerin ayırt edicilik indekslerinin .30 ile .59 arasında farklılaştığı raporlanmıştır (Ateşgöz, 2020). Testin madde güvenirlik analizi için Cronbach Alpha katsayı değeri .66 olarak hesaplanmıştır. Bu değerin görece düşük olduğu söylenebilir. Madde sayısının az olması Cronbach Alpha değerini etkilemiş olabilir. SAGES2 bataryasının yalnızca bir alt boyutunun uyarlanmış olması da güvenirliğin görece düşük çıkmasının nedenleri arasında gösterilebilir. Güvenirlik analizlerinde öncül araştırmalar için kabul edilebilir değer olarak .60 (Nunnally, 1978), iyi uyum değeri olarak .70 ve üzeri Cronbach Alpha değerleri önerilmektedir (Nunnally ve Bernstein, 1994). Fen bilimleri alt ölçeği uyarlanan SAGES-2'nin kabul edilebilir düzeyde güvenirliğe sahip olduğu ifade edilebilir.

3.4 Veri Toplama Süreci

ÇABİYAT'ı yenileme sürecinde son adım ana uygulamayı kapsamaktadır. Ana uygulama öncesinde Eskişehir İl Milli Eğitim Müdürlüğü'nden araştırmanın yürütülmesine ilişkin gerekli izinler alınmıştır. Uygulama öncesi Eskişehir ilinde belirlenen araştırmanın yapılacağı A ve B okulları ziyaret edilerek okul yönetimi ile çalışmanın içeriği ve süreci hakkında bir toplantı yapılmıştır. Okul idaresinden bir oda ve sınıf öğretmenlerinin süreç hakkında bilgilendirilmesi talep edilmiştir. Araştırma katılımına ilişkin veli izin formları öğrenci sayısı kadar çoğaltılarak okul idaresine teslim edilmiştir (Ek 3). Velilerden gelen izin belgeleri doğrultusunda, velisi bulunduğu öğrencinin çalışmaya katılması uygun bulunmayan öğrenci, çalışma

kapsamı dışında bırakılmıştır. Uygulama 15.03.2024 - 15.06.2024 tarihleri arasında 3 aylık bir zaman diliminde, eğitim-öğretimin devam ettiği her gün saat 9.00 – 17.00 arasında gerçekleştirilmiştir. Uygulamalar A okulunda boş bir odada, B okulunda ise Rehber Öğretmenin odasında yapılmıştır. Odaların fiziki şartlarının bireysel değerlendirmeye uygun oldukları görülmüştür.

ÇABİYAT-R; 9 kanıt değerlendirme, 3 deney tasarlama ve 3 hipotez geliştirme becerisini ölçen toplam 15 maddeden oluşmaktadır. Küçük yaş gruplarının dikkat süreleri hesaba katıldığında, testin uzun olacağı ve öğrencilerin motivasyonlarının etkilenebileceği öngörülmüştür. Bu yüzden test ikiye bölünmüş ve her bir öğrenciye iki ayrı form şeklinde, iki ayrı oturumda uygulanmıştır. İlk form hipotez geliştirme ve deney tasarlama görevlerinin yer aldığı 6 maddeden, ikinci form ise kanıt değerlendirme görevlerinin yer aldığı 9 maddeden oluşmaktadır. Öğrenciler sınıflarından bireysel çağırılarak ÇABİYAT-R'nin ilk formu ve SAGES 2 bataryası ilk oturumda uygulanmıştır. Bu oturum ortalama 20-25 dakika sürmüştür. Tüm öğrenciler için ilk oturum tamamlandıktan sonra ikinci oturum için her bir öğrenci tekrar çağırılmıştır. İlk oturum ile ikinci oturum arasında yaklaşık olarak bir buçuk aylık süre bulunmaktadır. İkinci oturum yaklaşık olarak 15-20 dakika sürmüştür. Uygulamada A okulundan anasınıfı düzeyinde 70, birinci sınıftan 95 ve ikinci sınıftan 87 öğrenciye; B okulundan yalnızca anasınıfı düzeyinde 23 öğrenciye ulaşılmıştır. İkinci oturumun uygulanamadığı anasınıfı, birinci ve ikinci sınıftan toplam 23 öğrenci ise veri setine dahil edilmemiştir. Sonuç olarak ana uygulama kapsamında 252 öğrenci ile veri toplama süreci tamamlanmıştır.

3.5 Verilerin Analizi ve Yorumlanması

Ana uygulama sonunda elde edilen veriler araştırma soruları kapsamında analiz edilmiştir. Testin psikometrik özellikleri incelenmeden önce, kanıt değerlendirme becerisine yönelik revize sürecinde geliştirilen 9 maddeden seçilmesi planlanan 3 maddeye ilişkin madde analizlerine yer verilmiştir. Madde analizlerinde; betimsel istatistik, madde ayırt edicilik, maddeler arası korelasyon, madde güvenirlik ve regresyon analizleri gerçekleştirilmiştir. Madde seçimi tamamlandıktan sonra, testin son halinde yer alan 9 madde ile (3 kanıt değerlendirme, 3 deney tasarlama ve 3 hipotez geliştirme) testin psikometrik özelliklerini değerlendirmek üzere geçerlik ve güvenirlik analizleri yapılmıştır.

3.5.1 Madde analizleri

Testte yer alacak maddelerin istatistiksel hesaplarının yapılması, teste dahil edilecek ya da çıkarılacak maddelerin belirlenmesi amacıyla madde analizlerinin yapılması önerilmektedir (Raubenheimer, 2004). Madde analizleri SPSS 23.0 istatistik analiz programı ile yürütülmüştür. Betimsel istatistik analizlerinde sınıflara göre her maddenin minimum ve maksimum değerleri ve sıklıkları, madde ortalamaları ve standart hataları, maddelerin açıkladıkları varyans değerleri, çarpıklık ve basıklık katsayıları yorumlanmıştır. Maddeler arası ayırt edicilik için öğrencilerin her bir maddeden elde ettikleri akıcılık puanları sınıf düzeylerine göre kıyaslanmıştır. Madde seçiminde bir diğer yöntem olarak basit regresyon analizleri önerilmektedir (Desboulets, 2018; Tezbaşaran, 2008). Bu kapsamda öğrencilerin kanıt değerlendirme alt ölçeğinde yer alan her bir madde için aldıkları akıcılık puanı ile testin tamamından aldıkları toplam akıcılık puanları hesaplanmıştır. Sonrasında kanıt değerlendirme madde akıcılık puanının, test toplam akıcılık puanını yordama derecesine bakılmıştır. En yüksek regresyon katsayısına sahip maddeler sıralanarak (Özbaşı, 2024), madde seçimi için ek kanıtlar oluşturulmuştur. Madde seçiminde diğer bir kriter olarak erkek ve kız öğrencilerinin puanları arasındaki ortalamalara bakılmış ve bir grup lehine anlamlı bir fark olup olmadığı kontrol edilmiştir. Yapılan analiz cinsiyet bağlamında madde yanlılığından ziyade öncül bir analiz olarak gerçekleştirilmiş, kanıt değerlendirme maddeleri için gruplar arasında büyük farklar olup olmadığını öğrenmek amacıyla yürütülmüştür. Madde seçiminde son olarak maddeler arası korelasyon katsayıları hesaplanmış ve madde güvenilirlik analizleri gerçekleştirilmiştir.

3.5.2 Geçerlik analizleri

Geçerlik çalışmaları genel olarak kapsam geçerliği, yapı geçerliği ve ölçüt geçerliği şeklinde sınıflandırılmaktadır (Devellis ve Thorpe, 2022; Hair vd., 2019; Huck, 2008). Çalışmada kapsam geçerliği kapsamında, madde havuzu oluşturulma sürecinin başından animasyonların oluşturulması süreci sonuna kadar alan uzmanlarının görüşleri alınmıştır. Yapı geçerliği kapsamında açımlayıcı grafik analizi, açımlayıcı yapısal eşitlik modellemesi, yakınsama ve ıraksama geçerliklerine bakılmıştır. Ölçüt geçerliği kapsamında ise uyum geçerliği ve bilinen gruplar

geçerliğine bakılmıştır. Geçerlik analizleri kapsamında ayrıca cinsiyete göre ölçme değışmezliğı analizleri de değerdendirilmiştir.

3.5.2.1 Yapı geçerliğı

Cronbach ve Meehl (1955) yapı geçerliğini, bir değışkenin diğerd değışkenler ile teorik ilişkisi olarak yorumlamaktadırlar. Eğerd ölçek net bir kuramsal yapıya sahip değild ise yapı geçerliğinden söz edilemez (Cronbach ve Meehl, 1955). Uygulamalı araştırmalarda yapı geçerliğı sıklıkla faktör analizi yöntemleri ile ilişkilendirilmektedir (Brown, 2015). Alan yazın incelendiğinde yapı geçerliğı için açımlayıcı ve doğrulayıcı faktör analizlerinin ardı sıra uygulanması önerilmektedir (Kline, 2011; Tabachnick ve Fidell, 2013). Mevcut araştırmada yapı geçerliğıin sağlanması için Açımlayıcı Grafik Analizi (Exploratory Graph Analysis- EGA, Golino ve Epskamp, 2017) ve Açımlayıcı Yapısal Eşitlik Modellemesinden (Exploratory Structural Equation Models- ESEM, Asparouhov ve Muthen, 2009) yararlanılmıştır. EGA, son yıllarda çok değışkenli verilerde kullanılan bir psikometrik ağ yaklaşımıdır (Christensen ve Golino, 2021). Açımlayıcı Yapısal Eşitlik Modellemesi (AYEM) ise temelde doğrulayıcı bir teknik olmakla birlikte aynı zamanda açımlayıcı öğeleri de barındıran, son yıllarda uygulamalı olarak test edilmiş bir analiz türüdür (Marsh vd., 2009; 2016).

3.5.2.1.1 Açımlayıcı grafik analizi- EGA

Çok değışkenli veri kümelerinde boyut ya da faktör sayılarını belirlemek, psikometri biliminde uzun süredir tartışılan konulardan biri olmuştur (Preacher ve MacCallum, 2003; Ruscio ve Roche, 2012; Timmerman ve Lorenzo-Seva, 2011; Velicer ve Jackson, 1990). 1960'lı yıllardan itibaren bireylerin kişilik özellikleri, davranışları ya da zekâ ölçümleri ile ilgili psikoloji alanlarında gerçekleştirilen çalışmalarda temel boyut sayısının doğru tahmini için çeşitli teknikler kullanılmıştır. Kaiser-Guttman'ın K1 Kuralı (Guttman, 1954; Kaiser, 1960) Paralel Analiz tekniğı (Horn, 1965), Birikinti Yamaç Grafiğı (Cattell, 1966), En Küçük Ortalamalı Kısmi Korelasyon Tekniğı (MAP- Velicer, 1976) ve Bayesian Bilgi Kriteri (BIC- Schwarz, 1978) yaygın olarak kullanılan geleneksel faktör belirleme tekniklerindendir. Kullanılan tekniklerin ise birtakım sınırlılıklara sahip olduğu belirtilmektedir. K1 kuralı olarak bilinen Kaiser ölçütünün, olması gerekenden çok sayıda faktör sayısı üretmesi (Zwick ve Velicer, 1986), yamaç birikinti grafiğı tekniğinin öznel

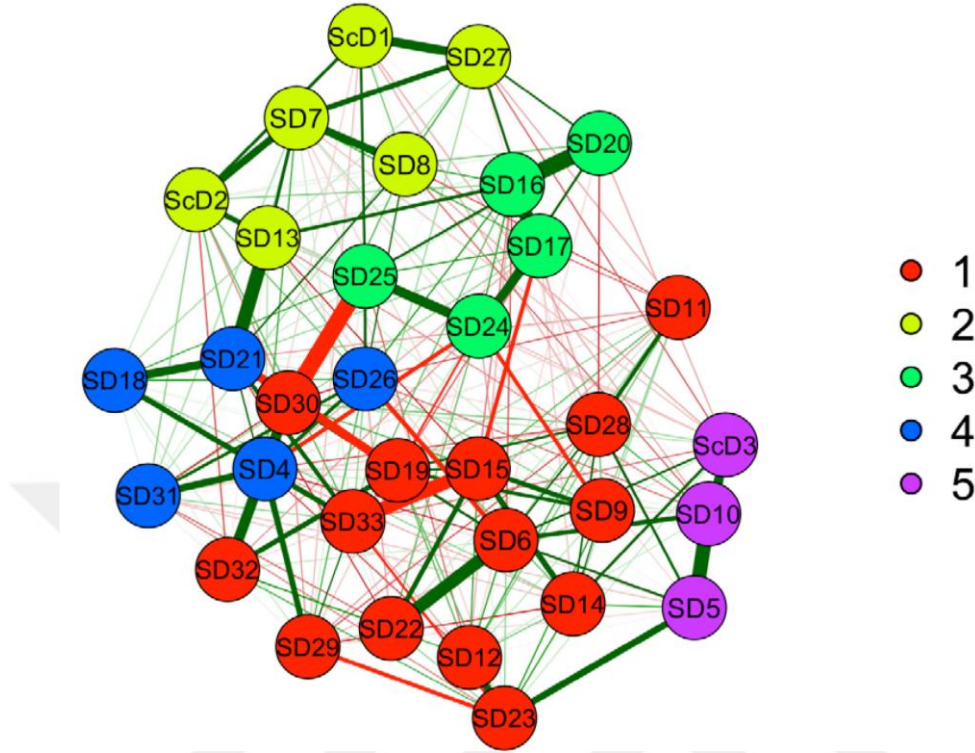
yorumlama hatalarına sahip olması (Ledesma ve Valero-Mora, 2007), paralel analiz tekniğinin rastgele değerler üzerinden kıyaslama yapması (O'connor, 2000), BIC tekniğinin örneklem büyüklüğünden etkilenmesi (Burnham ve Anderson, 2004), MAP tekniğinin veri setindeki varyans değerlerini gözden kaçırmaması (Zwick ve Velicer, 1986) ve diğer tekniklere göre karmaşık olması (O'connor, 2000) söz konusu sınırlılıklar kapsamında değerlendirilmektedir.

Faktör sayısını belirlemede ortaya çıkan sınırlılıkların üstesinden gelebilmek amacıyla, ağ psikometrisine ve modellemesine dayanan yeni bir yaklaşım olarak EGA önerilmektedir (Golino ve Epskamp, 2017). EGA, veri setlerinde boyut sayısını tahmin ederek sunduğu grafiksel model ile veri kümelerini bir araya toplamaya çalışan alternatif bir faktör belirleme tekniği olarak kullanılmaktadır. Ağ modellemeleri, *kümeler* ve kümeler arasında kısmi korelasyonu gösteren *çizgiler (kenar)* ile gösterilmektedir. Kümeler, gözlenen değişkenleri temsil ederken kümelerden oluşan topluluklar ise bir faktöre işaret etmektedir. Örnek bir EGA görseline Şekil 3.3' te yer verilmiştir.

EGA ile geleneksel yöntemlerin karşılaştırıldığı simülasyon çalışmalarında, EGA'nın faktör sayısı belirlemede diğer tekniklere kıyasla oldukça iyi performans gösterdiği rapor edilmiştir (Golino vd., 2020; Golino ve Epskamp, 2017; Cosemans et al., 2022; Markos ve Tsigilis, 2024).

ÇABİYAT-R testi net bir kuramsal alt yapıya sahip olup faktör sayısı ve faktörler altında toplanacak madde sayıları bağlamında güçlü bir önsel (a priori) bilgiye dayanmaktadır. Bu kapsamda test için açıklayıcı faktör analizi yapılmadan; EGA ile madde boyut arasındaki ilişkiye bakılarak yapıya ilişkin genel bir çerçeve sunulmaya çalışılmıştır. Huck (2008), yapılması planlanan faktör analizi türü ne olursa olsun analizin yapılabilmesi için öncelikli olarak varsayımların kontrol edilmesi gerektiğini vurgulamaktadır. Bu bağlamda kayıp veri, örneklem büyüklüğü, tek ve çok değişkenli normallik, doğrusallık, uç değerlerin kontrolü, çoklu doğrusallık ve tekillik varsayımlarının karşılanıp karşılanmadığı kontrol edilmiştir (Tabachnick ve Fidell, 2013).

Şekil 3.5 Örnek bir EGA grafiği



Kaynak: Golino vd., 2020, s. 44

Yapılan betimsel analiz bulguları doğrultusunda 9 madde için herhangi bir kayıp veriye ulaşılmamıştır. Faktör analizleri için örneklem büyüklüğünün belirlenmesi önemli bir adımdır (Akbulut, 2010). Literatürde örneklem büyüklüğünün belirlenmesine yönelik çeşitli görüşler bulunmaktadır. Kimi araştırmacılar doğrudan olması gereken örneklem sayılarını paylaşıırken (Comrey, 1988; Guadagnoli ve Velicer, 1988); kimi araştırmacılar madde-birey bağlamında önerilerde bulunmuşlardır (Floyd ve Widaman, 1995; Streiner, 1994). ÇABİYAT-R testinin psikometrik özellikleri doğrulayıcı faktör analizinin temel alındığı bir yaklaşımla incelenmiştir. Literatürde doğrulayıcı faktör analizi için 200 kişiye ulaşılmasının kabul edilebilir düzeyde olduğu belirtilmektedir (Anderson ve Gerbing, 1984; Jackson, 2001). Örneklem büyüklüğü belirlemede önerilen diğer bir yöntem ise simülasyon çalışmalarının kullanılmasıdır (Walf vd., 2013). Çalışmada modelin test edilmesi için gerekli olan minimum örneklem büyüklüğü, yapısal eşitlik modelleri için geliştirilen *a priori örneklem büyüklüğü hesaplama aracı*

kullanılarak belirlenmiştir (Soper, 2024). Modelin son hali 9 gözlenen değişken ve 3 gizil faktör ile test edilmiştir. Hesaplama aracında bekl

enen etki büyüklüğü 0.3, anlamlılık düzeyi 0.05 ve istatistiksel güç 0.8 olarak belirlenmiştir. Hesaplama sonucunda, modelin test edilebilmesi için önerilen minimum örneklem büyüklüğünün 200 olduğu görülmüştür. Çalışmada 252 kişiye ulaşılmıştır. Bu sayı, doğrulayıcı faktör analizleri için ulaşılması gereken örneklem büyüklüğü kriterini (Anderson ve Gerbing, 1984; Jackson, 2001), madde-birey ölçütünü (Cohen vd., 2013; Goodwin, 1999; Stevens, 2002) ve simülasyon sonucunda önerilen örneklem büyüklüğü kriterini (Soper, 2024) karşılamaktadır. Örneklem büyüklüğünün belirlenmesinde bir diğer adım Kaiser-Meyer-Olkin (KMO) değerinin hesaplanmasıdır. KMO ile Barlett Küresellik Test sonuçlarının da raporlanması gerektiği önerilmektedir (Watkins, 2018). Çalışmada KMO değeri, .848 olarak hesaplanmış ve Barlett testinin anlamlı çıktığı görülmüştür [$\chi^2(36) = 750,588$; $p < .001$]. Hesaplanan KMO değeri, faktör analizinin yapılabilmesi için belirlenen örneklem sayısının mükemmel düzeyde olduğuna işaret etmektedir (Kaiser ve Rice, 1974).

Tek değişkenli normallik varsayımı çarpıklık ve basıklık katsayıları üzerinden değerlendirilmiştir. Çalışmada maddelerin çarpıklık katsayı değerleri .183 ile .945; basıklık katsayı değerleri ise -.692 ile .867 arasında değişmektedir. Huck (2000) çarpıklık katsayısının -1 ile +1, basıklık katsayısının ise -1 ile +2 arasında olması gerektiğini belirtmektedir. Bu bağlamda elde edilen çarpıklık ve basıklık katsayılarının tek değişkenli normallik varsayımını sağladığı görülmüştür.

Çok değişkenli normallik varsayımı için Mardia'nın çarpıklık ve basıklık katsayıları incelenmiş (Mardia, 1970) ve Energy Test yapılmıştır (Székely ve Rizzo, 2005). *Mardia çarpıklık* = 247.22, $p < .001$, *Mardia Basıklık* = 2.54, $p < .001$ ve Energy Testi $E = 2.43$, $p < .001$ olarak hesaplanmıştır. Elde edilen sonuçların istatistiksel olarak anlamlı oldukları görülmüştür. Sonuçların anlamlı çıkması çok değişkenli normallik varsayımlarının karşılanamadığı şeklinde yorumlanmaktadır (Mardia, 1970; Székely ve Rizzo, 2005). Söz konusu varsayımın sağlanamaması ise bir sorun olarak görülmemektedir (Tabachnick ve Fidell, 2019). Elde edilen sonuç, ileriki adımlarda yapılacak olan faktör analizlerinde doğru kestirim yönteminin belirlenmesi için önemlidir.

Doğrusallık varsayımının karşılanması saçıma diyagramları incelenmiştir. Değişken çiftleri arasında incelenen saçıma diyagramlarında noktaların düz çizgiye paralel şekilde dağıldıkları görülmüştür. Bu bağlamda doğrusallık varsayımının sağlandığı söylenebilir. Çalışmada uç değerlerin varlığı, veri setinde Mahalanobis ve Cook uzaklık değerleri tespit edilerek belirlenmiştir. Tabachnick ve Fidell (2013) 9 değişken için $p < .01$ güven aralığında, Mahalanobis kritik değerinin 21,67 olduğunu belirtmektedirler. Cook değerinin ise 1'den yüksek olması gerektiği kabul edilmektedir (Cook, 1977). Bu bağlamda belirlenen kriterlere uymayan 8 öğrenci analiz dışında bırakılarak örneklem sayısı 244'e düşürülmüştür.

Çoklu doğrusallık ve tekillik varsayımlarının kontrolü; maddeler arası korelasyon katsayıları, TOL (tolerance), VIF (Variance Inflation Factor) ve CI (Condition Index) değerleri üzerinden incelenmiştir. Maddeler arası korelasyon katsayılarının .196 ile .585; tolerans değerlerinin 0.55 ile 0.64, VIF değerlerinin 1.56 ile 1.82 ve CI değerlerinin 1.00 ile 9.53 arasında değiştikleri görülmüştür. Çoklu doğrusallık ve tekillik varsayımlarının karşılanabilmesi için maddeler arası korelasyon katsayılarının .90'dan düşük (Tabachnick ve Fidell, 2013), TOL değerlerinin .10'dan büyük (Hair vd., 2019), VIF değerlerinin 10'dan (Kline, 2011) ve CI değerlerinin ise 30'dan küçük olması gerekmektedir (Belsley vd., 1980). Bu bağlamda varsayımların karşılandığı ve faktör analizleri için yeterli kanıtın sağlandığı söylenebilir.

Varsayımların karşılanmasının ardından EGA analizi R ortamında (R Core Team, 2019), Golino ve Christensen (2020) tarafından geliştirilen “*EGAnet*” paketi kullanılarak gerçekleştirilmiştir. EGA analizi Glasso (Grafiksel En Küçük Mutlak Daralma ve Seçim Operatörü- Graphical Least Absolute Shrinkage and Selection Operator) algoritması ile yürütülmüştür. Glasso algoritması, değişkenler arasında kısmi korelasyonların varlığına dayanmaktadır. Bu yöntem, değişkenler arasındaki zayıf ilişkilerin önüne geçerek daha kararlı yapılar oluşturmayı hedefler (Epskamp vd., 2012; Friedman vd., 2012; Golino vd. 2020). ÇABİYAT-R alt testlerinin teorik olarak birbirleriyle ilişkili olmaları sebebiyle Glasso algoritmasının kullanılması uygun görülmüştür. Glasso algoritması ile EGA analizi gerçekleştirilerek grafiksel model ve matris kenar ağırlıkları hesaplanmıştır.

EGA modelinde tanımlanan yapının, boyutların ve maddelerin tutarlılığını ortaya koymak için yeni bir yaklaşım olarak Bootstrap Exploratory Graph Analysis

(BootEGA) önerilmektedir (Christensen ve Golino, 2021). BootEGA yaklaşımı, EGA sonuçları arasında boyutların tutarlılığı ve maddelerin kararlılığını değerlendirmek için kullanılmaktadır. Bu yaklaşım ile benzer veri setleri üzerinden yapılan analizler sonucunda boyutlar arasında bir dalgalanma olup olmadığı kontrol edilmektedir (Christensen ve Golino, 2021). Bu bağlamda ÇABİYAT-R testinin madde kararlılığı ve yapı tutarlılığının BootEGA analizi ile test edilmesine karar verilmiştir. Analiz; R ortamında “EGAnet” paketi, “bootEGA” fonksiyonu ve veri setinde yer alan aynı sayıda vakanın 500 kez tekrarlanması ile yürütülmüştür.

3.5.2.1.2 Açımlayıcı yapısal eşitlik modellemesi

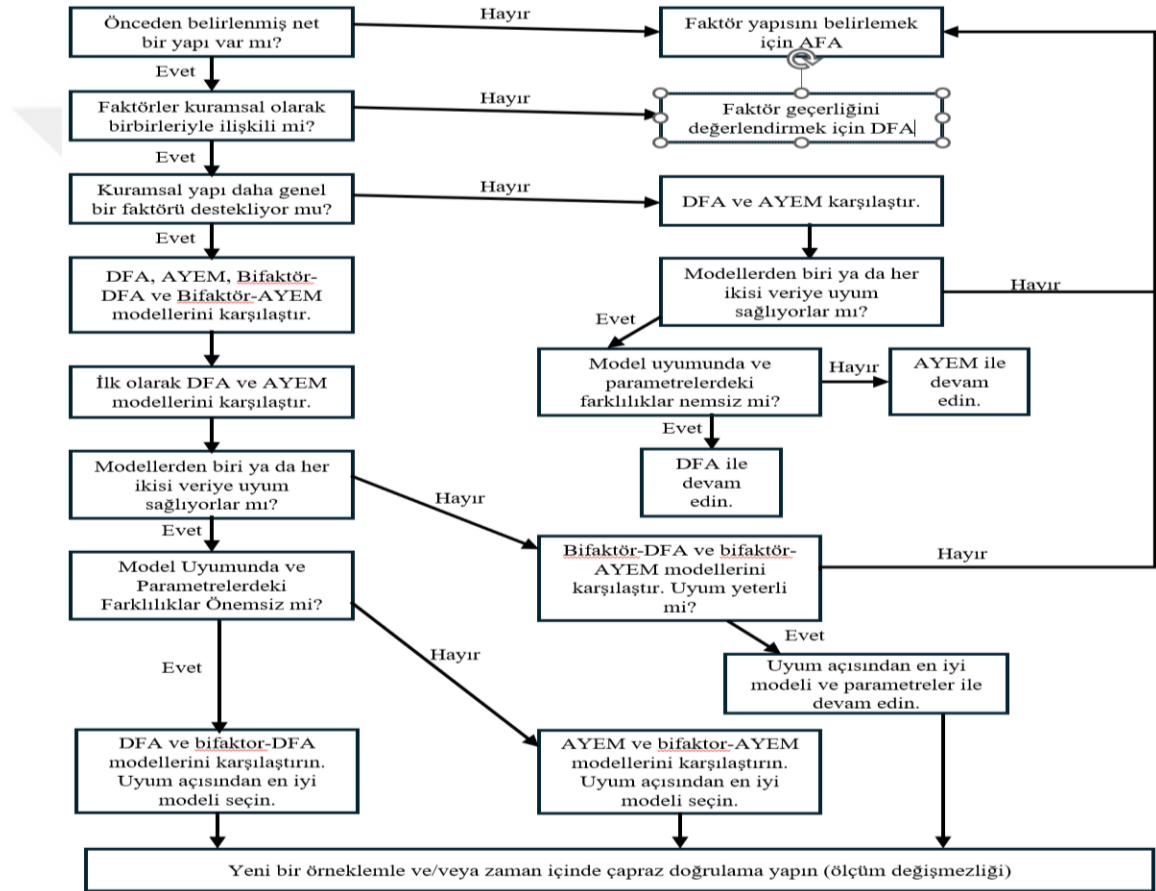
Araştırmalarda farklı türden faktör analiz teknikleri kullanılmasına rağmen en bilinenleri açımlayıcı (AFA) ve doğrulayıcı (DFA) olarak kabul edilmektedir (Huck, 2008). AFA ve DFA analizlerindeki temel ayrım hipotezin varlığı ile ilgilidir. Eğer araştırmacı analizin ortaya ne koyacağı ile ilgili hiçbir önseziye sahip değilse AFA, analizin yapılabilmesine hipotez rehberlik ediyorsa DFA kullanılmaktadır (Huck, 2008). Açımlayıcı yapısal eşitlik modellemesi (AYEM) ise AFA ve DFA’nın sunduğu avantajları bir araya getiren istatistiksel bir tekniktir (Asparouhov ve Muthén, 2009; Marsh ve Alamer, 2024). AYEM karmaşık faktör yapılarını ortaya çıkarırken aynı zamanda model uyumunu ve modelin değerlendirilmesine olanak tanımaktadır (Morin vd., 2013).

DFA, gizil faktörler arasındaki çapraz yüklerin 0’ a eşitlenmesi (Marsh, 2007; Morin vd., 2016), faktörlerin katışıksız yüklere sahip olmaları (Asparouhov ve Muthén, 2009; Marsh vd., 2009), hedeflenen faktör altında yer alan maddenin diğer maddeler ile bir ilişkiye sahip olmadığı varsayımına dayandırılması sebebiyle (McDonald, 2014) eleştirilmektedir. Çok boyutlu yapıların değerlendirilmesinde ölçekteki madde çapraz yüklerinin sıfıra eşitlenmelerinin, gerçekçi ve pratikte bir karşılığının olmadığı (Marsh vd., 2014; Morin vd., 2016, 2020), sonuç olarak DFA ile yürütülen analizlerin şişirilmiş faktör yüklerine ve olması gerekenden daha iyi uyum sağlayan modellemelerin oluşturulmasına neden olabileceği savunulmaktadır (Asparouhov ve Muthén, 2009; Morin vd., 2020). AYEM’ in söz konusu sınırlılıkların önüne geçebileceği düşünülmektedir. AYEM; keşfedici yönü ile maddeler arasında çapraz yüklemelere izin veren, daha esnek modellerin oluşturulduğu, doğrulayıcı bir yaklaşım sunmaktadır (Asparouhov ve Muthén, 2009; Marsh, vd., 2014; Morin vd.,

2014). Son dönemlerde AYEM ve DFA analizlerinin psikoloji ve eğitim alanlarında karşılaştırıldığı pek çok araştırmaya rastlanmıştır. Araştırma sonuçlarına göre AYEM' in, DFA' ya kıyasla daha iyi bilgilendirici olduğu raporlanmıştır (Alamer ve Marsh, 2022; Guay vd., 2015; Kruk vd., 2023). ÇABİYAT-R testinin yapı geçerliği DFA ve AYEM ile sağlanmıştır. AYEM' in hangi şartlarda kullanılabileceğine ilişkin şema Şekil 3.6' da yer almaktadır.

Şekil 3.6

AYEM analizi için karar ağacı



DFA: Doğrulayıcı faktör analizi. AFA: Açımlayıcı faktör analizi: AYEM: Açımlayıcı yapısal eşitlik modellemesi.

Kaynak: Alamer ve Marsh, 2022, s.1482

Şema incelendiğinde ÇABİYAT-R' nin net bir kuramsal yapıya sahip olması, faktörlerinin kuramsal olarak birbirleri ile ilişkili olmaları ve üst düzey bir faktör barındırmasından dolayı ÇABİYAT-R için DFA ve AYEM modellerinin birlikte karşılaştırılmasına karar verilmiştir. Ayrıca AYEM analizi ile gerçekleştirilen

alıřmalarda DFA modellerinin de test edildikleri grlmřtr (Alamer, 2022; Prokofieva vd., 2023; Swami vd., 2023).

alıřmada AYEM analizi yrtlrken Zyl ve Koolester'in (2022) nerdikleri adımlar takip edilmiřtir. İlk adım kuramsal alt yapıya uygun bir řekilde modelin ve uygun rneklem sayısının belirlendiėi planlama ařamasıdır. ABİYAT-R iin 3 faktrl ikinci dzey DFA ve 3 faktrl ikinci dzey AYEM modelleri test edilmek zere belirlenmiřtir. alıřmada modellerin test edilmesi iin gerekli olan minimum rneklem byklė karřılanmıř (Soper, 2024) ve analizler 244 kiřinin verileri zerinden yrtlmřtr. İkinci adım veri hazırlıėının saėlanması, analizin yrtlmesi iin uygun yazılım programının ve uygun kestirim ve dndrme metodlarının seilmesiştir. Veriler, faktr analizi iin varsayımların kontrol edildiėi EGA ncesinde hazır hale getirilmiřtir. Varsayımların kontrol sırasında ok deėiřkenli normallik varsayımının karřılanamadıėı grlmřtr (*Mardia arpıklık* =247.22, $p<.001$; *Mardia Basıklık* =2.54, $p<.001$ ve Energy Testi, $E=2.43$, $p<.001$). ok deėiřkenli normallik varsayımlarının karřılanamadıėı durumlarda MLR (Maximum Likelihood with Robust Standard Errors) ve WLSMV (Weighted Least Squares Mean and Variance Adjusted) kestirim yntemlerinin kullanılması nerilmektedir (Brown, 2015; Muthn ve Muthn, 2009).

Kestirim yntemlerinden MLR' nin srekli (Finney vd., 2016), WLSMV' nin ise kategorik deėiřkenler iin uygun teknikler oldukları belirtilmektedir (Bandalos ve Webb, 2005). alıřmada, ok deėiřkenli normallik varsayımının saėlanamaması ve veri setinin srekli deėiřkenlerden oluřması sebebiyle MLR kestirim ynteminin kullanılması uygun grlmřtr. AYEM ve DFA analizleri iin Mplus v 8.6 istatistik programından yararlanılmıřtır (Muthn ve Muthn, 2021). AYEM ve DFA modellerinin kod dizilimleri De Beer ve Van Zyl'in (2019) geliřtirdikleri kod reticisi ile oluřturulmuřtur. Mplus analiz programı varsayılan olarak Oblique dndrme tekniėini kullanmaktadır. Oblique dndrmesi, faktrlerin birbirleriyle iliřkili olduėu varsayımına dayanır (Prokofieva vd., 2023). ABİYAT-R testinde yer alan faktrlerin kuramsal olarak birbirleriyle iliřkili olmaları sebebiyle alıřmada Oblique tekniėinin kullanılması uygun grlmřtr. nc ve son adımda modeller iin uyum indekslerine karar verilmesi, model uyum indekslerinin raporlanması ve modellerin karřılařtırılarak en uygun modelin benimsenmesi yer almaktadır.

Çalışmalarda model uyum indeksleri olarak sıklıkla raporlanan ve modelleri kıyaslamada kullanılan kriterler incelenmiştir. ÇABİYAT-R'nin yapı geçerliği için uyum indeksleri olarak; χ^2/sd , CFI, TLI, RMSEA, SRMR AIC ve BIC değerlerinin raporlanmasına karar verilmiştir (Alamer, 2022; Prokofieva vd., 2023; Swami vd., 2023). Model uyum indeksleri değerlendirildiğinde iki modelin de ÇABİYAT-R testine yüksek uyum sağladıkları görülmüştür.

3.5.2.1.3 Yakınsama ve İraksama Geçerlikleri

Yapı geçerliğinin sağlanmasında yakınsama ve ıraksama geçerliklerinin değerlendirilmesi ilk olarak Campell ve Fiske (1959) tarafından önerilmiştir. Araştırmacılar söz konusu geçerlikleri; Çoklu Özellik- Çoklu Yöntem Matrisi (Multitrait-Multimethod Matrix- MTMM) yöntemi ile değerlendirmişlerdir (Campell ve Fiske, 1959). Yakınsama geçerliği bir yapının farklı şekillerde ölçülmesi sonucunda benzer sonuçlar elde edilip edilmediğini; ıraksama geçerliği ise ölçülen yapının kendinden farklı diğer yapılardan ayrışıp ayrışmadığının incelenmesidir (Boeteng vd., 2018). Fornell ve Larcker (1981) iki kriterin değerlendirilmesinde ortalama açıklanan varyans (AVE) ve birleşik güvenirlik (CR) katsayılarının hesaplandığı farklı bir yaklaşım önermişlerdir. AVE, ölçüm hataları kapsamında hesaplanan varyans miktarını belirtmektedir. AVE, bir faktör altında toplanan maddelerin faktör yüklerinin kareleri alınarak ortalamalarının hesaplanması ile elde edilmektedir. CR ise bir yapının genel güvenirliği ile ilgilidir ve faktör yükleri ile hata varyansları üzerinden türetilen bir formülle hesaplanmaktadır. Araştırmacılar ayrıca yakınsak ve ıraksak geçerlik kapsamında MSV (Maximum Shared Variance- En Yüksek Ortak Varyans) kriterinin de karşılanması gerektiğini önermişlerdir. Bu kriter gereği bir faktör kendi maddeleriyle olan ilişkisi kadar başka bir faktörle yüksek ilişki barındırmamalıdır. Diğer bir ifade ile faktöre ilişkin AVE katsayısı MSV katsayısından büyük olmalıdır. Çalışma kapsamında anlaşılır ve daha yorumlanabilir olmalarından dolayı AVE, CR ve MSV katsayılarının hesaplanmaları uygun görülmüştür (Fornell ve Larcker, 1981). AVE kapsamında hesaplanan değerler faktörler arası korelasyon katsayıları ile karşılaştırılmıştır. Korelasyon katsayıları için Pearson Momentler Çarpımı Korelasyon Katsayısı kullanılmıştır. AVE, CR ve MSV değerleri ikinci düzey üç faktörlü DFA ve ikinci düzey üç faktörlü AYEM modellerinin her ikisi için de ayrı ayrı hesaplanarak raporlaştırılmıştır.

3.5.2.2 Ölçüt geçerliği

Bir ölçeğin belirlenen kriterler bağlamında deneysel olarak incelenmesi ölçüt geçerliğini kapsamaktadır (DeVellis ve Thorpe, 2022; Nunnally ve Bernstein, 1994). Çalışmada ölçüt geçerliği kapsamında uyum geçerliği (convergent validity) ve bilinen grup geçerliği (known -group validity) incelenmiştir (Netemeyer vd., 2003). Uyum geçerliği kapsamında öğrencilerin fen bilimlerine yönelik alan bilgisi, bilinen grup geçerliği kapsamında ise öğrencilerin sınıf düzeyleri ele alınmıştır.

3.5.2.3 Uyum geçerliği

Uyum geçerliği, geliştirilen ölçeğin teorik olarak benzer bir yapıyı ölçen farklı bir ölçekle ilişkili olup olmadığının göstergesidir (Clark ve Watson, 2019). Çalışmada, uyum geçerliğini sağlamak amacıyla ÇABİYAT-R'nin alt test ve toplam test puanları ile SAGES 2'nin fen bilgisi alt test puanları, Pearson Momentler Çarpımı Korelasyon Katsayısı üzerinden karşılaştırılmışlardır.

3.5.2.3 Bilinen grup geçerliği

Bilinen gruplar geçerliği, geliştirilen bir ölçeğin gruplar arasındaki puanların anlamlılığını ayırt etmede kullanılmaktadır (DeVellis ve Thorpe, 2022; Netemeyer vd., 2003). Bilinen gruplar geçerliğini sağlamak amacıyla, ÇABİYAT-R alt test ve toplam test puanları öğrencilerin sınıf düzeyleri bağlamında değerlendirilmiş ve bu kapsamda 244 öğrenci üzerinden bağımsız gruplar tek faktörlü ANOVA analizleri yapılmıştır. Analiz öncesinde normallik varsayımları kontrol edilmiş ve varyansların homojenliği test edilmiştir. Normallik varsayımları için öğrencilerin sınıf düzeylerine göre çarpıklık ve basıklık katsayıları yorumlanmıştır. Alt testler ve toplam test için sınıf düzeylerine göre elde edilen çarpıklık değerleri -.065 ile .863 arasında, basıklık değerleri ise -.968 ile .608 arasında değişim göstermektedir. Huck (2008), çarpıklık ve basıklık katsayılarının -1 ile +1 arasında değişmesi gerektiğini belirtmektedir. Elde edilen sonuçlara göre verilerin sınıf düzeyinde normal dağılım sergilediği söylenebilir. Varyansların homojenliği için Levene's Testi yapılmıştır. Levene's Testi sonuçlarına göre hipotez geliştirme alt testinde varyansların homojen dağıldığı ($p>.05$); kanıt değerlendirme, deney tasarlama alt testleri ile test toplam puanında ise varyans homojenliğinin sağlanamadığı görülmüştür ($p<.05$). Varyans homojenliğinin sağlandığı durumlar için Scheffe, homojenliğin sağlanamadığı durumlar için

Tamhane's T2 izleme testleri yapılmıştır (Huck, 2008). Seçilen iki izleme testi de tutucu sonuçlar vermiştir (Hair vd., 2019; Tamhane, 1979).

3.5.2.4 Maddelerin cinsiyet bağlamında yanlılık analizi

Ölçek maddelerinin; cinsiyet, etnik köken, din ve ekonomik durum açısından bir grubu olumsuz etkilememesi veya potansiyel olarak bir grup lehine ön yargı oluşturmaması gerekir (Weiner vd., 2012). Bu bağlamda madde yanlılığı analizlerinin yapılması önerilmektedir (Mellenbergh, 1989). ÇABİYAT-R'nin madde yanlılığı cinsiyet değişkeni üzerinden ölçme değişmezliği analizleri kapsamında gerçekleştirilmiştir. Ölçme değişmezliği bir yapının gruplar arasında psikometrik özellikleri bağlamında eşdeğerliğini ifade etmektedir (Putnick ve Bornstein, 2016). Diğer bir ifade ile ölçme değişmezliği, bir aracın farklı grup bireyleri arasında aynı şekilde yorumlanıp yorumlanamayacağının göstergesidir (Holden vd., 2020). Ölçme değişmezliğinin, madde tepki kuramı (IRT) ve yapısal eşitlik modellemesi (SEM) olarak iki ayrı yaklaşımla test edilebildiği görülmektedir (Raju vd., 2002; Widaman, 2014). Mevcut çalışmada ölçme değişmezliği SEM yaklaşımı benimsenerek DFA kapsamında test edilmiştir. Bu kapsamda ölçme değişmezliği için öncül analizlerde varsayımlar karşılanmıştır.

Yapısal eşitlik modellemesine göre ölçme değişmezliğinin hiyerarşik yapıda ve temel olarak dört adımda gerçekleştirildiği görülmektedir (Putnick ve Bornstein, 2016; Widaman ve Reise, 1997). Her adımda modele bir kısıtlama getirilerek ölçme değişmezliğinin sağlandığı noktada bir sonraki adıma geçilmektedir. İlk adım olan *yapısal değişmezlik* (Configural Invariance), ölçme değişmezliğinin en liberal basamağı olarak ifade edilmekte ve faktör yapısının gruplar içerisindeki değişmezliği test edilmektedir. Bu adımda modelde herhangi bir kısıtlama bulunmamaktadır (Millsap, 1993). Eğer yapısal değişmezlik sağlanıyorsa bu kez faktör yüklerinin eşdeğerliklerinin test edildiği ikinci adıma, *metrik değişmezliğine* (Metric Invariance) geçilir. Metrik değişmezlik iki grup için eşdeğer olacak şekilde faktör yüklerinin sınırlandırılmasıyla test edilmektedir (Widaman ve Reise, 1997). Test sonucunda model uyum indeksleri bir önceki adımda test edilen yapısal değişmezlik modeli uyum indekslerine kıyasla büyük ölçüde daha kötü değilse gruplar arasında metrik değişmezliğin sağlandığı anlamına gelmektedir (Eid ve Kutscher, 2014).

Diğer bir ifade ile bu durum maddelerin faktörlere iki grup için de aynı şekilde yüklendiklerini göstermektedir.

Metrik değişmezlik sağlandıktan sonra iki grup için madde ortalamalarının kesişim noktalarının sabitlendiği ek bir kısıtlama getirilerek *skaler değişmezlik* test edilmektedir (Scalar Invariance). Eğer model metrik değişmezlik modeline göre kötü uyum gösteriyorsa en az bir maddenin gruplar arasında farklılık yarattığı ve skaler değişmezliğin sağlanamadığı yorumu getirilmektedir (Putnick ve Bornstein, 2016). Yeni model bir önceki model olan metrik değişmezlik modeline kıyasla büyük ölçüde kötü değilse bu kez madde kesişmelerinin gruplar arasında sınırlandırılmasının modeli kötü etkilemediği ve skaler değişmezliğin sağlandığı anlamına gelmektedir (Meredith, 1993). Skaler değişmezlik de sağlanıyorsa en tutucu adım olan *hata varyans değişmezliği* (Residual Invariance) ya da farklı bir terimle *katı* değişmezliğine geçilmektedir (Strict Invariance). Bu adımda önceki modellerde sabit bırakılan sınırlamalara ek olarak gruplar arasındaki hata varyanslarının sınırlandırılması test edilmektedir. Yeni model skaler değişmezlik sonucunda elde edilen model uyumu ile karşılaştırılır. Model uyumu daha kötü ise gruplar arasında en az bir madde için hata varyansının farklı olduğu, büyük ölçüde kötü uyum yoksa hata varyans sınırlandırılmasının modeli etkilemediği ve katı değişmezliğinin sağlandığı şeklinde yorumlanmaktadır (Putnick ve Bornstein, 2016). ÇABİYAT-R'nin ölçme değişmezliği literatürde belirtilen şekilde adım adım gerçekleştirilmiş ve son olarak ÇABİYAT-R'de yer alan gizil faktörlerin ortalamaları üzerinden erkek ve kız öğrenciler arasında anlamlı bir fark olup olmadığı kıyaslanmıştır.

3.5.3 Güvenirlik analizleri

Çalışmanın güvenilirlik analizleri kapsamında iç tutarlılık ve kodlayıcılar arası güvenilirlik analizleri yapılmıştır. İç tutarlılık güvenirliliği için Cronbach Alpha katsayısı hesaplanmıştır. Cronbach Alpha katsayısı ile madde silindiğinde Cronbach Alpha değerleri ve düzeltilmiş madde toplam korelasyon katsayıları da incelenmiştir.

ÇABİYAT-R'ın uygulanmasında öğrencilerin her bir madde için verdikleri doğru yanıtlar araştırmacı tarafından kontrol edilerek forma işaretlenmiştir. Sonrasında her bir madde için öğrencilerin akıcılık puanları hesaplanmıştır. Maddelerin kodlanmasında ya da puanlanmasında kodlayıcılar arası güvenilirlik analizlerinin yapılması önerilmektedir (Cohen, 1960; Shrout ve Fleiss, 1979). Uygulama

verilerinden rastgele seçilen 99 kişinin akıcılık puanları tekrar hesaplanmak üzere bir başka uzmana verilmiştir. Araştırmacı ve uzman arasındaki uyum Sınıf İçi Korelasyon Katsayısı (ICC) hesaplanarak yorumlanmıştır.



4 BULGULAR

Bu bölümde ÇABIYAT-R'nin madde analizi ve psikometrik özelliklerine ilişkin bulgulara yer verilmiştir. Öncelikle madde analizleri bulgularına yer verilmiş, sonra açıklayıcı yapısal eşitlik modellemesi yapılarak testin psikometrik özellikleri incelenmiştir.

4.1 Madde Analizlerine İlişkin Bulgular

Madde analizleri kapsamında betimsel istatistik, madde ayırt edicilik, regresyon analizleri, maddeler arası korelasyon ve madde güvenirlik analizleri gerçekleştirilmiştir. 244 kişi üzerinden 15 madde ile yapılan analizlere ilişkin betimsel istatistik bulgularına Tablo 4.1' de yer verilmiştir.

Tablo 4.1

Maddelere ilişkin betimsel istatistik bulguları

Maddeler	Bileşenler	En Düşük Puan	En Düşük Sıklık	En Yüksek puan	En Yüksek Sıklık	$\bar{X}$	ss	Varyans	Çarpıklık	Basıklık
Çalar Saat	KD1	0	32	5,0	21	2,471	1,50	2,25	-,091	-,973
Bıçak	KD2	0	37	5,0	9	2,078	1,37	1,89	,154	-,825
Mıknatıs	KD3	0	29	6,0	8	2,512	1,57	2,47	,176	-,645
Şeker	KD4	0	52	7,0	2	2,164	1,72	2,98	,555	-,337
Yuvarlanma	KD5	0	35	5,0	23	2,225	1,52	2,34	,220	-,979
Terleme	KD6	0	77	7,0	1	1,910	1,70	2,90	,473	-,692
Yağmur	KD7	0	80	6,0	2	1,615	1,51	2,29	,638	-,469
Balon	KD8	0	59	9,0	1	2,410	2,00	4,02	,669	-,001
Gölge	KD9	0	104	7,0	3	1,615	1,82	3,33	,960	,038
Hamster	DT1	0	18	5,0	6	2,045	1,14	1,30	,280	-,215
Kum havuzu	DT2	0	58	4,0	5	1,123	0,92	0,85	,945	,867
Tünel	DT3	0	34	5,0	1	1,590	1,10	1,21	,588	-,168
Sinek	HG1	0	19	5,0	2	1,971	1,05	1,12	,183	-,180
Suluk	HG2	0	10	6,0	3	2,500	1,26	1,60	,372	-,144
Gemi	HG3	0	14	6,0	3	2,352	1,31	1,73	,395	-,186

KD: Kanıt değerlendirme, DT: Deney tasarlama, HG: Hipotez geliştirme

Betimsel analiz bulgularına göre çarpıklık ve basıklık katsayılarının -1 ile +1 arasında değiştiği ve verilerin normal dağılım sargıladıkları görülmüştür (Huck, 2008). En yüksek üretilen doğru sayısı 4 (Kum Havuzu-DT2) ile 9 (Balon-KD8), maddelere hiçbir doğru yanıt üretemeyen öğrenci sayısı 10 (Suluk-HG2) ile 104 (Gölge- KD9), maddelere en yüksek yanıt üreten öğrenci sayısı ise 1 (Tünel-DT3) ile 23 (Yuvarlanma-KD5) arasında değişmektedir. Kanıt değerlendirme görevi için en yüksek yanıt üretilen diğer maddeler Çalar Saat (KD1- 21), Bıçak (KD2- 9) ve Mıknatıs (KD3-8) şeklinde sıralanmaktadır. Kanıt değerlendirme maddelerine

verilen doğru yanıtların frekansları genel olarak incelendiğinde; Çalar Saat (KD1), Bıçak (KD2), Mıknatıs (KD3) ve Yuvarlanma (KD5) maddelerinin kolay; Terleme (KD6), Yağmur (KD7) ve Gölge (KD9) maddelerinin ise öğrenciler için görece zor maddeler oldukları söylenebilir. Kanıt değerlendirme maddelerine hiç yanıt üretemeyen öğrenci sayısının deney tasarlama ve hipotez geliştirme maddelerine kıyasla daha fazla olduğu görülmektedir. Maddelerin açıkladıkları varyanslar genel olarak incelendiğinde kanıt değerlendirme maddelerinin, deney tasarlama ve hipotez geliştirme maddelerine kıyasla daha yüksek varyansa sahip oldukları görülmüştür. En yüksek varyans 4.02 ile Balon (KD8); en düşük varyans ise 0,85 Kum Havuzu (DT2) maddesine aittir. Kanıt değerlendirme görevine ilişkin en düşük varyans Bıçak (KD2) maddesine aittir. Balon (KD8) maddesinden sonra sırasıyla en yüksek varyanslar: Gölge (KD9) 3,33; Şeker (KD4) 2,98 ve Terleme (KD6) 2,90 maddelerine aittir.

Madde analizlerinin diğer adımında sınıf düzeyi bağlamında madde ayırt edicilikleri kontrol edilmiştir. Kanıt değerlendirme görevine yönelik geliştirilmiş 9 maddenin sınıf düzeylerine göre madde ayırt ediciliklerinin kontrolleri için tek faktörlü ANOVA analizleri yapılmıştır. Analize geçmeden önce Levene's varyans homojenlik testi kontrol edilmiştir. Varyans homojenlik testleri sonucunda Bıçak (KD2) ve Yuvarlanma (KD5) maddelerinde varyans homojenliğinin sağlandığı ($p>0.05$), diğer maddelerde ise homojenliğin sağlanamadığı görülmüştür ($p<0.05$). Bu doğrultuda gruplar arasında farkın kaynağını ortaya çıkarmak amacıyla Bıçak ve Yuvarlanma için Scheffe, varyans homojenliğin sağlanamadığı maddelerde ise Tamhane's T2 izleme testlerinin yapılması uygun görülmüştür. Grupların kanıt değerlendirme madde puanlarına ilişkin ANOVA analizleri sonuçları Tablo 4.2' de yer almaktadır.

Tablo 4.2 incelendiğinde Çalar Saat (KD1), Yuvarlanma (KD5) ve Gölge (KD9) maddeleri dışında kalan diğer tüm maddelerde gruplar arası anlamlı farklara rastlanmaktadır. Çalar Saat (KD1) maddesinde anasınıfı ortalaması ($\bar{X} = 2.08$) birinci sınıf ortalamasından ($\bar{X} = 2.0$) büyüktür. Yuvarlanma (KD5) ve Gölge (KD9) maddelerinde de anasınıfı ortalamaları birinci sınıfların ortalamalarından yüksek bulunmuştur ancak ortalamalar arasında istatistiksel olarak anlamlı bir farklılık sağlanamamıştır ($p>0.05$). Maddelerin etki büyüklük değerlerinin .104 ile .233 arasında değiştikleri görülmektedir. Bu bağlamda Bıçak (KD2) ve Gölge (KD9)

maddelerinin orta, diğer maddelerin ise geniş etki büyüklüğüne sahip oldukları söylenebilir (Cohen, 1988). Madde ayırt edicilik analizleri genel olarak değerlendirildiğinde Çalar Saat (KD1), Yuvarlanma (KD5) ve Gölge (KD9) maddeleri dışındaki diğer tüm maddelerin sınıf düzeyleri bağlamında ayırt ediciliklerinin iyi oldukları söylenebilir. Madde ayırt edicilikleri ile etki büyüklük değerleri göz önünde bulundurulduğunda, Şeker (KD4) ve Balon (KD8) maddelerinin pratik ve teorik bağlamda iyi çalıştıkları söylenebilir.

Tablo 4.2

Sınıf düzeylerine göre kanıt değerlendirme maddelerinin ayırt edicilikleri

Madde	Varyans Kaynağı	Kareler Toplamı	sd	KO	F	p.	İzleme Testi	Fark	p.	η^2
Çalar Saat KD1	G. Arası	99,856	2	49,928	26,802	.001	Tamhane	0-1	.972	.182
	Grup İçi	448,943	241	1,863				0-2	.001*	
	Toplam	548,799	243					1-2	.001*	
Bıçak KD2	G. Arası	58,588	2	29,294	17,521	.001	Scheffe	0-1	.005*	.127
	Grup İçi	402,933	241	1,672				0-2	.001*	
	Toplam	461,520	243					1-2	.027*	
Mıknatıs KD3	G. Arası	98,291	2	49,145	23,562	.001	Tamhane	0-1	.018*	.164
	Grup İçi	502,672	241	2,086				0-2	.001*	
	Toplam	600,963	243					1-2	.001*	
Şeker KD4	G. Arası	169,252	2	84,626	36,669	.001	Tamhane	0-1	.001*	.233
	Grup İçi	556,190	241	2,308				0-2	.001*	
	Toplam	725,443	243					1-2	.001*	
Yuvarlanma KD5	G. Arası	104,128	2	52,064	27,014	.001	Scheffe	0-1	.123	.183
	Grup İçi	464,474	241	1,927				0-2	.001*	
	Toplam	568,602	243					1-2	.001*	
Terleme KD6	G. Arası	104,635	2	52,318	20,966	.001	Tamhane	0-1	.002*	.148
	Grup İçi	601,381	241	2,495				0-2	.001*	
	Toplam	706,016	243					1-2	.010*	
Yağmur KD7	G. Arası	94,010	2	47,005	24,426	.001	Tamhane	0-1	.001*	.169
	Grup İçi	463,777	241	1,924				0-2	.001*	
	Toplam	557,787	243					1-2	.025*	
Balon KD8	G. Arası	225,521	2	112,76	36,162	.001	Tamhane	0-1	.001*	.231
	Grup İçi	751,495	241	3,118				0-2	.001*	
	Toplam	977,016	243					1-2	.043*	
Gölge KD9	G. Arası	84,356	2	42,178	14,012	.001	Tamhane	0-1	.074	.104
	Grup İçi	725,431	241	3,010				0-2	.001*	
	Toplam	809,787	243					1-2	.009*	

KO: Ortalamanın karesi. η^2 :Kısmi eta-kare

Madde analizlerinin diğer bir adımı basit doğrusal regresyon analizlerini kapsamaktadır. Öğrencilerin kanıt değerlendirme, deney tasarlama ve hipotez geliştirme alt testleri için toplam 15 madde üzerinden akıcılık puanları

hesaplanmıştır. Her bir öğrencinin akıcılık puanları toplanarak toplam test puanları elde edilmiştir. Sonrasında her bir kanıt değerlendirme maddesinin toplam test puanını yordama düzeylerine bakılmıştır. Kanıt değerlendirme maddelerine ilişkin basit doğrusal regresyon analizi sonuçlarına Tablo 4. 3'te yer verilmiştir.

Tablo 4.3

Basit doğrusal regresyon analizi sonuçları

Model	Maddeler	R ²	R ² Düzeltilmiş	ANOVA		Katsayılar		
				F	p	Sabit	Beta	Standardize β
1	KD4-Şeker	.582	.580	336,719	.000	16,963	6,294	.763*
2	KD8-Balon	.547	.546	292,794	.000	17,905	5,261	.740*
3	KD6-Terleme	.542	.540	286,543	.000	18,821	6,158	.736*
4	KD5-Yuvarlanma	.541	.539	285,638	.000	15,322	6,857	.736*
5	KD7-Yağmur	.529	.527	272,189	.000	19,527	6,846	.728*
6	KD3-Mıknatıs	.523	.521	265,162	.000	14,114	6,555	.723*
7	KD2-Bıçak	.519	.517	261,595	.000	15,090	7,456	.721*
8	KD9-Gölge	.460	.458	206,067	.000	22,030	5,296	.678*
9	KD1-Çalar Saat	.316	.314	111,970	.000	17,396	5,335	.562*

KD: Kanıt Değerlendirme

Tablo 4.3'te maddelerin varyans değerleri (F) en yüksekten en düşüğe doğru sıralanmaktadır. Tablo incelendiğinde kanıt değerlendirme maddelerinin toplam test puanını $R^2_{\text{Düzeltilmiş}} = .314 - .580$ değerleri arasında açıkladıkları görülmüştür ($p < .001$). Diğer bir ifadeyle kanıt değerlendirme maddeleri toplam test puanı üzerinde %32 ile %58 arasında değişim yaratmışlardır. Söz konusu yüzdelerinin orta ve yeterli derecede kabul edildikleri söylenebilir (Chin, 1998). Toplam test puanını yordayan en yüksek maddenin Şeker ($R^2_{\text{Düzeltilmiş}} = .580$, $F(1, 243) = 336,719$, $p < .001$); en düşük maddenin ise Çalar Saat ($R^2_{\text{Düzeltilmiş}} = .316$, $F(1, 243) = 111,970$, $p < .001$) olduğu görülmektedir. Madde analizlerinde F değerleri yüksekten düşüğe doğru sıralanarak madde seçimi için kanıt sunabilmektedir (Desboulets, 2018; Tezbaşaran, 2008). Regresyon analizi sonuçları genel olarak değerlendirildiğinde Şeker ve Balon maddelerinin oldukça iyi çalıştıkları, Çalar Saat ve Gölge maddelerinin ise yeterince iyi çalışmadıkları söylenebilir. Terleme ve Yuvarlanma maddelerinin açıkladıkları

varyans değerlerinin birbirlerine oldukça yakın oldukları ve iki maddeye ilişkin standardize edilmiş Beta katsayılarında da benzer sonuçlar alındığı görülmüştür.

Madde analizlerinde kanıt değerlendirme maddeleri için erkek veya kız öğrenciler lehine anlamlı bir fark olup olmadığı bağımsız örneklem t testi analizleri ile kontrol edilmiştir. Bağımsız örneklem t testi analizlerinde çok sayıda karşılaştırma yapılması durumunda Tip 1 hatanın görülme olasılığı artabilmektedir. Tip 1 hatanın önüne geçebilmek amacıyla Bonferroni düzeltilmesi önerilmektedir (Dunn, 1961). Bonferroni düzeltilmesi, analizde benimsenen anlamlılık düzeyinin yapılan analiz sayısına bölünmesiyle hesaplanmaktadır. Çalışmada anlamlılık düzeyi $p=0.006$ (0.05/9) olarak benimsenmiştir. Analiz sonuçlarına Tablo 4.4'te yer verilmiştir.

Tablo 4.4

Cinsiyete göre kanıt değerlendirme madde ortalamaları ve anlamlılık değerleri

Madde	Cinsiyet	N	$\bar{X}$	SS	$\Delta \bar{X}$	t	p
Çalar Saat	E	142	2,599	1,4735	,3045	1,566	,119
KD1	K	102	2,294	1,5324			
Bıçak	E	142	2,148	1,4289	,1675	,936	,350
KD2	K	102	1,980	1,3048			
Mıknatıs	E	142	2,669	1,6231	,3749	1,846	,066
KD3	K	102	2,294	1,4798			
Şeker	E	142	2,261	1,7122	,2312	1,031	,304
KD4	K	102	2,029	1,7489			
Yuvarlanma	E	142	2,387	1,5703	,3873	1,962	,051
KD5	K	102	2,000	1,4488			
Terleme	E	142	1,937	1,7227	,0641	,289	,773
KD6	K	102	1,873	1,6867			
Yağmur	E	142	1,556	1,4017	-,1397	-,710	,478
KD7	K	102	1,696	1,6638			
Balon	E	142	2,472	1,9595	,1483	,569	,570
KD8	K	102	2,324	2,0737			
Gölge	E	142	1,746	1,8506	,3151	1,332	,184
KD9	K	102	1,431	1,7828			

KD: Kanıt Değerlendirme. $\Delta \bar{X}$: Ortalamalar arası fark.

Tablo 4.4 incelendiğinde Yağmur (KD7) maddesi hariç tüm maddelerde erkek öğrencilerin kız öğrencilerden daha yüksek puan aldıkları fakat maddelerin tamamında ortalamalar arası puan farklılıklarının anlamlı olmadığı görülmüştür ($p>.006$). Bu bağlamda kanıt değerlendirme maddelerinin tamamında potansiyel olarak bir cinsiyet etkisinin olmadığı söylenebilir.

Tablo 4.5*Maddeler arası korelasyon katsayı sonuçları*

	KD1	KD2	KD3	KD4	KD5	KD6	KD7	KD8	KD9	DT1	DT2	DT3	HG1	HG2	HG3
KD1	-	,334**	,495**	,411**	,441**	,388**	,369**	,318**	,354**	,196**	,121	,254**	,151*	,176**	,296**
KD2		-	,530**	,570**	,544**	,448**	,507**	,519**	,455**	,304**	,283**	,403**	,354**	,366**	,341**
KD3			-	,520**	,523**	,487**	,446**	,528**	,466**	,276**	,250**	,319**	,258**	,329**	,423**
KD4				-	,567**	,585**	,512**	,506**	,469**	,330**	,332**	,364**	,322**	,361**	,405**
KD5					-	,614**	,528**	,490**	,466**	,218**	,236**	,372**	,273**	,283**	,410**
KD6						-	,552**	,494**	,461**	,281**	,234**	,337**	,331**	,369**	,364**
KD7							-	,627**	,494**	,241**	,233**	,250**	,406**	,340**	,367**
KD8								-	,531**	,226**	,248**	,320**	,323**	,340**	,397**
KD9									-	,297**	,269**	,355**	,252**	,208**	,242**
DT1										-	,556**	,542**	,318**	,294**	,353**
DT2											-	,546**	,213**	,196**	,231**
DT3												-	,353**	,327**	,310**
HG1													-	,559**	,550**
HG2														-	,502**
HG3															-

* p<.05 ** p<.01 KD1: Çalar saat, KD2: Bıçak, KD3: Mıknatıs, KD4: Şeker, KD5: Yuvarlanma, KD6: Terleme, KD7: Yağmur, KD8: Balon, KD9: Gölge, DT1: Hamster, DT2: Kum havuzu, DT3: Tünel, HG1: Sinek, HG2: Suluk, HG3: Gemi. KDT: Kanıt değerlendirme maddeleri toplamı. DTT: Deney tasarlama maddeleri toplamı. HGT: Hipotez geliştirme maddeleri toplamı. TT: Toplam test puan

Madde analizlerinin son aşaması maddeler arası korelasyon katsayıları ile güvenilirlik analizlerini barındırmaktadır. Maddeler arasındaki ilişkiye Pearson Momentler Çarpımı Korelasyon Katsayısı ile bakılmıştır. Korelasyon katsayılarına Tablo 4.5’te yer verilmiştir.

Çalar Saat (KD1) ile Kum Havuzu (DT2) arasındaki ilişki ($r=.121$, $p=>.05$) dışında kalan diğer tüm korelasyonlar anlamlıdır. Öte yandan Çalar Saat (KD1) ile Sinek (HG1) arasındaki ilişki ($r=.151$) $p=.05$ düzeyinde; diğer korelasyon katsayıları $p=.01$ düzeyinde istatistiksel olarak anlamlıdır. Maddeler arası korelasyon katsayılarının .121 ile .614 arasında değiştikleri görülmüştür. Korelasyon katsayılarına ek kanıt sunmak amacıyla düzeltilmiş madde toplam korelasyon değerleri ve güvenilirlik analizi sonuçlarına da Tablo 4.6’ da yer verilmiştir.

Tablo 4.6

Madde güvenilirlik analizi sonuçları

Madde	Düzeltilmiş Madde Toplam Korelasyonları	Madde silindiğinde Cronbach Alpha	Alt ölçek Cronbach Alpha
KD1-Çalar Saat	0,48	0,90	
KD2- Bıçak	0,67	0,89	
KD3- Mıknatıs	0,66	0,89	
KD4-Şeker	0,71	0,89	
KD5- Yuvarlanma	0,68	0,89	0,89
KD6- Terleme	0,67	0,89	
KD7- Yağmur	0,67	0,89	
KD8- Balon	0,67	0,89	
KD9- Gölge	0,60	0,89	
DT1- Hamster	0,45	0,90	
DT2- Kum havuzu	0,41	0,90	0,78
DT3- Tünel	0,53	0,90	
HG1-Sinek	0,50	0,90	
HG2- Suluk	0,49	0,90	0,77
HG3- Gemi	0,56	0,90	

KD: Kanıt değerlendirme. DT: Deney tasarlama. HG: Hipotez geliştirme

Güvenirlik analizlerinde ÇABIYAT-R testi Cronbach Alpha katsayısı $\alpha=.901$ olarak hesaplanmıştır. Alt ölçekler için Cronbach Alpha katsayıları kanıt değerlendirme $\alpha=.893$; deney tasarlama $\alpha=.779$ ve hipotez geliştirme $\alpha=.77$ olarak bulunmuştur. Toplam test ve alt ölçekler bağlamında elde edilen güvenilirlik katsayılarının tamamı

kabul edilebilir ($\alpha = .70$) düzeyin üzerindedir (Nunnally ve Bernstein, 1999). Madde silinmesi durumunda toplam test Cronbach Alpha katsayılarında olumlu bir değişikliğin olmayacağı görülmüştür. Düzeltilmiş madde toplam korelasyonlarının .30 ve üzerinde olması gerektiği önerilmektedir (Pallant, 2020). Tablo incelendiğinde, düzeltilmiş madde korelasyon katsayılarının tamamının iyi uyum gösterdikleri söylenebilir.

Kanıt değerlendirme maddelerinin seçimi

Madde analizleri genel olarak değerlendirildiğinde kanıt değerlendirme alt ölçeği için geliştirilen 9 maddeye ilişkin birtakım kararlar alınmıştır. Maddelerin değerlendirildiği kriterlere ve kararlara ilişkin özet Tablo 4.7’de yer verilmiştir.

Tablo 4.7

Kanıt değerlendirme maddelerinin seçimine ilişkin alınan kararlar

Maddeler	Değerlendirme Kriterleri						
	Betimsel Değerler	Maddeler Arası Ayırt Edicilik	Basit Regresyon Değerleri	Madde Korelasyonu	Cinsiyet Farklılığı	Madde Güvenirliliği	Karar
Çalar Saat KD1	×	×	×	×	✓	×	×
Bıçak KD2	×	✓	×	✓	✓	✓	×
Mıknatıs KD3	×	✓	×	✓	✓	✓	×
Şeker KD4	✓	✓	✓	✓	✓	✓	✓
Yuvarlanma KD5	×	×	✓	✓	✓	✓	×
Terleme KD6	✓	✓	✓	✓	✓	✓	✓
Yağmur KD7	×	✓	×	✓	✓	✓	×
Balon KD8	✓	✓	✓	✓	✓	✓	✓
Gölge KD9	×	×	×	✓	✓	×	×

✓: Kriteri karşılıyor; ×: Kriteri karşılamıyor

Tablo 4.7’de görüldüğü üzere kanıt değerlendirme becerisi için Şeker, Terleme ve Balon maddeleri seçilmiştir. Maddelerin seçimine ilişkin gerekçeler şu şekildedir:

- Betimsel analiz istatistik sonuçları incelendiğinde Yuvarlanma maddesinden 23 ve Çalar Saat maddesinden 21 öğrencinin en yüksek puana ulaştıkları görülmüştür. Diğer taraftan Gölge maddesi görece yüksek bir varyansa (3,33) sahip olmasına rağmen 104 öğrenci maddeye hiç yanıt üretememiştir. Maddelere verilen en yüksek ve en düşük sıklık oranları göz önünde bulundurulduğunda Yuvarlanma ve Çalar Saat maddelerinin oldukça kolay ve Gölge maddesinin ise oldukça zor bir madde olduğu düşünülmüştür. Sınıf düzeyinde yapılan madde ayırt edicilik analizleri incelendiğinde Çalar Saat, Yuvarlanma ve Gölge maddelerinin anasınıfı ve birinci sınıf öğrencileri arasında istatistiksel olarak anlamlı bir fark ortaya koyamadıkları görülmüştür. Diğer bir ifadeyle söz konusu maddeler grup farklılıklarını ayırt etmede yetersiz kalmışlardır. Basit Doğrusal Regresyon analizi sonuçları incelendiğinde Çalar Saat ($R^2_{\text{Düzeltilmiş}}=.316$, $F_{(1, 242)}=111,970$ $p<.001$) ve Gölge ($R^2_{\text{Düzeltilmiş}}=.458$, $F_{(1, 242)}=206,067$ $p<.001$) maddelerinin diğer maddelere kıyasla düşük varyans açıkladıkları görülmüştür. Ayrıca maddeler arası korelasyon bulgularında, Çalar Saat maddesinin deney tasarlama alt ölçeğinde yer alan Kum Havuzu maddesi ile düşük ve istatistiksel olarak anlamlı olmayan korelasyona sahip olduğu sonucuna ulaşılmıştır ($r=.12$ $p>.05$). Elde edilen bulgular doğrultusunda Çalar Saat, Yuvarlanma ve Gölge maddelerinin testten çıkarılmasına karar verilmiştir.
- Betimsel bulgular kapsamında Balon (4,02), Şeker (2,98) ve Terleme (2,90) maddelerinin Bıçak (1,89), Mıknatıs (2,47) ve Yağmur (2,29) maddelerine kıyasla görece daha yüksek varyans açıkladıkları görülmüştür. Elde edilen sonuçlar toplam test puanının yordandığı regresyon analiz sonuçları ile de benzerlik göstermektedir. Basit doğrusal regresyon analizi sonuçlarına göre Şeker ($R^2_{\text{Düzeltilmiş}}=.58$, $F_{(1, 242)}=336, 719$ $p<.001$); Balon ($R^2_{\text{Düzeltilmiş}}=.546$, $F_{(1, 242)}=292,794$ $p<.001$) ve Terleme ($R^2_{\text{Düzeltilmiş}}=.542$, $F_{(1, 242)}=286, 543$ $p<.001$) toplam test puanını yordayan en yüksek ilk üç değişkene işaret etmektedir. Terleme ve Yağmur maddelerine ait F varyans değişimlerinin birbirlerine yakın olduğu fark edilmiştir. İki madde arasındaki farklılığın anlamlı olup olmadığını test etmek için $R^2_{\text{Düzeltilmiş}}=.542_{\text{Terleme}}$ ve $R^2_{\text{Düzeltilmiş}}=.527_{\text{Yağmur}}$ katsayıları üzerinden model karşılaştırması yapılmıştır. Elde edilen sonuçlarına göre, $F_{(1, 240)}=6,81$, $p<.01$, iki model arasındaki fark istatistiksel olarak anlamlı bulunmuştur. Yani, Terleme maddesi toplam test puanını Yağmur maddesinden

daha iyi açıklamaktadır. Sınıf düzeyinde yapılan ayırt edicilik analizleri incelendiğinde Şeker, Terleme ve Balon maddelerinin etki büyüklükleri sırasıyla $\eta^2=.233$; $\eta^2=.148$ ve $\eta^2=.231$ olarak hesaplanmıştır. Elde edilen değerlerin geniş etkiye sahip oldukları söylenebilir ($\eta^2>.14$, Cohen, 1988). Diğer taraftan Bıçak maddesi, gruplar arasında istatistiksel olarak anlamlı bir fark yaratmasına rağmen etki büyüklüğünün orta düzeyde kaldığı görülmüştür ($\eta^2=.127$). Maddelere yönelik cinsiyet bağlamında yapılan analizler sonucunda herhangi bir grup lehine anlamlı bir fark elde edilememiştir. Yalnızca madde seçimine yönelik olarak kabaca bir yorumla maddelerin potansiyel bir cinsiyet etkisi yaratmadıkları söylenebilir. Betimsel istatistik, ayırt edicilik ve regresyon analiz sonuçları bağlamında Bıçak, Mıknatıs ve Yağmur maddelerinin analizden çıkarılmasına karar verilmiştir. Madde analiz bulguları genel olarak değerlendirildiğinde istatistiksel, teorik ve pratik faydaların gözetilmesi adına kanıt değerlendirme alt ölçeği için Şeker, Terleme ve Balon maddelerinin analize dahil edilmesine karar verilmiştir. Bu bağlamda ÇABİYAT-R testinin psikometrik özellikleri her bir alt ölçeğe ait 3'er ve toplamda 9 madde üzerinden yürütülmüştür.

4.2 ÇABİYAT-R' nin Geçerliği

Bu bölümde ÇABİYAT-R testinin yapı geçerliği, ölçüt geçerliği ve madde yanlılığı analiz bulgularına yer verilmiştir.

4.2.1 ÇABİYAT-R' nin yapı geçerliği

Bu bölümde ÇABİYAT-R' nin yapı geçerliği için açımlayıcı grafik analizi, açımlayıcı yapısal eşitlik modellemesi, yakınsak ve ıraksak geçerlik analizleri yapılarak bulgularına yer verilmiştir.

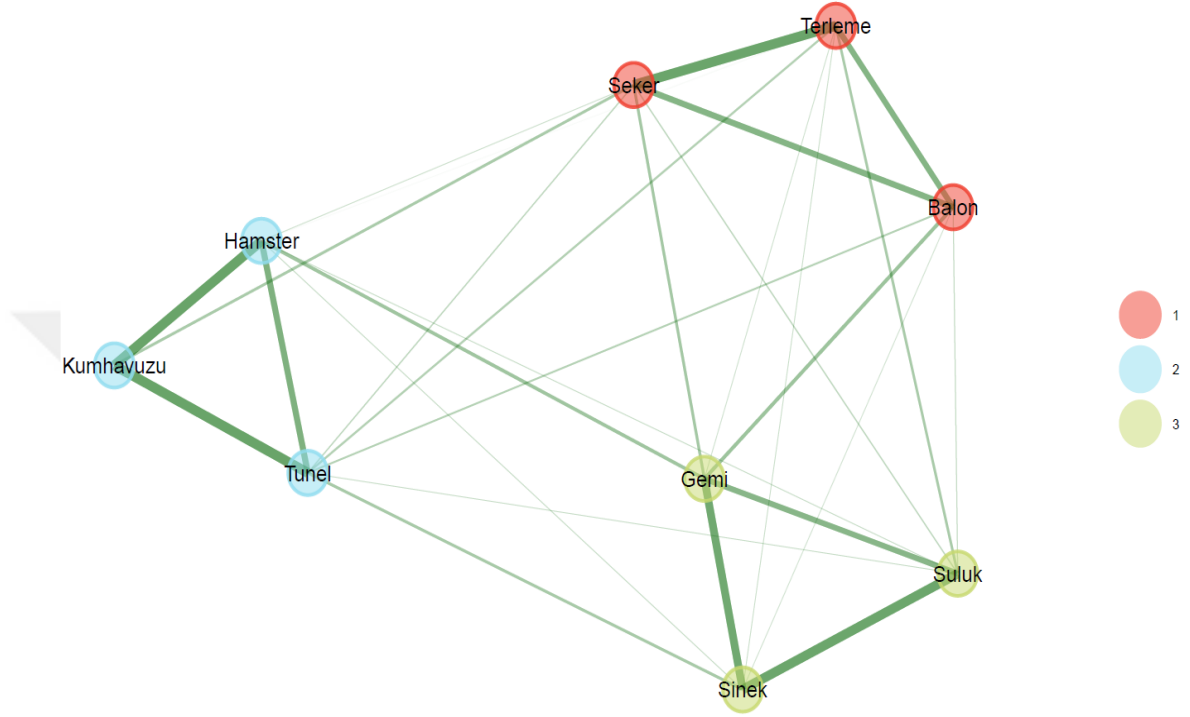
4.2.1.1 Açımlayıcı grafik analizi bulguları

ÇABİYAT-R' in açımlayıcı grafik analizi (EGA) 244 öğrenci ve 9 madde üzerinden gerçekleştirilmiştir. EGA analizi *Glasso (Grafiksel En Küçük Mutlak Daralma ve Seçim Operatörü- Graphical Least Absolute Shrinkage and Selection Operator)* algoritması ile yürütülmüştür. EGA analizinde ilk olarak düğümler arasındaki kenar matris ağırlık değerleri hesaplanmıştır. Matris değerlerinin hesaplanmasının ardından

maddeler arasındaki kısmi korelasyonları gösteren EGA modeli grafiğe dökülmüştür. EGA modeli grafiğine Şekil 4.1’ de yer verilmiştir.

Şekil 4.1.

ÇABİYAT-R testi EGA modeli grafiği



Şekil 4.1 EGA grafiğinde görüldüğü üzere, üçer maddeden oluşan 3 ayrı küme bulunmaktadır. Grafiğin, ÇABİYAT-R için öngörülen 3 faktörlü yapıya işaret ettiği söylenebilir. Kırmızı küme altında toplanan Şeker, Terleme ve Balon kanıt değerlendirme faktörünü; mavi küme altında toplanan Kum Havuzu, Hamster ve Tünel deney tasarlama faktörünü; yeşil küme altında toplanan Gemi, Suluk ve Sinek maddeleri ise hipotez geliştirme faktörünü temsil etmektedir. Aynı küme altında toplanan maddelerin birbirlerine daha kalın çizgiler ile bağlanmaları yapının homojenliğini göstermektedir (Christensen ve Golino, 2021). Yapı incelendiğinde Şeker-Terleme, Kum Havuzu-Hamster, Kum Havuzu-Tünel, Gemi-Sinek ve Sinek-Suluk maddeleri arasındaki kenarların görece daha kalın olduğu görülmektedir. Bu durum söz konusu madde çiftleri arasında daha yüksek bir ilişkinin varlığına işaret etmektedir. Diğer taraftan farklı kümeler altında toplanan madde çiftlerinin birbirleriyle olan kenar ilişkilerinin daha ince oldukları söylenebilir (örneğin Şeker-

Suluk vb.). Bu durumun ayrı faktörler altında toplanan maddelerin görece daha zayıf korelasyona işaret ettikleri söylenebilir. EGA modeli kenar matris değerlerine Tablo 4.8’de yer verilmiştir.

Tablo 4.8

ÇABIYAT-R EGA modeli kenar matris ağırlıkları (w)

Maddeler	KD1	KD2	KD3	DT1	DT2	DT3	HG1	HG2	HG3
KD1 Şeker	-	.35	.22	.05	.07	.05	.00	.04	.11
KD2-Terleme		-	.22	.01	.00	.07	.04	.09	.02
KD3- Balon			-	.00	.00	.05	.03	.04	.13
DT1- Hamster				-	.36	.24	.04	.03	.12
DT2- Kum Havuzu					-	.36	.00	.00	.00
DT3- Tünel						-	.10	.04	.00
HG1-Sinek							-	.34	.31
HG2-Suluk								-	.21
HG3- Gemi									-

Tablo 4.8 incelendiğinde kenar matris ağırlıklarının $w=.00$ ile $w=.36$ arasında değiştikleri görülmektedir. Tabloda yer alan “0” değerleri söz konusu madde çiftleri arasında bağın olmadığını göstermektedir. En yüksek kenar matris ağırlık değerleri Hamster- Kum Havuzu ve Kum Havuzu- Tünel maddeleri arasındadır (.36). Bu maddeyi .35 ile Şeker-Terleme ve .34 ile Sinek-Suluk madde çiftleri takip etmektedir. Farklı faktör altında toplanmalarına rağmen Gemi-Şeker ($w=.11$), Gemi-Balon ($w=.13$) ve Gemi-Hamster ($w=.12$) madde çiftleri arasındaki kenar çizgilerinin nispeten daha kalın kenarlar ile birbirlerine bağlandıkları söylenebilir. Bu durum madde çiftlerinin aynı faktör altında olmasalar da diğer madde çiftlerine kıyasla daha yüksek ilişkiye sahip oldukları şeklinde yorumlanabilir.

ÇABIYAT-R testinin yapı ve faktör tutarlılıkları ile madde kararlılıkları bootEGA analizi ile test edilmiştir. Analiz, veri setinde yer alan aynı sayıda vaka ile 500 kez tekrarlamalı olarak gerçekleştirilmiştir. BootEGA betimsel istatistik sonuçlarına Tablo 4.9’da yer verilmiştir.

Tablo 4.9

BootEGA betimsel istatistik sonuçları

n.boots	Median.dim	SE.dim	CI. dim	Lower.CI	Upper.CI	Lower. Quantile	Upper. Quantile
500	3	0.14	0.28	2.71	3.28	3	3

BootEGA betimsel istatistik tablosu incelendiğinde 500 tekrarlar ile yürütülen analizler sonrasında yapı %95 güven aralığında en düşük 2.71, en çok 3.28 ve ortalama olarak 3 faktöre işaret etmektedir. Standart hata değerinin 0.14 olarak hesaplanması, faktör sayısındaki değişimin düşük olduğu şeklinde yorumlanabilir (Christensen ve Golino, 2021). Faktör sayısına ilişkin daha yorumlanabilir olan frekans dağılımına Tablo 4.10'da yer verilmiştir.

Tablo 4.10.

BootEGA analizi faktör sayısı frekans dağılımları

Faktör Sayısı	Frekans
2	0.022
3	0.978

Faktör frekans dağılım tablosu incelendiğinde %97,8'inde yani 500 tekrar analizinin yaklaşık 490 tanesinde 3 faktörlü yapı elde edilirken %2,2'sinde yani 500 tekrar analizinin yaklaşık 10 tanesinde iki faktörlü yapı elde edilmiştir. Dolayısıyla 3 faktörlü yapının 2 faktörlü yapıya kıyasla daha tutarlı ve baskın olduğu görülmüştür. Bu bağlamda ÇABİYAT-R için yapısal tutarlılık kriterinin sağlandığı, 3 faktörlü yapının ÇABİYAT-R için daha istikrarlı olduğu söylenebilir. Yapısal tutarlılık analizinde faktörlerin tutarlılık düzeyleri de incelenmiştir. Analiz sonucunda birinci, ikinci ve üçüncü faktörler için sırasıyla 0.990, 1.00 ve 0.997 tutarlılık katsayılarına ulaşılmıştır. Elde edilen katsayı değerlerinin .75'in üzerinde olmalarından dolayı faktörlerin yüksek tutarlılık gösterdikleri söylenebilir (Golino vd., 2021).

Yapısal tutarlılıktan sonra bootEGA analizi ile ÇABİYAT-R madde kararlılıkları da incelenmiştir. Faktörler bağlamında madde kararlılık değerlerine Tablo 4.11 ve Şekil 4.2'de yer verilmiştir.

Tabloda 0 değerleri yorumlamada karmaşıklık yaratmamak adına silinmiştir. Tablo 4.10 incelendiğinde Şeker, Terleme, Hamster, Kum Havuzu ve Tünel'in 1.00, Balon'un 0.99, Sinek ve Suluk maddelerinin 0.978 ve son olarak Gemi maddesinin 0.974 düzeyinde madde kararlılığına sahip oldukları görülmüştür. Balon (.010), Sinek (.022), Suluk (.022) ve Gemi (.026) maddelerinde ise önemli olmayan düzeyde ($\leq .80$) çapraz yüklerin olduğu görülmüştür (Christensen ve Golino, 2019; 2021). Şekil 4.2 incelendiğinde de deneysel olarak maddelerin yüksek bir tutarlılıkla (0.97 - 1.00) aynı faktör altında toplandıkları görülmektedir. Kırmızı ve 1 numaralı düğümler

kanıt değerlendirme görevleri olan Şeker, Terleme ve Balon maddelerini; Mavi ve 2 numaralı düğümler deney tasarlama görevleri olan Hamster, Kum Havuzu ve Tünel maddelerini; son olarak yeşil ve 3 numaralı düğümler ise hipotez geliştirme görevleri olan Sinek, Suluk ve Gemi maddelerini barındırmaktadır.

Tablo 4.11

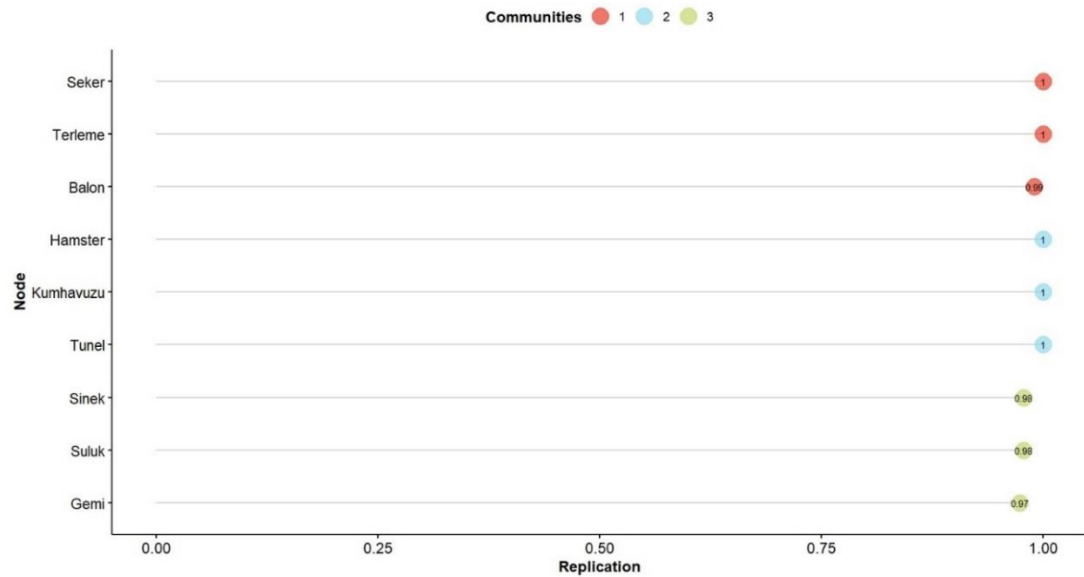
BootEGA analizi faktörler arası madde kararlılık katsayı sonuçları

Madde	Faktör 1	Faktör 2	Faktör 3
KD1-Şeker	1.000		
KD2-Terleme	1.000		
KD3-Balon	0.990		0.010
DT1-Hamster		1.000	
DT2-Kum Havuzu		1.000	
DT3-Tünel		1.000	
HG1-Sinek	0.022		0.978
HG2-Suluk	0.022		0.978
HG3-Gemi	0.026		0.974

EGA ve bootEGA analizleri genel olarak değerlendirildiğinde ÇABİYAT-R testinin öngörülen 3 faktörlü bir yapıya hizmet ettiği söylenebilir. Yapısal tutarlılık analizine göre 3 faktörlü yapı yaklaşık olarak %98 düzeyinde tutarlı sonuçlar vermiştir. Madde kararlılık katsayılarının .70'in üzerinde olmalarından dolayı tüm maddelerin oldukça kararlı oldukları ve belirgin bir şekilde tek bir faktör altında toplandıkları görülmüştür (Christensen ve Golino, 2021; Golino ve Christensen, 2020).

Şekil 4.2.

BootEGA analizi deneysel olarak boyutlarda yer alan maddelerin kararlılık katsayıları



4.2.1.2 Açıklayıcı yapısal eşitlik modellemesi bulguları

ÇABİYAT-R testinin teorik altyapısı kanıt değerlendirme, deney tasarlama ve hipotez geliştirme olmak üzere 3 faktörden oluşmaktadır. Yapılan analizlerde üç faktörlü boyutun yapıya uygunluğunu kontrol etmek ve yapıya ilişkin ek kanıtlar sunmak amacıyla farklı modeller de test edilmiştir. Bu bağlamda ÇABİYAT-R için tek faktörlü DFA ve AYEM, iki faktörlü DFA ve AYEM, üç faktörlü DFA ve AYEM, üç faktörlü ikinci düzey DFA ve AYEM olmak üzere toplamda 9 model test edilmiştir. İki faktörlü DFA ve AYEM için kanıt değerlendirme maddeleri bir faktör, hipotez geliştirme ve deney tasarlama maddeleri bir faktör olacak şekilde belirlenmiştir. Modelleri test etme sürecinde kestirme ve döndürme tekniklerinden MLR ve Oblique benimsenmiştir. Modellerin karşılaştırıldığı uyum indeks değerlerine Tablo 4.12’de yer verilmiştir.

Tablo 4.12’de görüldüğü üzere tek faktörlü DFA ile AYEM ve iki faktörlü DFA ile AYEM modellerinin uyum indeks kriterlerini karşılamadıkları görülmüştür. Diğer taraftan uyum indeks değerleri Model 5 için (Üç faktörlü DFA); $\chi^2/sd = 1.01$, $p > .05$; CFI = 0.99; TLI = 0.99; RMSEA = 0.00; SRMR = 0.03; AIC = 6761.21; BIC = 6866.12, Model 6 için (üç faktörlü AYEM); $\chi^2/sd = 0.63$, $p > .05$; CFI = 1.00; TLI = 1.01; RMSEA = 0.00; SRMR = 0.01; AIC = 6768.48; BIC = 6915.36, Model 7 için (Üç faktörlü ikinci düzey DFA); $\chi^2/sd = 0.96$, $p > .05$; CFI = 1.00; TLI = 1.00; RMSEA = 0.00; SRMR = 0.03; AIC = 6761.21; BIC = 6866.12 ve Model 8 için (üç faktörlü ikinci düzey AYEM); $\chi^2/sd = 0.63$, $p > .05$; CFI = 1.00; TLI = 1.01; RMSEA = 0.00; SRMR = 0.01; AIC = 6768.48; BIC = 6915.36 şeklinde elde edilmişlerdir.

DFA ve AYEM uyum indeksleri kabul uyum ölçütleri bağlamında değerlendirildiğinde 3 faktörlü birinci ve 3 faktörlü ikinci düzey modellerinin mükemmel düzeyde uyum gösterdikleri söylenebilir. Değerler incelendiğinde üç faktörlü birinci ve ikinci düzey ile yürütülen DFA ve AYEM analiz bulgularının birbirlerine oldukça yakın oldukları ifade edilebilir. ÇABİYAT-R testinin kuramsal çerçevesi ikinci düzeyden hiyerarşik bir yapıya daha uygun olması sebebiyle devam analizler ve incelemeler Model 7 ve Model 8 üzerinden ilerletilmiştir.

Tablo 4.12

Uyum kabul ölçütleri ve model uyum indeks sonuçları

	χ^2/sd	CFI	TLI	RMSEA	SRMR	AIC	BIC
Uyum Kabul Kriterleri	<2 ; $p>0.05$	≥ 0.95 ; mükemmel uyum	>0.95 ; mükemmel uyum	≤ 0.05 ; mükemmel uyum	≤ 0.05 ; mükemmel uyum	En düşük değer	En düşük değer
Kaynak	Bollen, 2014; Kline, 2015	Bentler, 1990	Hu ve Bentler, 1999	Brown ve Cudeck, 1993	Tabachnick ve Fidell, 2012	Akaike, 1974	Schwarz, 1978
Model 1: Tek faktörlü DFA	7.08; $p=.00$	0.77	0.70	0.15	0.08	6922.08	7016.50
Model 2: Tek Faktörlü AYEM	6.83 $p=.00$	0.76	0.69	0.15	0.08	6922.08	7016.50
Model 3: İki faktörlü DFA	5.15 $p=.00$	0.85	0.79	0.13	0.07	6866.79	6964.71
Model 4: İki faktörlü AYEM	5.13 $p=.00$	0.88	0.78	0.13	0.05	6828.49	6950.89
Model 5: Üç faktörlü DFA	1.01 $p=.44$	0.99	0.99	0.00	0.03	6761.21	6866.12
Model 6: Üç faktörlü AYEM	0.63 $p=.81$	1.00	1.01	0.00	0.01	6768.48	6915.36
Model 7: Üç faktörlü ikinci düzey DFA	0.96; $p=.51$	1.00	1.00	0.00	0.03	6761.21	6866.12
Model 8: Üç faktörlü ikinci düzey AYEM	0.63; $p=.81$	1.00	1.01	0.00	0.01	6768.48	6915.36

İkinci düzeyden 3 boyutlu DFA ve AYEM modelleri arasındaki uyum indeksleri farkları incelendiğinde $\Delta CFI = 0.00$; $\Delta TLI = 0.01$; $\Delta RMSEA = 0.00$; $\Delta SRMR = 0.02$; $\Delta AIC = 7.274$; $\Delta BIC = 49.24$ hesaplanmıştır. Modeller arasında anlamlı bir farkın olup olmadığına yönelik ek kanıt sunmak amacıyla Satorra-Bentler ki-kare anlamlılık analizi gerçekleştirilmiştir. Satorra-Bentler ki-kare anlamlılık testi MLR kestirim yöntemlerinde modellerin ölçeklenmiş ki-kare ve serbestlik dereceleri üzerinden gerçekleştirilen bir model anlamlılık testidir (Satorra ve Bentler, 2001). Analiz sonuçlarına göre ki-kare farkı ($\Delta\chi^2$) = 14.02 ve serbestlik dereceleri farkı 12 olarak bulunmuştur. Serbestlik derece değerinin 12 olması durumunda $p < .05$ düzeyinde kritik ki-kare değerinin 21.03 olduğu belirlenmiştir. İki model arasındaki ki-kare değerinin kritik değerden düşük olması ($14.02 < 21.03$) sebebiyle modeller arasında anlamlı bir fark olmadığı sonucuna ulaşılmıştır. Modellere ilişkin madde faktör yük, açıklanan madde varyans ve madde benzersizlik (artık madde varyans değeri) değerlerine Tablo 4.13'te yer verilmiştir.

Tablo 4.13

Üç faktörlü ikinci düzey DFA ve AYEM modelleri faktör yükleri, açıklanan ve artık varyans değerleri

Faktör	Madde	Üç Faktörlü İkinci Düzey DFA			Üç Faktörlü İkinci Düzey AYEM				
		λ	R^2	δ	λ_{KD}	λ_{DT}	λ_{HG}	R^2	δ
KD	KD1	0.78*	0.61	0.39	0.77*	0.07*	-0.05*	0.61	0.39
	KD2	0.74*	0.55	0.45	0.76*	-0.04	0.02	0.56	0.44
	KD3	0.67*	0.44	0.56	0.62*	-0.02	0.09	0.44	0.56
DT	DT1	0.74*	0.55	0.45	-0.06	0.72*	0.10	0.55	0.45
	DT2	0.71*	0.51	0.49	0.02*	0.83*	-0.14*	0.60	0.40
	DT3	0.76*	0.58	0.42	0.05	0.65*	0.08	0.54	0.46
HG	HG1	0.75*	0.56	0.44	-0.10	0.02	0.84*	0.63	0.37
	HG2	0.72*	0.52	0.48	0.07	-0.01	0.67*	0.51	0.49
	HG3	0.73*	0.54	0.46	0.12*	0.03*	0.62*	0.51	0.49

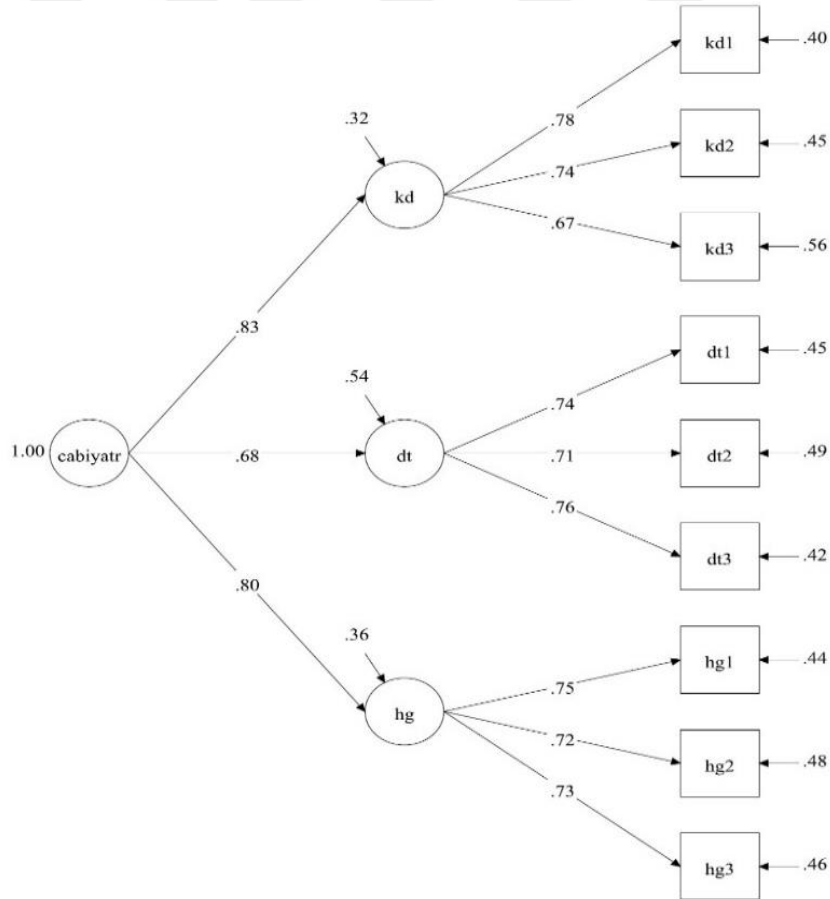
λ : Faktör yükleri. δ : Madde benzersizliği/artık madde varyansı. λ_{KD} : Kanıt değerlendirme faktör yükleri. λ_{DT} : Deney tasarlama faktör yükleri. λ_{HG} : Hipotez geliştirme faktör yükleri

Tablo 4.13 incelendiğinde üç faktörlü DFA modeli için faktör yüklerinin $\lambda = .67$ ile $\lambda = .78$ arasında, maddelerin açıkladıkları varyans değerlerinin ise $R^2 = .44$ ile $R^2 = .61$ arasında değiştikleri görülmüştür. Diğer taraftan üç faktörlü AYEM sonuçları

incelendiğinde faktör yüklerinin $\lambda=.62$ ve $\lambda=.84$ arasında değiştiği görülmektedir. KD1 maddesinin DT alt ölçeğinde $\lambda=0.07$ ve HG alt ölçeğinde $\lambda=-0.05$; DT2 maddesinin, KD alt ölçeğinde $\lambda=.0.02$ ve HG alt ölçeğinde $\lambda=-0.14$; HG3 maddesinin KD alt ölçeğinde $\lambda=0.12$ ve DT alt ölçeğinde $\lambda=0.03$ $p<.01$ düzeyinde anlamlı fakat oldukça zayıf çapraz yük değerlerine rastlanmıştır. İki modele ilişkin madde benzersizliği olarak ifade edilen artık madde varyanslarının $\delta=.39$ ile $.56$ arasında değiştikleri görülmektedir. Bu bağlamda faktör yük (λ), madde bazında açıklanan varyans (R^2) ve madde benzersizlik değerlerinin (δ) iki model için de iyi uyum gösterdikleri söylenebilir. ÇABIYAT-R için AFA ve AYEM modellerinden elde edilen teorik alt yapılar Şekil 4.3' ve Şekil 4.4'te yer verilmiştir.

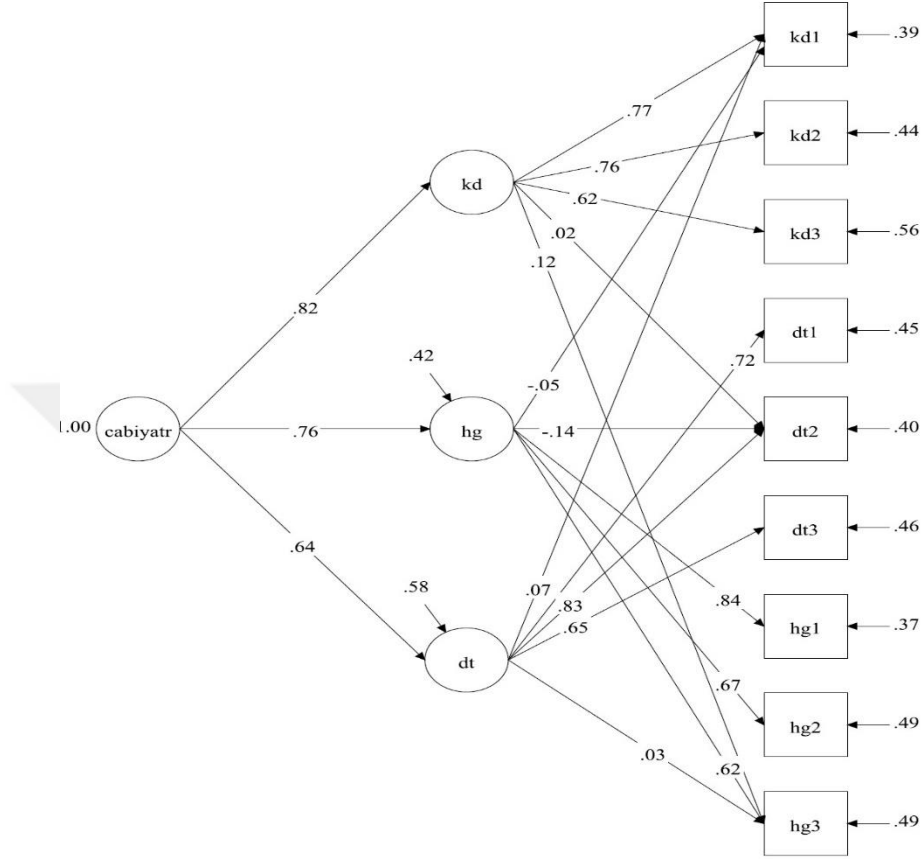
Şekil 4.3

ÇABIYAT-R DFA modelinin teorik alt yapısı



Şekil 4.4

ÇABIYAT-R AYEM modelinin teorik alt yapısı



4.2.1.3 Yakınsama ve ıraksama geçerliklerine ilişkin bulgular

Yakınsama ve ıraksama geçerliklerine ilişkin AVE, CR ve MSV değerleri üç faktörlü ikinci düzey DFA ve AYEM sonuçlarından elde edilen katsayılar üzerinden hesaplanmıştır. AVE, gizil bir yapıyı temsil eden bir dizi madde arasındaki yakınsamanın özetini ifade etmektedir. Ayrıca AVE, faktörlerin yapıyı açıklama yüzdeliği olarak da yorumlanmaktadır (Hair vd., 2019). CR ise her bir maddenin faktöre katkısını ayrı ayrı ele alan bir iç tutarlılık ölçüsü olarak kabul edilmektedir (Jöreskog, 1971). MSV ise bir faktörün diğer faktörlerle olan en yüksek ortak varyansını ifade etmektedir. Analizler sonucunda elde edilen katsayılar Tablo 4.14’te yer verilmiştir.

Tablo 4.14*Modellere ilişkin AVE ve CR katsayı sonuçları*

Bileşen	Madde	Üç Faktörlü DFA				Üç Faktörlü AYEM			
		AVE	$\sqrt{AVE}$	CR	MSV	AVE	$\sqrt{AVE}$	CR	MSV
	KD1								
KD	KD2	0.535	0.73	0.775	0.435	0.518	0.72	0.762	0.385
	KD3								
	DT1								
DT	DT2	0.548	0.74	0.784	0.312	0.548	0.74	0.783	0.277
	DT3								
	HG1								
HG	HG2	0.538	0.73	0.777	0.435	0.518	0.72	0.756	0.385
	HG3								

KD: Kanıt değerlendirme alt ölçeği. DT: Deney tasarlama alt ölçeği. HG: Hipotez geliştirme alt ölçeği

Kanıt değerlendirme, deney tasarlama ve hipotez geliştirme alt ölçeklerine ilişkin AVE değerleri üç faktörlü DFA için sırasıyla 0.535, 0.548, 0.538 ve CR değerleri yine aynı sırayla 0.775, 0.784 ve 0.777 olarak hesaplanmıştır. Üç faktörlü AYEM için alt ölçeklere ilişkin AVE değerleri sırasıyla 0.518, 0.548, 0.518 ve CR değerleri aynı sırayla 0.762, 0.783 ve 0.756 olarak hesaplanmıştır. Her iki modelde yer alan bileşenlerin AVE değerleri göz önünde bulundurulduğunda açıkladıkları varyans yüzdelерinin %50'nin üzerinde olduğu görülmektedir. Literatürde AVE değerinin .50 (Fornell ve Larcker, 1981), CR değerinin ise .70 ve üzerinde olması beklenmektedir (Hair vd., 2019). Bu bağlamda iki modelin de ÇABİYAT-R testinin yakınsama geçerliğine ilişkin kanıt sundukları görülmüştür.

Tablo 4.14'te ayrıca faktörlere ilişkin AVE değerlerinin kareköklerine yer verilmiştir. ÇABİYAT-R testinin ıraksama geçerliğini kontrol edebilmek amacıyla bileşenler arasındaki Pearson Momentler Çarpımı Korelasyon Katsayısı incelenmiştir. Üç faktörlü ikinci düzey DFA modeli için KD-DT arasındaki korelasyon katsayısı $r=.559$, $p<.001$; KD-HG arasındaki korelasyon katsayısı $r=.66$, $p<.001$ ve DT-HG arasındaki korelasyon katsayısı $r=.539$, $p<.001$ olarak hesaplanmıştır. Üç faktörlü ikinci düzey AYEM modeli için ise KD-DT arasındaki korelasyon katsayısı $r=.527$, $p<.001$; KD-HG arasındaki korelasyon katsayısı $r=.621$, $p<.001$ ve DT-HG

arasındaki korelasyon katsayısı $r=.49$, $p<.001$ olarak hesaplanmıştır. İraksama geçerliğinin sağlanabilmesi için alt ölçekler bağlamında hesaplanan $\sqrt{AVE}$ değerlerinin, faktörler arasındaki korelasyon katsayılarından büyük olmaları beklenmektedir (Fornell ve Larcker, 1981). Tablo incelendiğinde kanıt değerlendirme, deney tasarlama ve hipotez geliştirme alt ölçekleri için hesaplanan $\sqrt{AVE}$ değerlerinin alt ölçekler arasındaki korelasyon katsayılarından büyük oldukları sonucuna ulaşılmıştır.

Tablo 4.14'te ayrıca MSV değerlerine yer verilmiştir. Üç faktörlü ikinci düzey DFA modeli için MSV katsayıları kanıt değerlendirme, deney tasarlama ve hipotez geliştirme bileşenleri için sırasıyla 0.435, 0.784 ve 0.777 olarak; üç faktörlü ikinci düzey AYEM modeli için ise yine aynı sırayla 0.385, 0.277 ve 0.385 olarak hesaplanmıştır. Modellerin bileşenleri bağlamında elde edilen AVE değerlerinin MSV değerlerinden büyük oldukları görülmüştür. Elde edilen sonuçlar kanıt değerlendirme, deney tasarlama ve hipotez geliştirme alt ölçeklerinin birbirleri ile ilişkili olmalarına rağmen yapı olarak birbirlerinden ayrıştıklarını ortaya koymaktadır. Dolayısıyla iki modelde de iraksama geçerliğinin sağlandığına yönelik kanıtlar elde edilmiştir.

4.2.2 Ölçüt geçerliğine ilişkin bulgular

ÇABİYAT-R testinin ölçüt geçerliği için uyum ve bilinen grup geçerlikleri incelenmiştir.

4.2.2.1 Uyum geçerliğine ilişkin bulgular

ÇABİYAT-R'nin uyum geçerliği için öğrencilerin fen bilimleri alanındaki akademik başarılarının ölçüldüğü SAGES 2 bataryasından yararlanılmıştır. ÇABİYAT-R alt ölçekler ve toplam test bağlamında elde edilen akıcılık puanları ile SAGES 2 puanları arasındaki korelasyon değerleri incelenmiştir. Pearson Momentler Çarpımı Korelasyon Katsayı analizi sonuçlarına Tablo 4.15' te yer verilmiştir.

ÇABİYAT-R testinin alt ölçekleri ve toplam test puanlarının SAGES 2 puanları ile arasındaki korelasyon katsayılarının .51 ile .68 arasında değiştiği görülmektedir ($p<.001$). ÇABİYAT-R ve SAGES 2 arasında elde edilen korelasyon katsayılarının yüksek düzeyde ($r>.50$) ilişki göstermeleri (Cohen, 1988), uyum geçerliğinin sağlandığına yönelik kanıt oluşturmaktadır.

Tablo 4.15*ÇABİYAT-R ve SAGES 2 uyum geçerliği katsayı sonuçları*

Alt Ölçekler ve Toplam Test	SAGES 2
Kanıt Değerlendirme Toplam	.59**
Deney Tasarlama Toplam	.52**
Hipotez Geliştirme Toplam	.51**
Test toplam	.68**

Alt ölçekler ile SAGES 2 puanları arasında korelasyon katsayılarının büyüklükleri arasında anlamlı bir farkın olup olmadığını anlamak için Fisher r-z (*Fisher r to z transformation*) dönüşümü uygulanmıştır (Meng vd., 1992). Dönüşümler sonucunda alt ölçek korelasyon katsayıları arasındaki farkların istatistiksel olarak anlamlı olmadıkları görülmüştür ($p > .05$).

4.2.2.2 Bilinen grup geçerliğine ilişkin bulgular

ÇABİYAT-R'nin bilinen grup geçerliğini ortaya koymak için öğrencilerin testten aldıkları puanlar, sınıf düzeyleri bağlamında tek yönlü ANOVA ile analizi edilmiştir. Sonuçlara Tablo 4.16'da yer verilmiştir.

Yapılan ANOVA analizleri sonucunda ÇABİYAT-R testi alt ölçek ve toplam test puanlarında anasınıfı, birinci ve ikinci sınıf öğrencileri arasında anlamlı farklara rastlanmıştır ($p < .001$). Hangi gruplar arasında anlamlı farklar olduğunu ortaya koymak amacıyla izleme testleri yapılmıştır. İzleme testleri sonucunda deney tasarlama ve hipotez geliştirme alt ölçeklerinde birinci ve ikinci sınıf öğrencilerinin puanları arasındaki fark $p < .05$ düzeyinde; diğer karşılaştırmalarda $p < .001$ düzeyinde anlamlı farklar bulunmuştur. Etki büyüklük değerlerinin $\eta^2 = .169$ ile $\eta^2 = .344$ arasında değiştikleri görülmektedir. Bu bağlamda gruplar arasındaki puan farklarının yüksek bir etki düzeyine sahip oldukları söylenebilir (Cohen, 1988). Sınıf düzeyinde yapılan karşılaştırmalar genel olarak değerlendirildiğinde, ÇABİYAT-R testinde alt ölçekler ve toplam test puanlarında üst sınıfların daha yüksek puan aldıkları görülmektedir. Dolayısıyla sonuçlar ÇABİYAT-R'nin bilinen grup geçerliği

kapsamında test edilen gelişimsel geçerliğin sağlandığına yönelik kanıt oluşturmaktadır.

Tablo 4.16

Alt ölçek ve toplam test puanlarına göre sınıf düzeyinde yapılan ANOVA analizi sonuçları

Bileşenler	Varyans Kaynağı	Kareler Toplamı	sd	Ortalama Kare	F	p.	İzleme Testi	Fark	p.	η^2
KDT	G. Arası	1446,892	2	723,446	50,244	.001	Tamhane	0-1	.001**	.294
	Grup İçi	3470,043	241	14,399				0-2	.001**	
	Toplam	4916,934	243					1-2	.001**	
DDT	G. Arası	288,534	2	144,267	24,490	.001	Tamhane	0-1	.001**	.169
	Grup İçi	1419,679	241	5,891				0-2	.001**	
	Toplam	1708,213	243					1-2	.042*	
HGT	G. Arası	419,133	2	209,567	27,868	.001	Scheffe	0-1	.001**	.188
	Grup İçi	1812,289	241	7,520				0-2	.001**	
	Toplam	2231,422	243					1-2	.013*	
TT	G. Arası	5654,364	2	2827,182	63,178	.001	Tamhane	0-1	.001**	.344
	Grup İçi	10784,586	241	44,749				0-2	.001**	
	Toplam	16438,951	243					1-2	.001**	

KDT: Kanıt değerlendirme toplam. DDT: Deney tasarlama toplam. HGT: Hipotez geliştirme toplam. TT: Test toplam. 0: Anasınıfı. 1: Birinci sınıf. 2: İkinci sınıf

4.2.3 Cinsiyete göre ölçme değişmezliği analizi bulguları

ÇABİYAT-R testi için madde yanlılığı, cinsiyet değişkeni üzerinden DFA kapsamında ölçme değişmezliği analizleri ile incelenmiştir. Yapı geçerliği analizleri sonucunda DFA ve AYEM modelleri arasında anlamlı bir fark elde edilemediği için AYEM kapsamında ayrıca ölçme değişmezliği analizleri yürütülmemiştir. Ölçme değişmezliği analizleri hiyerarşik bir yapıya sahip olmasına rağmen ÇABİYAT-R için tüm adımlar başarı ile uygulanabilmektedir. Tablo 4.17’ de yapısal değişmezlik, metrik değişmezliği, skaler değişmezlik ve katı değişmezliği (hata varyans değişmezliği) modellerine ilişkin uyum indekslerine yer verilmiştir.

Tablo 4.17’ de görüldüğü üzere uyum indeksleri yapısal değişmezlik modeli için CFI = 0.969; TLI = 0.965; RMSEA = 0.054; SRMR = 0.086; metrik değişmezlik modeli için CFI = 0.990; TLI = 0.988; RMSEA = 0.032; SRMR = 0.047; skaler değişmezlik modeli için CFI = 0.990; TLI = 0.988; RMSEA = 0.032; SRMR = 0.047 ve son olarak katı değişmezlik modeli için CFI = 0.988; TLI = 0.987; RMSEA = 0.033; SRMR = 0.058 olarak elde edilmiştir. Modeller kabul uyum ölçütleri bağlamında değerlendirildiğinde yapısal değişmezlik modelindeki SRMR kriterinin iyi uyum değeri civarında ($0.05 \leq SRMR < 0.08$), diğer kriterlerin tamamının ise mükemmel

ya da mükemmele yakın seviyelerde $CFI \geq 0.95$; $TLI > 0.95$; $RMSEA \leq 0.05$; $SRMR \leq 0.05$ oldukları görülmektedir (Bentler, 1990; Brown ve Cudeck, 1993; Hu ve Bentler, 1999; Tabachnick ve Fidell, 2012). Model uyum indeks farkına yönelik Chen (2007) ve Cheung ve Rensvold (2002); $\Delta CFI \leq -0.01$; $\Delta TLI \leq -0.01$; $\Delta RMSEA \geq +0.015$; $\Delta SRMR \geq +0.03$ (Metrik) veya $+0.015$ (Skaler ve Katı) olması durumunda modelin reddedilmesi gerektiğini belirtmektedirler. Bu bağlamda modeller arasındaki uyum indeks farklarının tamamının kabul edilebilir aralıklarda olduğu görülmüştür.

Tablo 4.17

Ölçme değişmezliği modellerine ilişkin uyum indeksleri

Uyum İndeksleri	Modeller			
	Yapısal Değişmezlik	Metrik Değişmezlik	Skaler Değişmezlik	Katı Değişmezlik
CFI	0.969	0.990	0.990	0.988
ΔCFI	-	0.021✓	0.000✓	-0.002✓
TLI	0.965	0.988	0.988	0.987
ΔTLI	-	0.023✓	0.000✓	-0.001✓
RMSEA	0.054	0.032	0.032	0.033
$\Delta RMSEA$	-	-0.022✓	0.000✓	0.001✓
SRMR	0.086	0.047	0.047	0.058
$\Delta SRMR$	-	-0.039✓	0.000✓	0.011✓
Karar	-	✓	✓	✓

Tamam N:244; Erkek N: 142; Kız N:102

Ölçme değişmezliği sağlandıktan sonra gizil faktörlerin ortalamalarının kız ve erkek öğrenciler arasında bir fark yaratıp yaratmadığı incelenmiştir. Bu analiz sırasında faktör yükleri erkekler için 0'a sabitlenirken kız öğrencilerinin faktör yükleri ise serbest bırakılmıştır. Elde edilen model uyum indekslerinin tamamının kabul edilebilir düzeylerde olduğu görülmüş ve gizil faktör ortalamalarının karşılaştırılması için uygun oldukları belirlenmiştir $CFI = 0.986$; $TLI = 0.985$; $RMSEA = 0.035$; $SRMR = 0.060$. Tablo 4.18'de kız ve erkek öğrencilerin faktör ortalamaları arasındaki fark ve anlamlılık düzeyleri yer almaktadır.

Ölçüm değişmezlik analizleri genel olarak değerlendirildiğinde tüm aşamalarda değişmezliğin sağlandığı, test edilen modellerin tamamının mükemmel uyum seviyelerinde oldukları, modellerin cinsiyet bağlamında benzer sonuçlar ürettikleri ve faktör ortalamaları arasında anlamlı bir fark elde edilemediği görülmüştür. Dolayısıyla ÇABİYAT-R'de yer alan maddelerin kız ve erkek öğrenciler arasında herhangi bir grup lehine yanlılık barındırmadığı söylenebilir.

Tablo 4. 18

Kız ve erkek öğrenciler faktör ortalamaları karşılaştırmaları

Faktör	Kız Öğrenci Grubu (Ort.)	Erkek Öğrenci Grubu (Ort.)	t-değeri	p-değeri	Yorum
KD	-0.118	0.000	-0.031	0.975	Anlamlı değil
DT	-0.140	0.000	-0.035	0.972	Anlamlı değil
HG	-0.001	0.000	-0.010	0.992	Anlamlı değil

KD: Kanıt değerlendirme; DT: Deney tasarlama; HG: Hipotez geliştirme

4.3 ÇABİYAT-R'nin Güvenirliği

Çalışmanın güvenilirlik analizleri kapsamında iç tutarlılık ve kodlayıcılar arası güvenilirlik analizleri yapılmıştır. İç tutarlılık güvenirligi için Cronbach Alpha ve McDonalds güvenilirlik katsayıları hesaplanmıştır. Kodlayıcılar arası güvenilirlik analizleri için Cohen's Kappa ve Sınıf İçi Korelasyon Katsayısı (ICC) katsayıları yorumlanmıştır.

4.3.1 ÇABİYAT-R'nin iç tutarlılığına ilişkin bulgular

ÇABİYAT-R'nin iç tutarlılık güvenilirlikleri alt ölçeklerden ve testin toplamından elde edilen puanlar ile incelenmiştir. Tablo 4.19' da güvenilirlik analizleri sonucunda elde edilen katsayılara yer verilmiştir.

Tabloda görüldüğü üzere üretilen Cronbach Alpha katsayıları alt ölçekler bağlamında kanıt değerlendirme $\alpha = .765$, deney tasarlama $\alpha = .779$ ve hipotez geliştirme $\alpha = .770$ olarak hesaplanmıştır. Test toplam Cronbach Alpha katsayısı ise $\alpha = .83$ olarak bulunmuştur. Üretilen McDonald's Omega katsayıları kanıt değerlendirme, deney tasarlama, hipotez geliştirme ve toplam test için sırasıyla $\omega = .765$, $\omega = .782$, $\omega = .771$ ve $\omega = .84$ olarak bulunmuştur. Güvenirlik analizleri kapsamında ayrıca Stratified alpha (Katmanlı Alfa) katsayısı hesaplanarak $\alpha_{\text{strat}} = .875$ olarak bulunmuştur. Stratified Alpha'nın çok boyutlu ve alt ölçeklere sahip yapılarda kullanılması ve raporlanması önerilmektedir (Feldt ve Brennan, 1989). İç tutarlılık kapsamında elde edilen güvenilirlik katsayılarının tamamının .70'in üzerinde olması ÇABİYAT-R'nin iç tutarlılığı için yeterli kanıt oluşturmaktadır (Cortina, 1993; DeVellis ve Thorpe, 2022; Nunnally ve Bernstein, 1994).

Tablo 4.19

ÇABİYAT-R güvenirlik analizleri sonuçları

Ölçek	Madde	Düzeltilmiş Madde Toplam Korelasyonu	Madde silindiğinde Cronbach Alpha	Ölçek Cronbach Alpha (α)	McDonald Omega Katsayısı (ω)
Kanıt Değerlendirme	Şeker -KD1	.626	.655	.765	.765
	Terleme- KD2	.617	.667		
	Balon- KD3	.561	.738		
Deney Tasarlama	Hamster- DT1	.623	.699	.779	.782
	Kum Havuzu-DT2	.627	.702		
	Tünel- DT3	.615	.704		
Hipotez Geliştirme	Sinek- HG1	.639	.668	.770	.771
	Suluk- HG2	.598	.699		
	Gemi- HG3	.593	.710		
Test Toplam	Şeker -KD1	.630	.801	.830	.840
	Terleme- KD2	.593	.806		
	Balon- KD3	.555	.817		
	Hamster- DT1	.499	.817		
	Kum Havuzu-DT2	.448	.823		
	Tünel- D3	.551	.813		
	Sinek- HG1	.546	.814		
	Suluk- HG2	.545	.812		
	Gemi- HG3	.581	.808		

Madde silindiğinde hesaplanan Cronbach Alpha katsayıları incelendiğinde maddelerin analizden çıkartılması durumunda alt ölçek ve test toplama ait güvenirlik katsayılarının azaldığı görülmüştür. İç tutarlılığın bu halinin optimum düzeyde olduğu söylenebilir. İç tutarlılık analizlerinde Cronbach Alpha katsayısının örneklem büyüklüğü ve madde sayılarından etkilendiği bilinmektedir (Akbulut, 2010). Bu yüzden çalışmada güvenirlik analizlerine ek kanıtlar sunmak amacıyla (Cortina, 1993) Tablo 4.20’de korelasyon katsayılarına da yer verilmiştir.

Tablo 4.20 incelendiğinde maddeler arası korelasyon katsayılarının .196 ile .585 arasında, düzeltilmiş madde toplam korelasyon katsayılarının ise .448 ile .630 arasında değiştikleri görülmektedir ($p<001$). Maddeler arası ortalama korelasyon değeri ise .37 olarak hesaplanmıştır. Maddeler arası korelasyon katsayılarının .15 ile .50 arasında değişimleri (Briggs ve Cheek, 1986; Clark ve Watson, 1995), maddeler arası ortalama korelasyon değerinin .40 civarında olması (Clark ve Watson, 2019) ve düzeltilmiş madde toplam korelasyon katsayılarının da .30 üzerinde olmaları beklenmektedir (Aiken, 1997). Bu bağlamda ÇABİYAT-R testinin maddeler arası,

maddeler arası ortalama ve düzeltilmiş madde toplam korelasyon katsayılarının iç tutarlılığa ilişkin yeterli kanıt sundukları söylenebilir.

Tablo 4.20

ÇABİYAT-R korelasyon katsayı sonuçları

Maddeler	KD1	KD2	KD3	DT1	DT2	DT3	HG1	HG2	HG3
KD1	-	,585**	,506**	,330**	,332**	,364**	,322**	,361**	,405**
KD2		-	,494**	,281**	,234**	,337**	,331**	,369**	,364**
KD3			-	,226**	,248**	,320**	,323**	,340**	,397**
DT1				-	,556**	,542**	,318**	,294**	,353**
DT2					-	,546**	,213**	,196**	,231**
DT3						-	,353**	,327**	,310**
HG1							-	,559**	,550**
HG2								-	,502**
HG3									-
DMTK	,630**	,593**	,555**	,499**	,448**	,551**	,546**	,545**	,581**
KDT	,831**	,824**	,827**	,334**	,327**	,410**	,393**	,430**	,470**
DTT	,412**	,344**	,316**	,844**	,815**	,840**	,356**	,333**	,361**
HGT	,439**	,429**	,428**	,388**	,257**	,395**	,823**	,832**	,837**
TT	,748**	,719**	,712**	,599**	,536**	,641**	,633**	,648**	,682**
Ortalama		.726			.592			.654	

KD: Kanıt değerlendirme. DT: Deney tasarlama. HG: Hipotez geliştirme. DMTK: Düzeltilmiş madde toplam korelasyonu. KDT: Kanıt değerlendirme toplam. DTT: Deney tasarlama toplam. HGT: Hipotez geliştirme toplam. TT: Test toplam

Kanıt değerlendirme maddelerinin kanıt değerlendirme alt testi ile arasındaki korelasyon katsayıları .827 ile .831 arasında, deney tasarlama maddelerinin deney tasarlama alt testi ile arasındaki korelasyon katsayıları .815 ile .844 arasında ve hipotez geliştirme maddelerinin hipotez geliştirme alt testi ile arasındaki korelasyon katsayıları .823 ile .837 arasında değişmektedir ($p < .001$). Kanıt değerlendirme maddelerinin toplam test puanı ile aralarındaki korelasyon katsayıları .712 ile .748 arasında, deney tasarlama maddelerinin toplam test puanı ile aralarındaki korelasyon katsayıları .536 ile .641 ve hipotez geliştirme maddelerinin toplam test puanı ile aralarındaki korelasyon katsayıları .633 ile .682 arasında değişmektedir. Yapılan incelemelerde test maddelerinin ait olmadıkları alt testler ile aralarındaki korelasyon katsayılarının .257 ile .439 arasında değiştikleri görülmüştür. Test maddeleri ait olmadıkları alt testler ile zayıf ($r < .30$) ya da orta düzeyde ($.30 \leq r < .50$) ilişki kurarlarken ait oldukları alt test ve toplam test ile yüksek ilişki ($r \geq .50$) kurmaktadır (Cohen, 1988). Maddelerin, toplam test puanları ile aralarındaki korelasyon katsayılarını daha yorumlanabilir hale getirmek amacıyla alt ölçekler kapsamında ortalama korelasyon katsayıları hesaplanmıştır. Kanıt değerlendirme maddelerinin

toplam test puanı ile aralarındaki ortalama korelasyon katsayısı $r=.726$; deney tasarlama maddelerinin toplam test puanı ile aralarındaki ortalama korelasyon $r=.592$ ve hipotez geliştirme maddelerinin toplam test puanı ile aralarındaki ortalama korelasyon değeri $r=.654$ olarak hesaplanmıştır. Elde edilen ortalama korelasyon katsayı büyüklükleri arasında anlamlı bir farkın olup olmadığını öğrenmek amacıyla Fisher r-z dönüşümü uygulanmıştır (Meng vd., 1992). Yapılan değerlendirmeler sonucunda $r_{ortalamaKD}=.726$ ile $r_{ortalamaDT}=.592$ ($z=3.72$, $p<.05$) ve $r_{ortalamaKD}=.726$ ile $r_{ortalamaHG}=.654$ ($z=2.142$, $p<.05$) aralarındaki korelasyon katsayı büyüklüklerinin istatistiksel olarak anlamlı oldukları bulunmuştur. Diğer taraftan $r_{ortalamaDT}=.654$ ile $r_{ortalamaHG}=.592$ ($z=-1.576$, $p>.05$) korelasyon katsayısı arasında istatistiksel olarak anlamlı bir fark olmadığı görülmüştür. Bu kapsamda toplam test puanı ile en yüksek ilişkiyi kanıt değerlendirme maddeleri kurmaktadır. Tablo 4.17 ve Tablo 4.18’ den elde edilen bulgular değerlendirildiğinde ÇABİYAT-R’nin ve alt ölçeklerinin iç tutarlılık düzeylerinin iyi düzeyde olduğu söylenebilir.

4.3.2 Kodlayıcılar arası güvenirlik bulguları

ÇABİYAT-R’nin kodlayıcılar arası güvenirlik analizleri için araştırmacı dışında bir uzmana katılımcılar içeresinden rassal olarak seçilen 99 öğrencinin (her bir sınıf düzeyinde 33 kişi) kağıtları verilmiş ve öğrencilerin puanları tekrar değerlendirilmiştir. Kodlayıcılar arasındaki tutarlılık madde bazında ve toplam test puanları üzerinden sınıf içi korelasyon katsayısı (r_{ICC}) analizi ile yürütülmüştür. Tablo 4.21’de kodlayıcılar arası güvenirlik bulguları yer almaktadır.

Tablo 4.21

ÇABİYAT-R’nin kodlayıcılar arası güvenirlik sonuçları

Madde	r_{ICC}	F (98,98)	p
Şeker-KD1	1.00	-	-
Terleme-KD2	1.00	-	-
Balon-KD3	.970	64,899	.000
Hamster -DT1	.979	95,567	.000
Kum Havuzu-DT2	.981	103,061	.000
Tünel-DT3	1.00	-	-
Sinek-HG1	.978	89,964	.000
Suluk-HG2	1.00	-	-
Gemi-HG3	1.00	-	-
ÇABİYAT-R Toplam	.998	823,676	.000

Tablo 4.21’de görüldüğü üzere r_{ICC} katsayı değerlerinin .979 ile 1.00 arasında değıştikleri görülmektedir. Elde edilen katsayı değeri lerinin kodlayıcılar arası tutarlılık için literatürde kabul edilen mükemmel uyum seviyesinde olduğunu söylenebilir (Koo ve Li, 2016).



5 SONUÇ TARTIŞMA ve ÖNERİLER

Çalışma kapsamında SDDS modeli (Klahr, 2000; Klahr ve Dunbar, 1988) ve çoğul düşünmeye dayalı yaratıcılık çalışmaları (Guilford, 1950, 1967; Torrance, 1988) temel alınarak ÇABİYAT bilimsel yaratıcılık testine kanıt değerlendirme alt testi eklenmiş ve revizyonu yapılmıştır. Madde havuzunun oluşturulmasında görev alan alanında uzman 7 kişilik ekip toplamda 36 kanıt değerlendirme maddesi geliştirmiş, farklı uzmanların görüşleri doğrultusunda hedef gruptan toplam 54 öğrenci ile nitel verilerin toplandığı ön deneme adımına geçilmiştir. Ön deneme sonucunda maddelerde revizyona gidilerek 9 kanıt değerlendirme, 3 hipotez geliştirme ve 3 deney tasarlama olmak üzere 15 maddelik test ile toplamda 252 kişinin verilerinin toplandığı ana uygulama gerçekleştirilmiştir. Ana uygulamada elde edilen veriler üzerinden öncelikle madde analizleri gerçekleştirilerek 9 kanıt değerlendirme maddesinden 3 maddenin seçilmesine karar verilmiş, ÇABİYAT-R testinin 9 maddelik son hali ile (3 kanıt değerlendirme, 3 hipotez geliştirme ve 3 deney tasarlama) psikometrik özellikleri incelenmiştir. Testin psikometrik özellikleri bağlamında yapı geçerliği, ölçüt geçerliği, uyum geçerliği ve bilinen grup geçerliği analizleri ile iç tutarlılık ve puanlayıcılar arası güvenirlik analizleri yapılmıştır.

5.1 ÇABİYAT-R'nin Madde Analizlerine İlişkin Tartışma ve Sonuç

Kanıt değerlendirme alt testi için madde seçim süreci; betimsel istatistik, madde ayırt edicilik, regresyon analizleri, maddeler arası korelasyon ve madde güvenirlik analizleri yapılarak gerçekleştirilmiştir. Yapılan analizler sonucunda belirli ölçütlere dayalı olarak birtakım kararlar alınmış ve kanıt değerlendirme maddelerinden Çalar Saat (KD1), Bıçak (KD2), Mıknatıs (KD3), Yuvarlanma (KD5), Yağmur (KD7) ve Gölge (KD9) maddelerinin analizden çıkarılmalarına; Şeker (KD4), Terleme (KD6) ve Balon (KD8) maddeleri ile analizlere devam edilmesine karar verilmiştir. Öğrencilerin en yüksek ve en düşük aldıkları puanların frekansları kapsamında KD1, KD2, KD3 ve KD5 maddelerinin kolay; KD6, KD7 ve KD9 maddelerinin ise zor oldukları; KD1, KD5 ve KD9 maddelerinin anasınıfı ve birinci sınıf öğrencilerini ayırt edemedikleri; KD4, KD8 ve KD6 maddelerinin ayırt edicilik analizlerinde geniş etki büyüklüğüne sahip olmaları; KD1 maddesinin deney tasarlama alt ölçeğinde yer alan Kum Havuzu maddesi ile arasındaki ilişkinin zayıf ve istatistiksel olarak anlamlı çıkmaması; regresyon analizi sonuçlarında kanıt değerlendirme alt testini en iyi yordayan ilk üç maddenin KD4, KD8 ve KD6 olması ve KD4, KD8, KD6 madde

varyanslarının görece diğer maddelerin varyanslarından yüksek olmaları şeklinde elde edilen bir dizi bulguya dayanarak madde seçimleri yapılmıştır.

Yaratıcılık testlerinde maddelerin çok kolay olmaları durumunda tavan, çok zor olmaları durumunda ise taban etkisinin ortaya çıkabileceği belirtilmektedir (Plucker ve Zabelina, 2009). Bu bağlamda KD1, KD2, KD3 ve KD5 maddelerinin kolay; KD7 ve KD9 maddelerinin ise zor maddeler oldukları düşünülerek analiz dışı bırakılmıştır. KD6 maddesine yanıt veremeyen öğrenci sayısının fazla olmasına rağmen (77) diğer madde analizi bulgularına göre iyi çalışan bir madde olması sebebiyle analize dahil edilmesine karar verilmiştir. Geliştirilen madde, alt test ve toplam test puanlarının sınıf ve bilgi düzeyi bağlamında ayırt edebilmeleri beklenmektedir. KD1, KD5 ve KD9 maddelerinin anasınıfı ve birinci sınıf öğrencileri kapsamında bir fark yaratamaması maddelerin analiz dışında bırakılma sebeplerinden biri olmuştur.

Frekans değerleri genel olarak incelendiğinde kanıt değerlendirme görevlerinde maddelere yanıt üretemeyen öğrencilerin hipotez geliştirme ve deney tasarlama görevlerine yanıt veremeyen öğrencilere kıyasla daha fazla olduğu söylenebilir. Bu durumun nedenlerinden birisinin kanıt değerlendirme görev tipinin hipotez oluşturma ve deney tasarlama görev tiplerine göre daha zorlayıcı olduğu düşünülmektedir. Kanıt değerlendirme görev tipi için öğrencinin kendisine verilen hipotezi kavrayabilmesi, yine kendisine sunulan deney sonuçlarını değerlendirerek deney sonucu- hipotez arasındaki uyuma yönelik fikir üretmesi gerekmektedir (Koslowski, 1996; Kuhn vd.,1988). Burada örtük olarak değerlendirme, karar verme, yaratma şeklinde birden fazla üst düzey düşünme becerisinin devreye girdiği düşünülmektedir. Oysa hipotez geliştirme ve deney tasarlama görevleri için öğrencilerden deney düzeneğinin neden oluşturulduğu, deneyin yapılma amacı, deney düzeneğinde değişiklik yapabilmeye yönelik fikir üretmeleri beklenmektedir. Klahr ve Dunbar (1988) hipotez geliştirildikten sonra test edilmek üzere deneyin tasarlanarak deney sonuçlarından elde edilen kanıtların hipotezle uygunluğuna karar verildiğini ifade etmektedirler. Dolayısıyla kanıt değerlendirme becerisinin söz konusu modelde bir karar mercii rolünü üstlendiği söylenebilir. ÇABİYAT-R testinin teorik alt yapısında kanıt değerlendirme becerisi, hipotez-deney sonuçlarının değerlendirilmesi ve eş zamanlı olarak kanıtlar üretilmesini gerektirmektedir.

Kanıt değerlendirme görevlerinde görece daha fazla sayıda öğrencinin maddelere yanıt üretememesinin bir diğer nedeni çalışmanın veri toplama sürecinin iki ayrı form ile gerçekleştirilmesinden kaynaklanmış olabilir. Testin uzun olmasından dolayı öğrencilere ilk oturumda hipotez geliştirme ve deney tasarlama görevleri, ikinci oturumda ise kanıt değerlendirme görevi uygulanmıştır. Görevlere ilişkin yönergeler doğaları gereği kısmi benzerlik ve farklılıklar barındırmaktadır. Hipotez geliştirme ve deney tasarlama görevlerinin yönerge kalıpları birbirlerine benzerken kanıt değerlendirme görevinin yönergesi bu iki görevden kısmen ayrılmaktadır. Bu bağlamda kimi öğrenciler hipotez geliştirme ve deney tasarlama görevlerinin uygulandığı birinci oturum ile kanıt değerlendirme görevinin uygulandığı ikinci oturum yönerge farklılıklarını tam olarak kavrayamamış olabilirler. Bu durum kanıt değerlendirme maddeleri için görece daha fazla 0 frekansına neden olmuş olabilir. Ayrıca oturumlar arasında yaklaşık olarak bir buçuk aylık bir zaman farkının olması kimi öğrenciler için motivasyon kaybı yaratmış olabilir. Öğrenciler sözü edilen zaman dilimi içerisinde testten uzaklaşmış, ikinci kez değerlendirilme kaygı ve endişesi yaşamış olabilirler.

5.2 ÇABİYAT-R'nin Geçerliğine İlişkin Tartışma ve Sonuç

ÇABİYAT-R'nin geçerliğine ilişkin bulgular yapı geçerliği, ölçüt geçerliği, uyum geçerliği ve bilinen grup geçerliği bağlamında bu bölümde tartışılmıştır.

5.2.1 ÇABİYAT-R'nin yapı geçerliğine ilişkin tartışma ve sonuç

ÇABİYAT-R'nin kanıt değerlendirme, deney tasarlama ve hipotez geliştirme becerilerinden oluşan 3 bileşenli yapısı, son dönemlerde kullanılmaya başlayan açımlayıcı grafik analizi (Christensen ve Golino, 2021) ve açımlayıcı yapısal eşitlik modellemesi (Asparouhov ve Muthén, 2009; Marsh ve Alamer, 2024) ile test edilmiştir. Yapı geçerliği çalışmalarında öncelikle EGA analizi bulgularına ulaşılmış, ÇABİYAT-R'nin kuramsal yapısı için öngörülen faktör sayısına ilişkin genel bir çerçeve sunulmuştur. Ardından AYEM ile DFA modellerine de yer verilerek iki yaklaşıma ait bulgular kıyaslanmış, SDDS modelinin bütünsel olarak test edildiği ve ÇABİYAT-R'nin 3 faktörlü yapısının doğrulandığı görülmüştür.

5.2.1.1 ÇABIYAT-R'nin açımlayıcı grafik analizine (EGA) ilişkin tartışma ve sonuç

EGA analizi sonuçlarına göre 3 bileşenin her biri altında 3' er maddenin toplandığı görülmüştür (bkz. Şekil 4.1). Kanıt değerlendirme kümesinin altında Şeker, Terleme ve Balon; deney tasarlama kümesinin altında Kum Havuzu, Hamster ve Tünel; hipotez geliştirme kümesinin altında Gemi, Suluk ve Sinek maddeleri bir arada bulunmaktadır. Kanıt değerlendirme kümesi altında toplanan maddelerin kenar matris ağırlıkları .22 ile .35, deney tasarlama bileşeni altında toplanan maddelerin kenar matris ağırlıkları .24 ile .36 ve hipotez geliştirme bileşeni altında toplanan maddelerin kenar matris ağırlıkları .21 ile .34 arasında değişmektedir. Bileşenler altında kümelenen maddelerin kenar matris ağırlıklarının orta (.25) ve yüksek matris ağırlıklarına (.35) sahip oldukları görülmüştür (Christensen ve Golino, 2021). Sonuçlar ÇABIYAT-R'nin 3 faktörlü bir yapıya işaret ettiğini göstermektedir.

Mevcut çalışmada farklı bileşenler altında toplanan madde çiftlerinin .02 ile .13 arasında zayıf düzeyde (.15) kısmi kenar ağırlıklara (Christensen ve Golino, 2021) sahip oldukları görülmektedir (bkz Tablo 4.8) KD1-HG3 arasında .11, KD3-HG3 arasında .13, DT1-HG3 arasında .12 ve DT3-HG1 arasında .10 düzeyinde kısmi kenar ağırlıkları bulunmuştur. Bu durum, Şekil 4.1' de korelasyonun bulunmadığı (.00) kenar çizgilerine kıyasla kısmi korelasyonların olduğu (.00<w<.15) söz konusu madde çiftlerinin görece daha kalın çizgiler ile birbirlerine bağlanmasında da kendini göstermektedir. Her ne kadar madde çiftleri arasındaki değerler zayıf korelasyona ve daha ince çizgilere sahip olsalar da bu durum bileşenler arasında bir örtüşme mi var sorusunu akıllara getirebilir. ÇABIYAT-R'nin 3 faktörlü yapısı temelde bilimsel yaratıcılığı ölçmeyi hedeflemektedir. Faktörler birbirlerinden ayrılan ama aynı zamanda birbirleri ile ilişkili bilimsel süreçleri barındırmaktadır (Klahr, 2000; Klahr ve Dunbar, 1988). Bu bağlamda kanıt değerlendirme ile hipotez geliştirme ve deney tasarlama ile hipotez geliştirme bileşenleri arasında kısmi ilişkilerin beklenen bir durum olduğu söylenebilir. Bilimsel süreç becerilerinin konu edildiği farklı çalışmalarda da benzer sonuçlar alınmıştır. Ateşgöz ve Sak (2021) çalışmalarında öğrencilerin akıcılık puanları üzerinden hesaplanan deney tasarlama maddelerinin, hipotez geliştirme maddeleri ile aralarındaki korelasyon katsayılarının .32 ile .39 arasında değişen orta düzeyde ilişkiye (Cohen, 1988) sahip olduklarını

raporlamışlardır. Benzer şekilde Ayas (2017) akıcılık puanları üzerinden deney tasarlama ve hipotez geliştirme maddeleri arasında .258 ile .414 arasında değişen korelasyon katsayıları bulurken; Ayas ve Sak (2014) çalışmalarında hipotez geliştirme ile deney tasarlama maddeleri arasında .19 ile .25, hipotez geliştirme maddeleri ile kanıt değerlendirme maddesi arasında .26 ve .28, deney tasarlama maddeleri ile kanıt değerlendirme maddesi arasında .27 ve .34 zayıf ama istatistiksel olarak anlamlı korelasyon katsayıları rapor etmişlerdir.

Bilimsel yaratıcılık testlerinin geliştirildiği ve psikometrik özelliklerinin incelendiği çalışmalarda faktör belirleme tekniği olarak sıklıkla Kaiser K1 kuralının ve/veya paralel analizinin kullanıldığı görülmektedir (Ateşgöz ve Sak, 2021; Ayas, 2017, Ayas ve Sak, 2014; Chin ve Siew, 2015; Kanlı 2014; Siew vd., 2014). Oysa faktör belirlemede kullanılan söz konusu iki tekniğin faktör sayısının belirlenmesinde düşük doğruluk katsayılarına sahip oldukları rapor edilmektedir (Cosemans vd., 2022; Golino vd., 2020; Markos ve Tsigilis, 2024) Kanlı (2014) çalışmasında Kaiser ölçütü ve yamaç birikinti grafiğinin iki faktörlü bir yapı önerdiğini fakat geliştirilen testin 3 faktörlü bir yapıyı öngördüğü gerekçesiyle testin psikometrik özelliklerini bu bağlamda incelediğini raporlamıştır. Yalnızca Kaiser K1 ölçütü temel alınarak PCA analizi ile gerçekleştirilen bir başka çalışmada ikinci maddenin öz değerinin (eigenvalue) .999 olması ve maddenin Kaiser kriterini karşılamaması sebebiyle tek faktörlü bir yapının elde edildiği görülmüştür (Chin ve Siew, 2015). Çalışmada olası iki faktörlü yapının toplam varyansa katkısının ise yaklaşık %17’lik bir değişim yaratacağı görülmektedir. Faktör sayısı belirlemede benimsenen geleneksel tekniklerin, elde edilen analiz sonuçları ve araştırmacıların öngördükleri kuramsal yapılar arasında bir tutarsızlık yaratabilecekleri düşünülmektedir. Mevcut çalışmada ise geleneksel tekniklere kıyasla yüksek performans gösteren ve faktör belirlemede çok yüksek bir doğruluk yüzdesine sahip olan EGA yaklaşımı benimsenmiş, ÇABİYAT-R’nin faktör sayısına ilişkin daha güvenilir sonuçlar elde edilmiştir.

Çalışmada EGA modelinde tanımlanan yapının boyutların ve maddelerin tutarlılığını ortaya koymak, boyutlar arasında bir dalgalanma olup olmadığını kontrol etmek (Christensen ve Golino, 2021) ve EGA sonuçlarına ek bir kanıt sunmak amacıyla bootEGA analizi bulgularına da yer verilmiştir. 500 tekrarlıma ile yürütülen analizler sonrasında yapının %95 güven aralığında %97,8 oranında 3 faktöre, %2,2 oranında

2 faktöre işaret ettiği; 3 faktörlü yapının 2 faktörlü yapıya kıyasla daha tutarlı ve baskın olduğu görülmüştür. Faktörlerin tutarlık katsayılarının ise birinci, ikinci ve üçüncü faktörler için sırasıyla 0.990, 1.00 ve 0.997 oldukları ve elde edilen katsayı değerlerinin .75'in üzerinde olmalarından dolayı faktörlerin yüksek tutarlılık gösterdiği şeklinde yorumlanabilir (Christensen ve Golino, 2021; Golino vd., 2021).

BootEGA analizinde ÇABİYAT-R'nin madde kararlılıkları da incelenmiş, maddelerin 0.97-1.00 arasında değiştikleri ve yüksek oranda (≥ 0.80) kararlılığa sahip oldukları görülmüştür (Christensen ve Golino, 2019). Kanıt değerlendirmeye ait Balon (.010) maddesinin hipotez geliştirme; hipotez geliştirme faktörü altında toplanan Sinek (.022), Suluk (.022) ve Gemi (.026) maddelerinin ise kanıt değerlendirme faktörü altında çapraz yüklere sahip oldukları görülmüştür (bkz. Tablo 4.11). Söz konusu maddelerin %1-2 oranlarında çapraz faktörle ilişkili oldukları ve değerlerin ihmal edilebilir düzeyde ($\leq .80$) kararsızlığa sahip oldukları söylenebilir (Christensen ve Golino, 2019; 2021). Madde kararlılık değerlerinin incelendiği çalışmalarda da benzer yaklaşımların benimsendiği görülmüştür. Santiago vd., (2022) çocuklar ve gençler için geliştirilmiş güçler ve zorluklar envanterinin (SDQ) boyutlarını araştırmak amacıyla EGA ve bootEGA analizleri ile yaptıkları çalışmalarında madde kararlılık düzeyleri .80'nin altında olanları çok boyutlu madde olarak tanımlamışlardır. Christensen ve Golino (2021) madde kararlılık için belirledikleri .65 ile .75 arasındaki maddeleri analizden çıkartarak bootEGA analizi gerçekleştirmişler ve simülasyon çalışmaları sonucunda madde kararlılık düzeyi olarak .70 ve üzerini önermişlerdir. Bu bağlamda EGA ile bootEGA analizi sonuçlarına dayanarak, ÇABİYAT-R'nin 3 faktörlü yapısının tutarlı ve faktörler altında toplanan maddelerin oldukça kararlı oldukları söylenebilir.

5.2.1.2 ÇABİYAT-R' nin doğrulayıcı faktör (DFA) ve açımlayıcı yapısal eşitlik modellemesi (AYEM) analizlerine ilişkin tartışma ve sonuç

Yapı geçerliğinin ikinci adımı ÇABİYAT-R testinin DFA ve AYEM modelleri ile test edilmesini kapsamaktadır. DFA ve AYEM uyum indeksleri kabul uyum ölçütleri bağlamında değerlendirildiğinde, modellerin mükemmel düzeyde uyum gösterdikleri (Tablo 4.12) ve ÇABİYAT-R'nin 3 faktörlü yapısının iki yaklaşımla da doğrulandığı görülmüştür. DFA modeli için faktör yüklerinin $\lambda=.67$ ile $\lambda=.78$ ve AYEM modeli için faktör yüklerinin $\lambda=.62$ ile $\lambda=.84$ arasında değiştikleri görülmüştür. 200 ve üstü

örneklemlerde kabul edilebilir faktör yükleri .40 üzeri olarak tanımlanmaktadır (Hair vd., 2019). Bu bağlamda üç faktörlü DFA ve AYEM modelleri için faktör yüklerinin kabul edilebilir değerlerinin oldukça üstünde oldukları söylenebilir. DFA ve AYEM modellerinde maddelerin açıkladıkları varyans değerlerinin $R^2=.44$ ile $R^2=.61$ arasında ve artık madde varyanslarının ise $\delta=.39$ ile .56 arasında değerler aldıkları görülmüştür. R^2 değerlerinin .50 ve üzerinde olmaları durumunda madde bağlamında açıklanan varyans değerlerinin iyi oldukları şeklinde yorumlanmaktadır (Hair vd., 2019; Kline, 2015). Madde benzersizlik değeri olarak belirtilen artık madde varyans değerlerinin $.10 < \delta < .90$ aralığında olması gerektiği önerilmektedir (Zyl ve Koolester, 2022). AYEM modeli içerisindeki çapraz faktör yüklerinin -0.14 ile 0.12 arasında değiştikleri, KD1 (Şeker), DT2 (kum havuzu) ve HG3 (Gemi) maddelerine ait çapraz faktör yük değerlerinin ise anlamlı oldukları görülmektedir ($p < .01$). AYEM ile yürütülen çalışmalarda çapraz faktör yüklerinin .50'den küçük olması gerektiği belirtilmektedir (Prokofieva vd., 2023; Marsh vd., 2014; Zyl ve Koolester, 2022). AYEM ile yürütülen çalışmalarda da benzer sonuçların alındığı görülmektedir (Alamer, 2022; Garrido vd., 2020; Tóth-Király vd., 2017; Prokofieva vd., 2023). Çapraz faktör yüklerinin çok zayıf olmalarına rağmen anlamlı çıkmalarının sebebi çalışmada test edilen modellerin hiyerarşik bir yapıya sahip olmaları ve hiyerarşik modellerin daha esnek yapılara izin vermesinden kaynaklanmış olabilir. DFA ve AYEM modelleri genel olarak değerlendirildiklerinde faktör yük (λ), madde bazında açıklanan varyans (R^2) ve madde benzersizlik değerlerinin (δ) iyi uyum gösterdikleri söylenebilir.

Elde edilen sonuçlar SDDS modelinin kuramsal alt yapı olarak belirlendiği diğer çalışmalardan ayrılmaktadır. Ayas ve Sak (2014), Bermejo, vd. (2016) BÜT 6-8 için tek boyutlu, Ayas (2017) BÜT 3-5 ve Ateşgöz ve Sak (2021) ÇABİYAT testlerinde iki boyutlu yapılar elde ettiklerini raporlamışlardır. BÜT 6-8 hipotez geliştirme (2 adet), deney tasarlama (2 adet) ve kanıt değerlendirmeye (1 adet) yönelik geliştirilmiş maddelerden oluşmasına rağmen SDDS modelinin doğrulanamamasının ve tek boyutlu bir yapı elde edilmesinin nedeni testte yer alan maddelerin sayıca yetersizliği olabilir. Literatürde her bir faktör için en az 3 maddenin bulunması önerilmekte (Hinkin, 1995; Knekta vd., 2019; Leandre vd., 2012), aksi taktirde faktörün zayıf ve kararsız olabileceği belirtilmektedir (Costello ve Osborne, 2005). Ancak çok özel durumlar için bir bileşen altında en az iki maddenin kabul edilebileceği ifade

edilmektedir (Kline, 2023; Yang ve Green, 2011). Bu bağlamda BÜT' ün teorik olarak öngörülen bileşenleri altında hipotez geliştirme ve deney tasarlama için 2'şer, kanıt değerlendirme için tek madde bulunması tek boyutlu bir yapı elde edilmesine neden olmuş olabilir. Araştırmacılar SDDS modelinin işaret ettiği 3 bileşenli bir yapıdan ziyade tek boyutlu bir yapı elde edilmesine yönelik teorik bir tartışma yürütmüşlerdir. SDDS modeline atıfta bulunularak hipotez geliştirme, deney tasarlama ve kanıt değerlendirme bileşenlerinin hipotez alanı ve deney alanlarından bağımsız değil birbirleriyle etkileşimli bir şekilde ilerledikleri iddia edilmektedir (Ayas ve Sak, 2014). Araştırmacılar, hipotez geliştirme ve deney tasarlama alanlarının birbirinin ardına geldiğini; örneğin deney tasarlama görevinin yalnızca deneyi tasarlamaktan ibaret olmadığını, aynı zamanda deney tasarımı ile ilgili hipotez ve deney sonucunda ortaya çıkan kanıtlar arasındaki uyumu da değerlendirmek gerektiğini iddia etmektedirler. Dolayısıyla araştırmacılara göre her bir madde hipotez geliştirme, deney tasarlama ve kanıt değerlendirme bileşenlerinin birbirleri ile etkileşimlerinin olduğu bilişsel bir sürece benzetilmektedir. SDDS modelinin geliştirilmesinde uygulanan bir dizi ampirik süreç incelendiğinde katılımcılara temelde tek bir görev verildiği bilinmektedir (Klahr ve Dunbar, 1988; Klahr, 2000). Görev üzerinden katılımcıların hipotez oluşturmaları, hipotezi test etmeleri (deney tasarlama) ve deney sonuçlarının kurulan hipotezle uygunluğunu değerlendirmeleri beklenmektedir (kanıt değerlendirme). Görev kapsamında 3 bileşenin koordineli ve etkileşimli bir şekilde hareket etmesinin anlaşılır olduğu düşünülebilir. Oysa BÜT, 3 bileşene yönelik geliştirilmiş 5 farklı maddeden oluşmaktadır. Maddelerin geliştirilme ve sunulma biçimlerindeki farklılıklarından dolayı bilişsel süreçlerin araştırmacıların iddia ettiği şekli ile gerçekleşmeyeceği düşünülmektedir.

Ayas (2017) ve Ateşgöz ve Sak (2021) çalışmalarında hipotez geliştirme ve deney tasarlama bileşenlerinden oluşan iki boyutlu bir yapı elde etmeleri sebebiyle yine mevcut çalışmanın sonuçlarından ayrılmaktadır. Araştırmacılar temelde ifade edilebilecek 3 varsayımla SDDS modelinde revizyona giderek kanıt değerlendirme bileşenini teorik yapıdan çıkarmışlardır: 1) kanıt değerlendirmenin SDDS modelinde hipotez geliştirme ve deney tasarlama bileşenleri arasında aracı bir değişken olarak görev alması, 2) kanıt değerlendirme becerisinin küçük yaş gruplarına uygun olmaması ve 3) kanıt değerlendirmenin özü itibari ile evet/hayır şeklinde ifade

edilebilecek dikotomik bir yapıya sahip olması sebebiyle çoğul düşünme görevlerine uygun olmaması. Kanıt değerlendirmenin, bilimsel düşünme kuram ve modellerinde aracı bir değişkenden çok daha fazlasını ifade ettiği söylenebilir. Örneğin Kuhn vd. (1988) bilimsel keşif sürecini hipotez-kanıt arasındaki farkı ayırt ve koordine edebilme yeteneği olarak tanımlamaktadır. Klahr ve Simon (2000) hipotez ve deney arama alanlarını birbirlerine bağlı 3 temel bileşenden oluştuğunu belirterek kanıt değerlendirmeyi bu bileşenlerden biri olarak ele almaktadırlar. ÇABİYAT-R'nin yapı geçerliği incelendiğinde DFA bulguları kapsamında kanıt değerlendirme faktör yüklerinin .67 ile .78, her bir maddenin yapıya katkılarının %44-61 arasında değiştiği görülmüştür. Bu bağlamda kanıt değerlendirme bileşeni bilimsel süreç becerileri içerisinde bağımsız bir bileşen olduğu şeklinde yorumlanabilir.

Literatürde kanıt değerlendirme becerisinin küçük yaş grubu çocuklarına uygulandığı ve mükemmel uyuma dayalı görev tiplerinde başarılı sonuçlar alındığı çalışmalara rastlanmıştır (Koerber vd., 2005; Piekny ve Maehler, 2012; Piekny vd., 2013; Ruffman vd., 1993). Fakat araştırmalar incelendiğinde Ayas (2017) ve Ateşgöz ve Sak'ın (2021) belirttikleri gibi kanıt değerlendirme görevlerinin tek tip oldukları ve görevlerin çoğul düşünmeye dayalı olmadıkları söylenebilir. Bu durumunun sebebi olarak söz konusu çalışmaların yaratıcılık, bilimsel yaratıcılık yaklaşımından ziyade bilimsel düşünme ya da bilimsel muhakeme ile ilişkilendirilmelerinden kaynaklandığı düşünülmektedir. Diğer bir ifade ile çalışmaların özü itibari ile çoğul düşünmeye dayalı açık uçlu bir görev tasarımına yönelmedikleri görülmektedir.

DFA ve AYEM uyum indeksleri arasındaki farkların $\Delta CFI \leq 0.01$ ve $\Delta RMSEA < 0.015$ değerlerine kıyasla küçük ve eşit olmalarından dolayı önemsiz (Cheung ve Rensvold, 2002), $\Delta TLI \leq 0.01$ ve $\Delta SRMR < 0.01$ değerlerinden büyük olmaları sebebiyle (Tablo 4.12) anlamlı oldukları görülmüştür (Chen, 2007). Yapılan Satorra-Bentler ki-kare anlamlılık sonuçlarına göre iki model arasında anlamlı bir fark olmadığı sonucuna ulaşılmıştır ($p > .05$). Doğrulama yaklaşımlarının birlikte kullanıldığı çalışmalarda da benzer sonuçların alındığı görülmüştür. Marsh ve Alamer (2024) çalışmalarında DFA ve AYEM tabanlı modelleri karşılaştırmışlar ve modellerin her iki yaklaşımla da iyi uyum düzeylerine sahip olduklarını bulmuşlardır. DFA, AYEM ve her iki yaklaşımın dengeli bir yaklaşımı olarak ifade edilen SET-AYEM modellerinin çapraz yüklemelere sunduğu esneklik, model uyum indeksleri

ve model basitliđi gibi kriterler üzerinden karşılaştırıldıđı bir alıřmada üç model arasında belirgin bir farklılıđa rastlanmadıđı, her modelin kendine özgü birtakım avantaj ve dezavantajların olduđu belirtilmiřtir (Marsh vd., 2020). Diđer taraftan AYEM modellerinin daha esnek ve gereki yapılar sunduđu, DFA modellerine kıyasla uyum indeks deđerlerinin daha iyi olduđu rapor edilen farklı alıřmalara da rastlanmaktadır (Király vd., 2017; Liu ve Wang, 2016; Lim, 2022; Zyl ve Klooster, 2022). Mevcut alıřma ile literatürde gerekleřtirilen diđer alıřmalar arasında kısmi benzerlik ve farklılıkların bulunması örnekleme büyüklüđu, alıřmaların farklı disiplinlerde gerekleřtirilmesi ve kuramsal alt yapılardan kaynaklanan birtakım nedenlere bađlı olabilir. DFA ve AYEM analizlerinin yürütüldüđu alıřmalarda simülasyon ile bir araya getirilen büyük veri setlerine rastlanmaktadır. Ayrıca alıřmalarda madde ve faktör sayılarının mevcut alıřma ile kıyaslandıđında daha fazla oldukları görölmektedir. AYEM daha büyük veri gruplarında ve ok sayıda boyutluluđu olduđu veri setlerinde esnek modellemelere izin verirken DFA analizi en az karmařık modelin seilebildiđi bir diđer ifade ile daha basit ve sade modellemelere izin vermektedir (Marsh vd., 2020). ABİYAT-R'nin 9 madde ve 3 boyuttan oluřan sade ve basit kuramsal yapısı, DFA ve AYEM modellemeleri arasında bir fark yaratmamıř olabilir. İki yaklařımın karşılaştırıldıđı alıřmalar incelendiđinde test edilen yapıların psikoloji alıřmalarında kullanılan davranıř temelli bataryalar oldukları görölmektedir. ABİYAT-R'nin bilimsel süreç ve becerilerinin temel alındıđı açık uçlu bir yaratıcılık testi olmasından dolayı diđer alıřmalardan farklılařan bulgular elde edilmiř olabilir.

5.2.1.3 Yakınsama ve ıraksama geerliklerine iliřkin tartıřma ve sonuç

ABİYAT-R'nin yapı geerliđi kapsamında yakınsama ve ıraksama geerlikleri için AVE, CR, $\sqrt{AVE}$ ve MSV deđerleri incelenmiřtir. Kanıt deđerlendirme, deney tasarlama ve hipotez geliřtirme alt testleri için hem DFA hem de AYEM modellerinde .50 üzerinde AVE (Fornell ve Larcker, 1981) ve .70 üzerinde CR deđerleri (Hair vd., 2019) hesaplanmıř ve ABİYAT-R'nin yakınsama geerliđinin sađlandıđı görölmüřtür. Hesaplanan $\sqrt{AVE}$ deđerleri ile alt testlerin birbirleriyle olan korelasyon katsayıları incelenmiř ve $\sqrt{AVE}$ deđerlerinin faktörler arasındaki korelasyon katsayılarından büyük oldukları görölmüřtür. Ayrıca bileřenlere iliřkin AVE deđerlerinin MSV deđerlerinden büyük olmaları sebebiyle (Fornell ve Larcker,

1981) ıraksama geçerliğine yönelik kanıt sunulmuştur. Ateşgöz ve Sak (2021) çalışmalarında ÇABİYAT testinin deney tasarlama ve hipotez geliştirme alt ölçekleri için AVE ve CR değerlerini sırasıyla .635, .839, .676 ve .862 olarak bulmuşlardır. Çalışmadan elde edilen sonuçlar ile mevcut çalışma kıyaslandığında ÇABİYAT ve ÇABİYAT-R için tutarlı sonuçların alınması yakınsak ve ıraksak geçerlikleri bağlamında iki test için güçlü kanıtların varlığı şeklinde yorumlanabilir. Yapılan diğer çalışmalar incelendiğinde yakınsama ve ıraksama geçerliklerine ilişkin bulguların raporlanmadığı görülmektedir. Bu bağlamda araştırmacı, literatürde gerçekleştirilen çalışmalarda rapor edilen faktör yük değerleri ve faktörler arasındaki korelasyon katsayıları üzerinden AVE, CR ve $\sqrt{AVE}$ değerlerini hesaplayarak mevcut çalışma ile karşılaştırmıştır. Ayas (2017) çalışması için hipotez geliştirme faktörü AVE değeri .417 ve CR değeri .588, deney tasarlama faktörü AVE değeri .686, CR değeri ise .811 olarak hesaplanmıştır. Hipotez geliştirme faktörünün yakınsama geçerliğinin kabul edilebilir değerlerde olmadığı görülmektedir (Fornell ve Larcker, 1981). Çalışmadan elde edilen $\sqrt{AVE}$ değerlerinin (.64 ve .82) ise faktörler arasındaki korelasyon katsayılarından ($r=.302 - .435$) büyük olmaları sebebiyle ıraksak geçerliğinin sağlandığı söylenebilir. Kanlı (2017) 3 boyutlu kuramsal yapıyı test ettiği çalışmasında faktörlerin AVE değerleri sırasıyla .270, .347 ve .344; CR değerleri ise .740, .784 ve .676 olarak hesaplanmıştır. AVE değerleri faktör bazında açıklanan varyans değerlerini ifade etmektedir. Söz konusu çalışmada CR değerlerinin .70'in üzerinde olmaları sebebiyle faktör bazında güvenilir sonuçlar alındığı söylenebilir. Bu durumun nedeni faktörler altında toplanan madde sayıları ile ilgili olabilir. Bazı çalışmalarda yapıların yalnızca açıklandığı fakat doğrulanmadığı görülmekle birlikte AVE değerlerinin .50 ve CR değerlerinin ise .70'in üzerinde oldukları hesaplanmıştır (Hu ve Adey, 2002; Siew vd., 2014). Çalışmalarda yüksek AVE ve CR değerlerine ulaşılmasının nedeninin temel bileşenler analizi (PCA) yaklaşımının benimsenmesi olduğu düşünülmektedir. PCA yaklaşımı yüksek faktör yüklerine neden olmakta ve değişkenlerde maksimum varyansı açıklayacak şekilde bileşenler oluşturmaktadır (Hair vd., 2019).

5.2.2 ÇABİYAT-R'nin ölçüt geçerliğine ilişkin tartışma ve sonuç

ÇABİYAT-R testinin ölçüt geçerliği için uyum ve bilinen grup geçerliğinden elde edilen bulgular değerlendirilmiştir.

5.2.2.1 ÇABİYAT-R'nin uyum geçerliğine ilişkin tartışma ve sonuç

ÇABİYAT-R'nin uyum geçerliği için testin alt ölçekler ve toplamından elde edilen akıcılık puanları ile SAGES 2 puanları arasındaki ilişkiye bakılmış, korelasyon katsayılarının alt ölçekler için .51 ile .59 arasında değiştikleri ve toplam ölçek ile .68 düzeyinde ilişkiye sahip olduğu görülmüştür ($p<.001$). ÇABİYAT-R ve SAGES 2 arasında elde edilen korelasyon katsayıları yüksek düzeyde ($r>.50$) ilişki göstermektedir (Cohen, 1988). ÇABİYAT testinin geliştirildiği çalışmada da mevcut çalışma bulguları ile paralel sonuçlar alındığı; deney tasarlama, hipotez geliştirme ve toplam ölçek akıcılık puanlarının SAGES 2 puanları ile sırasıyla .65, .54 ve .71 düzeyinde korelasyona sahip olduğu raporlanmıştır (Ateşgöz ve Sak, 2021). İki çalışmada da SAGES 2 ile yüksek düzeyde elde edilmiş korelasyon katsayıları değerlendirildiğinde ÇABİYAT-R'nin uyum geçerliği için yeterli kanıtı sahip olduğu söylenebilir.

SAGES 2'nin araştırmada kullanılan alt ölçeği fen bilimleri alanında bilimsel bilgiyi ölçmektedir. Yaratıcılığın alana özgü olduğu ve alana özgü yaratıcılığın şekillenmesinde alana özgü bilgi, deneyim ve becerinin gerekli olduğu konusunda ise yaygın bir görüş birliği hakimdir (Alexander, 1992; Amabile, 1996; Feldhusen, 2006; Urban, 1991). Bu bağlamda ÇABİYAT-R bilimsel yaratıcılık testi ile bilimsel bilginin değerlendirildiği SAGES 2 bataryası arasında .50 üzerinde korelasyon katsayılarının oluşması önemli bir bulgudur. Özdemir ve Dikici (2017) yaptıkları çalışmada, bilimsel süreç becerilerinin bilimsel yaratıcılığı doğrudan etkilediğini, bilimin doğasını anlamaya yönelik geliştirilmiş bir ölçeğe (NSKS) ait *test edilebilirlik* alt boyutunun bilimsel süreç becerileri ve bilimsel yaratıcılık arasındaki ilişkiye kısmen aracılık ettiğini, NSKS' nin *yaratıcı* alt boyutunun ise bilimsel süreç beceri testinin bilimsel yaratıcılık testi üzerindeki etkisinde aracı rolü üstlendiğini rapor etmişlerdir.

Fen bilimlerine karşı tutum, bilimsel yaratıcılık ve bilimin doğası arasındaki ilişkinin değerlendirildiği bir başka çalışmada da bilimsel yaratıcılıktan yüksek puan alan öğrencilerin bilimin doğası hakkında daha gerçekçi oldukları; bilimsel yaratıcılıktan

düşük puan alan öğrencilerin ise bilimin doğası hakkında daha geleneksel görüşlere sahip oldukları ve iki grup arasında istatistiksel olarak anlamlı bir fark ($t=-4.701$, $p<.05$) bulunduğu rapor edilmiştir (Usta ve Akkanat, 2015). Mevcut çalışma sonuçları ile kısmi benzerlik gösteren diğer çalışmalarda ise bilimsel bilginin doğasına yönelik ölçekler ile bilimsel yaratıcılık arasında $r=0.245$ (Liang, 2002) ve $r=0.33$ (Jo, 2009) düşük ve orta düzeyde anlamlı korelasyona ulaşıldığı söylenebilir (Cohen, 1988).

Bilimsel yaratıcılık alanında çalışan araştırmacıların, öğrencilerin bilimsel yaratıcılık ile benzer başka disiplinlerdeki yetenekleri (Ayas ve Sak, 2014), fen bilimine yönelik ilgi ve tutumu (Ayas, 2017; Jo, 2009; Kanlı, 2014; Liang, 2002; Usta ve Akkanat, 2015) ve fen bilimlerindeki akademik kazanımlar arasındaki ilişkiyi (Ayverdi vd., 2012; Kanlı, 2014; Liang, 2002) değerlendirdikleri görülmektedir. Ayas ve Sak (2014), BÜT 6-8 ile matematik ve fen notları arasında sırasıyla .31, .35; Ayas (2017), BÜT 3-5 formunun akıcılık puanları ile matematik ve fen notları arasında sırasıyla .404 ve .409; Kanlı (2014) Y-BÇT ile matematik ve fen notları arasında sırasıyla .390 ve .487; bilimsel yaratıcılık ile akademik başarı arasında Huang vd. (2017) $r=.434$; ve Ayverdi vd. (2012) $r=.356$ düzeyinde korelasyon katsayıları rapor etmişlerdir. Bu bağlamda bilimsel yaratıcılık ile ilişkilendirilen testlerin yaklaşık olarak %9 ile %24'ü arasında öğrencilerin matematik ve fen notlarının açıklandığı söylenebilir. Çalışmalar genel olarak incelendiğinde bilimin doğası, karne notları ve akademik başarının bilimsel yaratıcılıkla ilişkilerinin mevcut çalışma sonuçlarıyla paralellik gösterdikleri söylenebilir.

5.2.2.2 ÇABİYAT-R'nin bilinen gruplar geçerliğine ilişkin tartışma ve sonuç

Bilinen gruplar geçerliği geliştirilen bir ölçeğin gruplar arasındaki puan farklarının anlamlılığını test etmede kullanılmaktadır (DeVellis ve Thorpe, 2022; Netemeyer vd., 2003). ÇABİYAT-R'nin bilinen gruplar geçerliği sınıf düzeyleri üzerinden incelenmiş, üst sınıfların alt sınıflardan alt test ve toplam testte daha yüksek puan aldıkları görülmüştür (Tablo 4.16). Elde edilen sonuçların yapılan diğer araştırmalar ile kısmi benzerlik ve farklılıkların olduğu söylenebilir. ÇABİYAT' ın geliştirildiği çalışma sonuçları incelendiğinde alt ölçeklerin ve testin tamamından elde edilen akıcılık, esneklik, bileşik yaratıcılık ve orijinallik puanlarının tamamında üst sınıfların daha yüksek puan aldıkları ve gelişimsel geçerliğin sağlandığı rapor

edilmiştir (Ateşgöz ve Sak, 2021). Hu ve Adey (2002) çalışmalarında 12 ve 13 yaşları arasındaki yaratıcılık puanlarının .01; 13-15 yaş arası öğrencilerin yaratıcılık puanlarının ise .05 düzeyinde anlamlı olarak ayırt edildiğini fakat yaşa bağlı olarak yaratıcılığın doğrusal bir yol izlemediğini raporlamışlardır. Aschauer vd., (2022) yaptıkları çalışmada kendi geliştirildikleri çoğul düşünmeye dayalı bilimsel yaratıcılık testinden üst sınıfların alt sınıflara kıyasla daha yüksek puan aldıklarını ve aradaki farkın anlamlı olduğunu rapor etmişlerdir. Üst sınıfların alt sınıflardan anlamlı bir şekilde daha yüksek puan alındığı mevcut araştırma bulguları ile örtüşen farklı çalışmalara da rastlanmıştır (Çeliker vd., 2015; Yang vd., 2016).

Ayas (2017) BÜT 3-5 formunun psikometrik özelliklerinin incelediği çalışmada hipotez geliştirme ve testin tamamından elde edilen puanların üst sınıflar lehine anlamlı farklılıklar oluşturduğunu ancak deney tasarlama alt testinin 3.ve 4. sınıfları istatistiksel olarak ayırt edemediği sonucuna ulaşmıştır. Araştırmacının da ifade ettiği gibi bu durumun sebebi çoğu çalışmada kendini gösteren 4.sınıf düşüşü olgusu (Camp, 1994; Raina, 1980; Saggar vd., 2019; Smith ve Carlsson, 1983) ile açıklanabilir. Kanlı (2014) Y-BÇT testine ilişkin gelişimsel geçerlik bulguları incelendiğinde toplam test bağlamında 5. ve 6. sınıflar arasında anlamlı olmayan sonuçlar elde edildiği gibi alt testler bağlamında elde edilen sonuçların da tartışmalı ve çelişkili oldukları görülmüştür. Özellikle analogik problem çözme alt boyutunun sınıflar arasında ayırt ediciliğinin tam olarak sağlanamadığı söylenebilir. Araştırmalar incelendiğinde yaratıcılığın bilgi, yaş ve sınıf düzeyi bağlamında mevcut araştırma bulguları ile kimi benzerlik ve farklılıkların yaşandığı görülmektedir. Literatürde yaratıcılığın gelişimsel bir süreci takip edip etmediği konusu tartışmalı olmakla birlikte (Anderson, 1992; Charles ve Runco, 2001; Sak ve Maker, 2006; Simonton, 1997), ÇABİYAT-R'nin anasınıfı, birinci ve ikinci sınıf öğrencilerini akıcılık puanları üzerinden iyi bir şekilde ayırt ettiği söylenebilir.

5.2.3 ÇABİYAT-R'nin cinsiyete göre ölçme değişmezliğine ilişkin tartışma ve sonuç

ÇABİYAT-R'nin maddelerinin potansiyel olarak kız ve erkek öğrenciler arasında bir grup lehine yanlı olup olmadığını incelemek adına literatürde önerildiği gibi madde yanlılık analizleri yapılmıştır (Mellenbergh, 1989). Madde yanlılık analizleri SEM yaklaşımı benimsenerek hiyerarşik yapılı ölçme değişmezliği analizleri ile

yürütülmüştür. Yapılan analizler sonucunda yapısal değişmezlik, metrik değişmezliği, skaler değişmezlik ve katı değişmezlik modellerinin tamamının sağlandığı görülmüştür. Ölçme değişmezliği sonucunda yapılan faktör ortalamaları arasında cinsiyet bağlamında bir farklılık olmadığı tespit edilmiştir ($p=.975$). Literatürde yaratıcılık puanlarının cinsiyet değişkeni üzerinden değerlendirildiği çok sayıda çalışmanın grup ortalamaları üzerinden ilerlediği ve madde yanlılığına yer verilmediği görülmektedir (Baer ve Kaufman 2008; Nakona vd., 2021; Taylor vd., 2024). Ateşgöz ve Sak (2021) çalışmalarında madde yanlılık analizlerini t testi üzerinden gerçekleştirerek erkek ve kız öğrenciler arasında anlamlı bir fark olmadığını ve bu bağlamda test maddelerinin yanlılık barındırmadıklarını raporlamışlardır. Söz konusu bulgunun metodolojik ve istatistiksel olarak sınırlılık barındırdığı düşünülmektedir. Madde yanlılık analizleri madde tepki kuramı tabanlı (Raju ve Ellis, 2002), çok gruplu doğrulayıcı faktör analizleri (Putnick ve Bornstein 2016), Mantel-Haenszel (MH) yöntemi (Mantel ve Haenszel, 1959) şeklinde farklı yaklaşımlar ile ele alınmaktadır. Diğer taraftan madde yanlılığı kapsamında ortalamalar üzerinden gerçekleştirilen değerlendirmeler çoğu psikometrist tarafından reddedilmesine (Camilli & Shepard, 1994; Hunter vd., 1984; Reynolds, 2000) rağmen yaygın bir şekilde kullanılmaya devam etmektedir. Hatta kimi araştırmacılar en tutarlı fakat en yıkıcı sonuçların ortalamalar üzerinden yapılan madde yanlılığı ölçümlerinden alındığını iddia etmektedirler (Warne vd., 2014).

Yaratıcılık alanında yapısal eşitlik modellemelerine dayalı gerçekleştirilen ölçme değişmezliği çalışmalarının sınırlı sayıda olduğu söylenebilir (Kim vd., 2006; Krumm vd., 2014; 2016; Miroshnik vd., 2022; Purc ve Laguna, 2019; Said-Metwaly vd., 2021). Araştırmalar her ne kadar genel yaratıcılığın ölçüldüğü çoğul düşünme testlerine dayandırılmış olsa da mevcut araştırma bulguları ile benzerlik gösterdikleri görülmektedir. TTCT testinin faktör yapısı ve ölçme değişmezliğinin incelendiği araştırmalarda yapısal, metrik, skaler, katı değişmezlik modellerinin sağlandığı ve grupların faktör ortalamaları arasında cinsiyete yönelik bir farklılık bulunmadığı rapor edilmiştir (Kim vd., 2006; Krumm; 2016; Said-Metwaly vd., 2021). Çalışmaların mevcut araştırma bulguları ile tutarlı olmalarının nedeni olarak çoğul düşünme tabanlı yaklaşımın benimsenmesi gösterilebilir.

Bilimsel yaratıcılık alanında yapılan birçok çalışmada mevcut çalışma sonuçları ile paralel bulgular elde edilmiş, erkek ve kız öğrencilerin puanları arasında istatistiksel olarak anlamlı bir farkın olmadığı raporlanmıştır (Ateşgöz ve Sak, 2021; Ayverdi vd., 2012; Baysal vd., 2013; Hu vd., 2010; Jo, 2009; Kanlı, 2017; Mohamed, 2006; Sternberg vd., 2020; Xu vd., 2023). Diğer taraftan sınırlı sayıda çalışma ile mevcut çalışma bulguları ayrılmaktadır. Ortaokul öğrencilerinin bilimsel yaratıcılık düzeylerinin farklı değişkenlere göre incelendiği bir çalışmada (Çeliker vd., 2015) kız öğrencilerin erkek öğrencilerden anlamlı bir şekilde daha yüksek puan aldıkları rapor edilmiştir ($p<.05$). Ortaokul öğrencileri ile yapılan bir başka çalışmada akıcılık ($t=3,916$; $p<0,05$), esneklik ($t=4,743$; $p<0,05$) ve orijinallik puanları ortalamalarında ($t=3,299$; $p<0,05$) kız öğrenciler lehine anlamlı farklılıklar bulunduğu ve kız öğrencilerin bilimsel yaratıcılıklarının erkek öğrencilerden daha yüksek olduğu rapor edilmiştir (Saçar, 2019). Mevcut araştırma bulgusu ile örtüşmeyen iki çalışma incelendiğinde benzer hedef gruplara benzer bir bilimsel yaratıcılık testinin uygulandığı görülmüş, bu bağlamda farklılığın katılımcılardan ve kullanılan yaratıcılık testinden kaynaklanabileceği düşünülmüştür. Ayrıca söz konusu farklılığın ölçüm aracından mı yoksa katılımcılardan mı kaynaklandığı noktası belirsizlik taşımaktadır.

Yaratıcılık alanında yapılan ve cinsiyet değişkeninin incelendiği çalışmalardan elde edilen sonuçların tutarsız oldukları söylenebilir. Baer ve Kaufman (2008) yaratıcılık çalışmalarında cinsiyet değişkeninin incelendiği 70'in üzerinde araştırmaya ulaşmışlar; bu çalışmalardan 35'inde kadın ve erkekler arasında bir farkın olmadığını, 4 araştırmada erkeklerin daha yüksek puanlar aldıklarını, 9 çalışmada kadınların daha yüksek puan aldıklarını ve 30 çalışmada ise kadınlar ve erkeklerin puanlarına ilişkin çelişkili sonuçların alındığını belirtmişlerdir. Derleme çalışmada kültüre dayalı bilimsel yaratıcılık testlerinin kullanıldığı 2 çalışmaya yer verilmiş (Sansanwal ve Sharma, 1993; Shukla ve Sharma, 1986), bu çalışmalarda kadın ve erkekler arasında bir farka rastlanmadığı rapor edilmiştir (Baer ve Kaufman, 2008). Çelişkili sonuçların alındığı çalışmalarda kullanılan testlerin büyük bir çoğunluğunun genel yaratıcılığı ölçmeyi hedefleyen çoğul düşünme testleri ve türevlerinin oldukları söylenebilir.

Söz konusu araştırmalar incelendiğinde yaratıcılıkta cinsiyet konusuna ilişkin ciddi bir külliyat olmasına rağmen net bir görüş birliğinin sağlanamadığı açıktır. Çalışmalarda kullanılan ölçüm araçlarının belirli bir faktör yapısına sahip olmalarına rağmen grup farklılıklarının ortalamalar üzerinden test edilmeleri bir sınırlılık olarak gösterilebilir. ÇABİYAT-R'nin madde yanlılık analizleri ise ölçme değişmezliği kapsamında ele alınmıştır. Madde yanlılığı için elde edilen sonuçların bilimsel yaratıcılık alanında yapılan çalışmalara kıyasla daha güçlü oldukları söylenebilir.

5.3 ÇABİYAT-R'nin Güvenirliğine İlişkin Tartışma ve Sonuç

Bu bölümde ÇABİYAT-R'nin iç tutarlılık ve kodlayıcılar arası güvenilirlik analizlerden elde edilen bulgular tartışılmıştır.

5.3.1 ÇABİYAT-R'nin iç tutarlılığına ilişkin tartışma ve sonuç

ÇABİYAT-R'nin alt testler bağlamında Cronbach Alpha katsayıları kanıt değerlendirme $\alpha = .765$, deney tasarlama $\alpha = .779$, hipotez geliştirme $\alpha = .770$ ve test toplam Cronbach Alpha katsayısı ise $\alpha = .83$ olarak bulunmuştur. Madde silindiğinde hesaplanan Cronbach Alpha katsayıları incelendiğinde (Tablo 4.19) alt testler ve toplam testin güvenilirlik katsayılarında bir yükselme olmadığı görülmekte ve maddelerin bu hali ile optimum iç tutarlılığa ulaştıkları düşünülmektedir. Elde edilen katsayıların literatürde kabul gören .70'in üzerinde olması (Cortina, 1993; Nunnally ve Bernstein, 1994) sebebiyle ÇABİYAT-R'nin iç tutarlılık güvenirliliğinin sağlandığı söylenebilir. Bilimsel yaratıcılık çalışmalarında elde edilen Cronbach Alpha katsayılarının mevcut çalışmada elde edilen değerlere yakın oldukları ve sonuçlar ile benzerlik gösterdikleri söylenebilir (Ateşgöz ve Sak, 2021; Ayas ve Sak, 2014; Chin ve Siew, 2015; Hu ve Adey, 2002, Kanlı, 2014, Mohamed, 2006). Diğer taraftan kimi araştırmaların alt boyutlarından elde edilen Cronbach Alpha katsayılarının .521 ile .681 (Xu vd., 2021); .68 (Siew vd., 2014); .433 ile .584 (Ayas, 2017) düzeylerinde kaldıkları ve ÇABİYAT-R'nin söz konusu çalışmalara kıyasla iç tutarlılığının oldukça iyi olduğu görülmüştür.

Ateşgöz ve Sak (2021) çalışmalarında 801 kişiden elde ettikleri akıcılık puan türü için hipotez geliştirme, deney tasarlama alt testleri ve toplam test Cronbach Alpha katsayılarını sırasıyla .84, .84 ve .88 olarak hesaplamışlardır. Elde edilen bulguların mevcut çalışma bulguları ile benzerlik göstermelerine rağmen araştırmacıların

raporladıkları Cronbach Alpha katsayılarının görece bu çalışmadan büyük oldukları söylenebilir. Bu durumunun nedeninin örneklem sayısı olduğu düşünülmektedir. Cronbach Alpha katsayısının örneklem sayısı (Bonett, 2002; Charter, 2003; Churchill ve Peter, 1984; Duhachek vd., 2005) ve madde sayısına (Churchill ve Peter, 1984; Peterson, 1994) bağlı bir katsayı olduğuna ilişkin araştırma bulguları bulunmaktadır. Ayas (2017) BÜT 3-5 formunda tüm puan türleri için hipotez geliştirme alt testinde .433 ile .584 arasında; deney tasarlama alt testinde .813 ile .828 arasında ve toplam testte .701 ile .743 arasında değişen Cronbach Alpha Katsayıları rapor etmiştir. Çalışmada her bir alt ölçek için 2’şer maddenin bulunması iç tutarlılık bulgularındaki zayıflığın nedeni olarak gösterilebilir.

Mevcut çalışmada Cronbach Alpha katsayılarına ek kanıt sunmak amacıyla McDonald’s Omega ve Stratified Alpha katsayıları da rapor edilmiştir. McDonald’s Omega katsayısının madde faktör yüklemeleri ve benzersizlik (uniqueness) değerleri kullanılarak hesaplanmasından dolayı (McDonald, 1999) özellikle çok boyutlu yapılarda Cronbach Alpha’ya göre daha iyi bir bileşik puan tahmini sağladığı (Graham, 2006; Zinbarg vd., 2005) ve bu yüzden raporlanması önerilmektedir (Chen vd., 2012). McDonald’s Omega katsayıları kanıt değerlendirme, deney tasarlama, hipotez geliştirme ve toplam test için sırasıyla $\omega=.765$, $\omega=.782$, $\omega=.771$ ve $\omega=.840$ olarak hesaplanmıştır. Elde edilen McDonald’s Omega katsayılarının kabul edilebilir .70 düzeyinin üzerinde oldukları görülmüştür (Kline, 2013, McNeish, 2017). Çok boyutlu ve alt ölçeklere sahip yapılarda raporlanması önerilen (Feldt ve Brennan, 1989) bir diğer güvenilirlik katsayısının Stratified Alpha olduğu söylenebilir. Çalışma kapsamında Stratified Alpha katsayısı $\alpha_{\text{strat}}=.875$ olarak hesaplanmış ve bu değer yüksek güvenilirliğe işaret ettiği görülmüştür (Cortina, 1993; DeVellis ve Thorpe, 2022; Nunnally ve Bernstein, 1994).

Güvenirlik analizlerinde Cronbach Alpha katsayısının yukarıda sözü edilen kimi sınırlılıklardan dolayı güvenilirlik analizlerinde düzeltilmiş madde korelasyon katsayıları, maddeler arası ve madde toplam korelasyonlarının da incelenmesi önerilmektedir (Clark ve Watson, 2019; Cortina, 1993). Düzeltilmiş madde toplam korelasyon katsayılarının .30 üzeri (Aiken, 1997), maddeler arası korelasyon katsayılarının .15 ile .50 arasında değişimleri (Briggs ve Cheek, 1986; Clark ve Watson, 1995), maddeler arası ortalama korelasyon değerinin .40 civarında olması

(Clark ve Watson, 2019) beklenmektedir. Tablo 4.18 incelendiğinde ÇABİYAT-R testinin maddeler arası, maddeler arası ortalama ve düzeltilmiş madde toplam korelasyon katsayılarının iç tutarlılığa ilişkin yeterli kanıt sundukları söylenebilir. İncelenen diğer çalışmalarda Ateşgöz (2020) .47 ile .72; Ayas (2017) .277 ile .708; Hu ve Adey (2002) .576 ile .766; Chin ve Siew (2015) .398 ile .753; Kanlı (2014) .370 ile .582 arasında değişen düzeltilmiş madde toplam korelasyonları hesaplandığı ve mevcut çalışma bulgularına benzer sonuçlar alındığı görülmüştür.

5.3.2 Kodlayıcılar arası güvenilirliğe ilişkin tartışma ve sonuç

ÇABİYAT-R güvenilirlik analizleri için kodlayıcılar arası güvenilirlik analizlerine başvurulmuş ve iki uzman arasındaki sınıf içi korelasyon katsayı değerleri .979 ile 1.00 arasında hesaplanmıştır. Şeker, Terleme, Tünel, Suluk ve Gemi maddelerinde ise iki uzman arasında tam uyumun sağlandığı görülmüştür. Elde edilen katsayı değerlerinin bilimsel yaratıcılık alanında yapılan kimi çalışmalar ile benzer (Aschauer vd., 2022; Ateşgöz ve Sak, 2021; Ayas, 2017) kimi çalışmalardan ise görece daha yüksek oldukları söylenebilir (Bermejo vd., 2016; Chin ve Siew, 2015; Hu ve Adey, 2002; Xu vd., 2024).

Hu ve Adey (2002) .793 ile .913; Chin ve Siew (2015) .780 ile .933; Bermejo vd. (2016) .793 ile .988; Xu vd. (2024) .814 ile .919 arasında değişen puanlayıcılar arası güvenilirlik katsayıları raporlamışlardır. ÇABİYAT-R alt testlerinde yer alan maddelerin tamamının doğru yanıt havuzu bulunmaktadır. Söz konusu doğru yanıt havuzu öncül araştırmaya (Ateşgöz, 2018; Ateşgöz ve Sak, 2021), alan uzmanlarının görüşlerine, mevcut araştırma için yapılan alanyazın okumalarına ve ön değerlendirmelerden elde edilen verilere dayanmaktadır. Bu kapsamda öğrencilerden gelen yanıtlar, doğru ya da yanlış olacak şekilde uygulayıcı tarafından işaretlenerek kısa bir sürede değerlendirilebilmektedir. Böylelikle farklı puanlayıcılar arasında oluşabilecek sistematik ya da metodolojik bir hatanın önüne geçildiği düşünülmektedir.

Geliştirilen bilimsel yaratıcılık testlerinin araştırmacıları tarafından her ne kadar alana özgü maddelerden oluştukları iddia edilse de maddeler incelendiğinde alan bilgisi özelinde sınırlılıklara sahip oldukları düşünülmektedir. Örneğin Xu ve arkadaşlarının (2024) çalışmalarında kullandıkları maddelerden biri “*plastik bir şişenin bilimsel alanda kullanılması*” ile ilgilidir. Belirtilen madde bazında çok

sayıda yanıt üretilebileceği açıktır. Fakat gelen yanıtlar kimi puanlayıcılar için doğru kabul edilebilecekken kimileri için yanlış olarak kabul edilebilir. Hatta gelebilecek yanıtların gerçekten bilimsel bir başlıkta değerlendirilip değerlendirilemeyeceği dâhi şüphelidir. Diğer taraftan ÇABİYAT-R maddeleri doğrudan fen bilimleri alanında fizik, kimya ve biyoloji disiplinlerine özgü alan bilgilerini barındırmaktadır. Bu bakımdan ÇABİYAT-R testinde yer alan maddelere verilebilecek yanıtların bilimsel doğruluklarının diğer bilimsel yaratıcılık test maddelerine kıyasla daha keskin oldukları söylenebilir. Sonuç olarak iç tutarlılık ve kodlayıcılar arası güvenilirlik analizlerinden elde edilen bulgular kapsamında ÇABİYAT-R'nin güvenilir bir test olduğu düşünülmektedir.

5.4 Öneriler

Bu bölümde ÇABİYAT-R'den elde edilen bulgular ışığında eğitimsel uygulamalara ve ileride yapılabilecek araştırmalara yönelik birtakım önerilere yer verilmiştir.

5.4.1 Eğitim uygulamalarına yönelik öneriler

1. Hedef grubu öğrencilerinin bilimsel yaratıcılıkları ve fen bilimleri alanındaki yetenekleri ölçülerek öğrenciler için uygun eğitim programları tasarlanabilir.
2. ÇABİYAT-R, Fen bilimleri alanında üstün yetenekli öğrencilerin tanınmasında farklı yetenek testleri ile kullanılabilir.
3. Fen bilimlerine dayalı bir öğretim programının etkililiği için ön test ve son test kapsamında ÇABİYAT-R'den yararlanılabilir.

5.4.2 İleriki araştırmalara yönelik öneriler

1. ÇABİYAT-R testi için yalnızca akıcılık puanı hesaplanmıştır. Alt testlerde yer alan her bir madde için alan uzmanlarının görüşleri alınarak “uygunluk” ve “alan bilgisi” gibi teste yeni puanlar eklenebilir.
2. ÇABİYAT-R çoğul düşünmeye dayalı olarak geliştirilmiştir. ÇABİYAT-R kuramsal alt yapısı temel alınarak her bir alt teste tekil düşünme becerisinin de ölçülebildiği maddeler eklenebilir. Böylelikle oluşacak yeni model için hem bilimsel süreç hem de yaratıcılık becerileri kapsamında çok daha kapsamlı bir yapı elde edilebilir.
3. ÇABİYAT-R testinin revizyonu için yeterli örneklem büyüklüğüne ulaşılmıştır. Fakat daha büyük örneklem grupları ile yapının tekrar test edilmesi ve genellenebilir sonuçlara ulaşılması önerilebilir.
4. ÇABİYAT-R kuramsal alt yapısı kullanılarak lise öğrencilerine yönelik bilimsel yaratıcılık testleri geliştirilebilir.

5. ÇABİYAT-R ve farklı teorik alt yapı ile geliştirilen bilimsel yaratıcılık testlerinin model karşılaştırmaları sağlanabilir.
6. ÇABİYAT-R'nin uyum geçerliği için fen bilimleri alanında akademik başarıyı ölçen bir bataryadan yararlanılmıştır. Bilimin doğası (NOS) üzerine geliştirilmiş farklı envanterler ile ÇABİYAT-R'nin uyum geçerliği tekrar test edilebilir.
7. ÇABİYAT-R ve zekâ testlerinin birlikte kullanıldığı araştırmalar yapılarak genel zeka ile ÇABİYAT-R puanları arasındaki ilişki incelenebilir.
8. ÇABİYAT-R'nin farklı yetenek alanlarındaki testler ile uyum geçerliği test edilebilir. Örneğin matematik alanında geliştirilmiş bir test ile ÇABİYAT-R arasındaki ilişkiler incelenebilir.
9. ÇABİYAT-R alana özgü geliştirilmiş bir yaratıcılık testidir. Genel yaratıcılığa ilişkin bir test ile olan ilişkisi incelenebilir.
10. ÇABİYAT-R testini alan öğrencilerin düşünme süreçlerini anlamaya yönelik nitel çalışmalar gerçekleştirilebilir. Böylelikle erken çocukluk dönemi bilimsel süreç becerilerinin mekanik işleyişi hakkında bilgi edinilebilir.
11. Literatürde kanıt değerlendirme becerisinin küçük yaş gruplarında ancak mükemmel uyum ya da uyumsuzluk olması durumunda ölçülebileceği iddia edilmektedir. ÇABİYAT-R'den elde edilen sonuçlar bu varsayımın aksini göstermektedir. Bu bağlamda ÇABİYAT-R kanıt değerlendirme becerisinin temel alındığı kanıt üretme görev tipinin farklı çalışmalarda da kullanılması önerilebilir.
12. ÇABİYAT-R'nin norm çalışması yapılarak standart bir ölçüm aracı haline getirebilir.
13. ÇABİYAT-R güvenirlik analizleri kapsamında test tekrar test analizlerinin yapılması önerilebilir.

KAYNAKÇA

- Abraham A. (2016). Gender and creativity: an overview of psychological and neuroscientific literature. *Brain imaging and behavior*, 10(2), 609–618. <https://doi.org/10.1007/s11682-015-9410-8>
- Ai, X. (1999). Creativity and academic achievement: An investigation of gender differences. *Creativity Research Journal*, 12(4), 329–337. https://doi.org/10.1207/s15326934crj1204_11
- Aiken, L. R. (2009). *Psychological testing and assessment*. Pearson Education India.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716–723.
- Alabbasi, A. M., Thompson, T. L., Runco, M. A., Alansari, L. A., & Ayoub, A. E. A. (2025). Gender differences in creative potential: A meta-analysis of mean differences and variability. *Psychology of Aesthetics, Creativity, and the Arts*, 19(1), 87–100. <https://doi.org/10.1037/aca0000506>
- Alacapınar, F. G. (2013). Grade level and creativity. *Eğitim Araştırmaları-Eurasian Journal of Educational Research*. 50, 247-266.
- Alamer, A. (2022). Exploratory structural equation modeling (ESEM) and bifactor ESEM for construct validation purposes: Guidelines and applied example. *Research Methods in Applied Linguistics*, 1(1), 100005. <https://doi.org/10.1016/j.rmal.2022.100005>
- Alamer, A., & Marsh, H. (2022). Exploratory structural equation modeling in second language research: An applied example using the dualistic model of passion. *Studies in Second Language Acquisition*, 44(5), 1477-1500. <https://doi.org/10.1017/S0272263121000863>
- Alexander, P. A. (1992). Domain Knowledge: Evolving Themes and Emerging Concerns. *Educational Psychologist*, 27(1), 33–51. <https://doi.org/10.1207/s15326985ep2701>
- Amabile, T. M. (1982). The social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43(5), 997–1013.
- Amabile, T. M. (1983a). A consensual technique for creativity assessment. T. M. Amabile (Ed.), *The social psychology of creativity* içinde, (pp. 37-63). New York, NY: Springer.
- Amabile, T. M. (1983b). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45(2), 357–376. <https://doi.org/10.1037/0022-3514.45.2.357>
- Amabile, T. M. (1996). *Creativity in context: Update to "The Social Psychology of Creativity."* Westview Press.
- Amsel, E., & Brock, S. (1996). The development of evidence evaluation skills. *Cognitive Development*, 11, 523–550.
- An, D., Song, Y., & Carr, M. (2016). A comparison of two models of creativity: Divergent thinking and creative expert performance. *Personality and Individual Differences*, 90, 78–84. <https://doi.org/10.1016/j.paid.2015.10.040>

- Anderson, M. (1992). *Intelligence and development: A cognitive theory*. Blackwell Publishing.
- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49, 155-173. <https://doi.org/10.1007/BF02294170>
- Arreola, N., & Reiter-Palmon, R. (2016). The effect of problem construction creativity on solution creativity across multiple real-world problems. *Psychology of Aesthetics, Creativity, and the Arts*, 10, 287-295.
- Aschauer, W., Haim, K., & Weber, C. (2022). A contribution to scientific creativity: A validation study measuring divergent problem solving ability. *Creativity Research Journal*, 34(2), 195-212. <https://doi.org/10.1080/10400419.2021.1968656>
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural equation modeling: a multidisciplinary journal*, 16(3), 397-438. <https://doi.org/10.1080/10705510903008204>
- Atesgoz, N. N., & Sak, U. (2021). Test of scientific creativity animations for children: Development and validity study. *Thinking Skills and Creativity*, 40, 100818.
- Atakaya, M. A., Sak, U., & Ayas, M. B. (2022). A Study on Psychometric Properties of Creativity Indices. *Creativity Research Journal*, 36(2), 348-364. <https://doi.org/10.1080/10400419.2022.2134550>
- Avcu, A. (2021). The Use of Exploratory Graph Analysis to Validate Trust in Relationships Scale. *International Journal of Assessment Tools in Education*, 8(3), 542-552. <https://doi.org/10.21449/ijate.831784>
- Ayas, M. B. (2017). *Bilimsel üretkenlik testinin 3, 4 ve 5. Sınıf öğrencilerine uygun formunun geliştirilmesi ve ön psikometrik özelliklerinin incelenmesi*. Yayınlanmamış Doktora Tezi. Eskişehir: Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü.
- Ayas, M. B. and Sak, U. (2014). Objective measure of scientific creativity: Psychometric validity of the Creative Scientific Ability Test. *Thinking Skills and Creativity*, 13, 195-205.
- Ayverdi, L., Asker, E., Aydın, S. Ö., & Sarıtaş, T. (2012). İlköğretim öğrencilerinin bilimsel yaratıcılıkları ile fen ve teknoloji dersi akademik başarıları arasındaki ilişkinin belirlenmesi. *İlköğretim online*, 11(3), 646-659.
- Azarpira, N., Amini, M., Kojuri, J. et al. Assessment of scientific thinking in basic science in the Iranian second national Olympiad. *BMC Res Notes* 5, 61 (2012). <https://doi.org/10.1186/1756-0500-5-61>
- Baer, J. (1991). Generality of creativity across performance domains. *Creativity Research Journal*, 4(4), 23-39.
- Baer, J. (1993). *Creativity and divergent thinking: A task-specific approach*. Hillsdale, NJ: Erlbaum.
- Baer, J. (1994). Divergent thinking is not a general trait: A multidomain training experiment. *Creativity Research Journal*, 7(1), 35-46. <https://doi.org/10.1080/1040041940953450>

- Baer, J. (1998). The case for domain specificity of creativity. *Creativity Research Journal*, 11(2), 173-177.
- Baer, J. (1999). Domains of Creativity. In M. A. Runco and S. R. Pritzker (Eds.), *Encyclopedia of Creativity* (Vol 1) (pp. 591-596). Academic Press
- Baer, J. (2003). The impact of the core knowledge curriculum on creativity. *Creativity Research Journal*, 15, 297-300. doi:10.1080/10400419.2003.9651422
- Baer, J. (2015). *Domain specificity of creativity*. Academic Press.
- Baer, J., & Kaufman, J. C. (2005). Bridging generality and specificity: *The amusement park theoretical (APT) model of creativity*. *Roeper Review*, 27(3), 158-163. <https://doi.org/10.1080/02783190509554310>
- Baer, J., & Kaufman, J. C. (2008). Gender differences in creativity. *The Journal of Creative Behavior*, 42(2), 75-105. <https://doi.org/10.1002/j.2162-6057.2008.tb01289.x>
- Baer, J. (2016). Implications of domain specificity for creativity assessment. In J. Baer (Ed.), *Domain specificity of creativity*, (pp. 103-140). Elsevier.
- Bandalos, D.L. & Webb, M.-Y. (April, 2005). Efficacy of the WLSMV Estimator for Coarsely Categorized and Nonnormally Distributed Data. American Educational Research Association, Montréal, Canada.
- Barraza, P., Preiss, D. D., & Pardo, M. (2019). The development of creativity in Chilean kindergarten and school children. *Studies in Psychology*, 40(3), 608-634. <https://doi.org/10.1080/02109395.2019.1655220>
- Bart, W. M., Hokanson, B., Sahin, I., & Abdelsamea, M. A. (2015). An investigation of the gender differences in creative thinking abilities among 8th and 11th grade students. *Thinking Skills and Creativity*, 17, 17-24. <https://doi.org/10.1016/j.tsc.2015.03.003>
- Baysal, Z., Baysal, Z. N., Kaya, N., & Üçüncü, G. (2013). İlkokul dördüncü sınıf öğrencilerinde bilimsel yaratıcılık düzeyinin çeşitli değişkenler açısından incelenmesi. *Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, 38(38), 55-64.
- Beghetto, R. A., Kaufman, J. C., & Baxter, J. (2011). Answering the unexpected questions: Exploring the relationship between students' creative self-efficacy and teacher ratings of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 5(4), 342-349. <https://doi.org/10.1037/a0022834>
- Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, 107(2), 238. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bermejo, M.-R., María-José, Ruiz-Melero, Esparza, J., Ferrando, M. and Pons, R. (2016). A new measurement of scientific creativity: The study of its psychometric properties. *Anales de Psicología*, 32(3), 652-661.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social,

- and behavioral research: a primer. *Frontiers in public health*, 6, 149. 1-18. <https://doi.org/10.3389/fpubh.2018.00149>
- Bollen, K. A. (2014). *Structural equations with latent variables*. John Wiley & Sons.
- Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics*, 27(4), 335–340. <https://doi.org/10.3102/10769986027004335>
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of personality*, 54(1), 106-148. <https://doi.org/10.1111/j.1467-6494.1986.tb00391.x>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
- Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M., & Wilson, M. (2010). A framework for analyzing scientific reasoning in assessments. *Educational Assessment*, 15, 142–174. doi:10.1080/10627197.2010.530562
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, 21(2), 230-258. <https://doi.org/10.1177/0049124192021002005>
- Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert, & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich longitudinal study* (pp. 38-54). Cambridge University Press.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Burns, J. C., Okey, J. R., & Wise, K. C. (1985). Development of an integrated process skill test: TIPS II. *Journal of Research in Science Teaching*, 22, 169–177. [https://doi:10.1002/\(ISSN\)1098-2736](https://doi:10.1002/(ISSN)1098-2736)
- Büyüköztürk, Ş., Akgün, Ö. E., Demirel, F., Karadeniz, Ş., & Çakmak, E. K. (2017). *Bilimsel araştırma yöntemleri* (23. baskı). Pegem Akademi.
- Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of personality*, 80(4), 796-846. <https://doi.org/10.1111/j.1467-6494.2011.00749.x>
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Sage.
- Camp, G. C. (1994). A longitudinal study of correlates of creativity. *Creativity Research Journal*, 7(2), 125–144. <https://doi.org/10.1080/10400419409534519>
- Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review*, 67(6), 380–400. <https://doi.org/10.1037/h0040373>
- Carey, S. (1985). *Conceptual change in childhood*. MIT press.
- Carlson, M. D., & Morrison, R. S. (2009). Study design, precision, and validity in observational studies. *Journal of palliative medicine*, 12(1), 77-82. <https://doi.org/10.1089/jpm.2008.9690>

- Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, Validity, and Factor Structure of the Creative Achievement Questionnaire. *Creativity Research Journal*, 17(1), 37–50. https://doi.org/10.1207/s15326934crj1701_4
- Chalmers, A. F. (1999). *What Is This Thing Called Science?* (3rd ed.). Open University Press.
- Charles, R. E., & Runco, M. A. (2001). Developmental trends in the evaluative and divergent thinking of children. *Creativity Research Journal*, 13(3-4), 417–437. https://doi.org/10.1207/S15326934CRJ1334_19
- Charter R. A. (2003). Study samples are too small to produce sufficiently precise reliability coefficients. *The Journal of general psychology*, 130(2), 117–129. <https://doi.org/10.1080/00221300309601280>
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245–276.
- Chae, S. (2003). Adaptation of a picture-type creativity test for pre-school children. *Language Testing*, 20(2), 179–188. <https://doi.org/10.1191/0265532203lt251oa>
- Chang, H.-P., Chen, C.-C., Guo, G.-J., Cheng, Y.-J., Lin, C.-Y., & Jen, T.-H. (2011). The development of a competence scale for learning science: Inquiry and communication. *International Journal of Science and Mathematics Education*, 9, 1213–1233. doi:10.1007/s10763-010-9256-x
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling: a multidisciplinary journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Chen, C., Kasof, J., Himsel, A. J., Greenberger, E., Dong, Q., & Xue, G. (2002). Creativity in drawings of geometric shapes: A cross-cultural examination with the consensual assessment technique. *Journal of Cross-Cultural Psychology*, 33(2), 171–187. <https://doi.org/10.1177/0022022102033002004>
- Chen, C., Himsel, A., Kasof, J., Greenberger, E., & Dmitrieva, J. (2006). Boundless creativity: Evidence for the domain generality of individual differences in creativity. *The Journal of Creative Behavior*, 40(3), 179–199. <https://doi.org/10.1002/j.2162-6057.2006.tb01272.x>
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, 80(1), 219–251. <https://doi.org/10.1111/j.1467-6494.2011.00739.x>
- Chen, Z., & Klahr, D. (1999). All other things being equal: acquisition and transfer of the control of variables strategy. *Child development*, 70(5), 1098–1120. <https://doi.org/10.1111/1467-8624.00081>
- Cheng, P. C.-H., & Simon, H. A. (1992). The right representation for discovery: Finding the conservation of momentum. In D. Sleeman & P. Edwards (Eds.), *Machine learning: Proceedings of the ninth international workshop* (vol. ML92, pp. 62–77). Morgan Kaufman.

- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5
- Cheung, P. C., & Lau, S. (2013). A Tale of Two Generations: Creativity Growth and Gender Differences Over a Period of Education and Curriculum Reforms. *Creativity Research Journal*, 25(4), 463-471. <https://doi.org/10.1080/10400419.2013.843916>
- Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, 35(6), 623-654. [https://doi.org/10.1002/\(SICI\)1098-2736\(199808\)35:6<623::AID-TEA3>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1098-2736(199808)35:6<623::AID-TEA3>3.0.CO;2-O)
- Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, 94(2), 327-343. <https://doi.org/10.1037/0022-0663.94.2.327>
- Chin, M. K., and Siew, N. M. (2015). The development and validation of a figural scientific creativity test for preschool pupils. *Creative Education*, 6(12), 1391-1402.
- Christensen, A. P., & Golino, H. (2019). Estimating the stability of psychological dimensions via Bootstrap Exploratory Graph Analysis: A Monte Carlo simulation and tutorial. <https://doi.org/10.3390/psych3030032>
- Christensen, A. P., & Golino, H. (2021). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A Monte Carlo simulation and tutorial. *Psych*, 3(3), 479-500. <https://doi.org/10.3390/psych3030032>
- Churchill, G. A., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research*, 21(4), 360-375. <https://doi.org/10.2307/3151463>
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological assessment*, 31(12), 1412-1427. <https://doi.org/10.1037/pas0000626>
- Claxton, A. F., Pannells, T. C., & Rhoads, P. A. (2005). Developmental Trends in the Creativity of School-Age Children. *Creativity Research Journal*, 17(4), 327-335. https://doi.org/10.1207/s15326934crj1704_4
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge. <https://doi.org/10.4324/9780203774441>
- Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of consulting and clinical psychology*, 56(5), 754. <https://doi.org/10.1037/0022-006X.56.5.754>

- Cook, R. D. (1977). Detection of Influential Observations in Linear Regression. *Technometrics*, 19(1), 15-18. <https://doi.org/10.1080/00401706.1977.10489>
- Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: spontaneous experiments in preschoolers' exploratory play. *Cognition*, 120(3), 341-349. <https://doi.org/10.1016/j.cognition.2011.03.003>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1), 98-104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Cosemans, T., Rosseel, Y., & Gelper, S. (2022). Exploratory graph analysis for factor retention: Simulation results for continuous and binary data. *Educational and psychological measurement*, 82(5), 880-910. <https://doi.org/10.1177/00131644211059089>
- Costello, A. B., & Osborne, J. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis. *Practical Assessment Research & Evaluation*, 10, 1-9.
- Cramond, B., Matthews-Morgan, J., Torrance, E. P., & Zuo, L. (1999). Why should the *Torrance Tests of Creative Thinking* be used to assess creativity? *Korean Journal of Thinking & Problem Solving*, 9(2), 77-101.
- Cramond, B., Matthews-Morgan, J., Bandalos, D., & Zuo, L. (2005). A Report on the 40-Year Follow-Up of the Torrance Tests of Creative Thinking: Alive and Well in the New Millennium. *Gifted Child Quarterly*, 49(4), 283-291. <https://doi.org/10.1177/001698620504900402>
- Creswell, J. W. (2013). *Research design qualitative quantitative and mixed methods approaches* (4th ed.). Thousand Oaks: Sage
- Crocker, S., & Buchanan, H. (2011). Scientific reasoning in a real-world context: The effect of prior belief and outcome on children's hypothesis-testing strategies. *British Journal of Developmental Psychology*, 29(3), 409-424. <https://doi.org/10.1348/026151010X496906>
- Cropley, A. J. (2000). Defining and Measuring Creativity: Are Creativity Tests Worth Using? *Roeper Review*, 23, 72-79. <https://doi.org/10.1080/02783190009554069>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>
- Cseh, G. M., & Jeffries, K. K. (2019). A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts*, 13(2), 159-166. <https://doi.org/10.1037/aca0000220>
- Csikszentmihalyi, M. (1999). Implications of a system perspective for the study of creativity. In R. Sternberg (Ed.), *Handbook of creativity* (pp. 313-328). Cambridge University Press.
- Csikszentmihalyi, M. (2014). *The systems model of creativity the collected works of Mihaly*. Springer.

- Cuevas, P., Lee, O., Hart, J., & Deaktor, R. (2005). Improving science inquiry with elementary students of diverse backgrounds. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 42(3), 337-357. <https://doi.org/10.1002/tea.20053>
- Çeliker, H. D., Tokcan, A., & Korkubilmez, S. (2015). Fen öğrenmeye yönelik motivasyon bilimsel yaratıcılığı etkiler mi?/does motivation toward science learning affect the scientific creativity?. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 12(30), 167-192.
- De Beer, L.T. & Van Zyl, L.E. (2019). ESEM code generator for Mplus. Retrieved from <http://www.surveymhost.co.za/esem/> doi: 10.6084/m9.figshare.8320250
- Desboulets, L. D. D. (2018). A review on variable selection in regression analysis. *Econometrics*, 6(4), 45. <https://doi.org/10.3390/econometrics6040045>
- Díaz, C., Dorner, B., Hussmann, H., & Strijbos, J. W. (2023). Conceptual review on scientific reasoning and scientific thinking. *Current Psychology*, 42, 4313–4325. <https://doi.org/10.1007/s12144-021-01786-5>
- Dikici, A. (2014). Relationships between thinking styles and behaviors fostering creativity: An exploratory study for the mediating role of certain demographic traits. *Educational Sciences: Theory and Practice*, 14(1), 179-201.
- Dudek, S. Z., Strobel, M. G., & Runco, M. A. (1993). Cumulative and proximal influences on the social environment and children's creative potential. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 154(4), 487–499. <https://doi.org/10.1080/00221325.1993.9914747>
- Duhachek, A., Coughlan, A. T., & Iacobucci, D. (2005). Results on the Standard Error of the Coefficient Alpha Index of Reliability. *Marketing Science*, 24(2), 294–301. <http://www.jstor.org/stable/40056958>
- Dumas, D., Organisciak, P., Maio, S., & Doherty, M. (2021). Four text-mining methods for measuring elaboration. *The Journal of Creative Behavior*, 55(2), 517–531. <https://doi.org/10.1002/jocb.471>
- Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science*, 17, 397–434.
- Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 365–395). The MIT Press.
- Dunbar, K., & Baker, L. M. (1994). Goals, analogy, and the social constraints of scientific discovery. *Behavioral and Brain Sciences*, 17(3), 538–539. doi:10.1017/S0140525X00035779
- Dunn, O. J. (1961). "Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293), 52-64. doi:10.1080/01621459.1961.10482090
- Edelsbrunner, P. A. (2017). *Domain-general and domain-specific scientific thinking in childhood: measurement and educational interplay* (Doctoral dissertation, ETH Zurich).

- Eichenberger, R. J. (1978). Creativity measurement through use of judgment criteria in physics. *Educational and Psychological Measurement*, 38(2), 421–427. <https://doi.org/10.1177/001316447803800224>
- Eid, M., & Kutscher, T. (2014). Statistical models for analyzing stability and change in happiness. In K. M. Sheldon & R. E. Lucas (Eds.), *Stability of happiness: Theories and evidence on whether happiness can change* (pp. 261–297). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-411478-4.00013-8>
- Einstein, A., & Infeld, L. (1938). *The evolution of physics*. New York, NY: Simon & Schuster.
- Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, 48(4), 1–18. <https://doi.org/10.18637/jss.v048.i04>
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, 23(4), 617–634. <https://doi.org/10.1037/met0000167>
- Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, 82, 904–927. <http://dx.doi.org/10.1007/s11336-017-9557-x>
- Fairchild, J., & Hunter, S. T. (2014). “We’ve got creative differences”: The effects of task conflict and participative safety on team creative performance. *The Journal of Creative Behavior*, 48(1), 64–87. <https://doi.org/10.1002/jocb.41>
- Feist, G. J. (1994). Personality and working style predictors of integrative complexity: A study of scientists' thinking about research and teaching. *Journal of Personality and Social Psychology*, 67(3), 474–484. <https://doi.org/10.1037/0022-3514.67.3.474>
- Feist, G. J. (2011). Creativity in science. In S. R. Pritzker, & M. A. Runco (Eds.), *Encyclopedia of creativity* (pp. 269–302). Elsevier.
- Feista, G. J. and Barron, F. X. (2003). Predicting creativity from early to late adulthood: Intellect, potential, and personality. *Journal of Research in Personality*, 3(37), 62–88.
- Friedlander, M. (1983). A natural science creativity test as a prediction of creative thinking in science [Abstract]. *Creative Child & Adult Quarterly*, 8(4), 211–215.
- Feldhusen, J. F. (2002). Creativity: The knowledge base and children. *High Ability Studies*, 13(2), 179–183. <https://doi.org/10.1080/1359813022000048806>
- Feldhusen, J. F. (2006). The Role of the Knowledge Base in Creative Thinking. In J. C. Kaufman & J. Baer (Eds.), *Creativity and reason in cognitive development* (pp. 137–144). Cambridge University Press. <https://doi.org/10.1017/CBO9780511606915.009>
- Feldman, D.H. (1999). The development of creativity. In R.J. Sternberg (Ed.), *Handbook of Creativity* (pp. 169–186). Cambridge University Press.
- Finney, S. J., Distefano, C., & Kopp, J. P. (2016). Overview of estimation methods and preconditions for their application with structural equation modeling. In K. Schweizer & C. Distefano (Eds.), *Principles and methods of test construction: Standards and recent advances* (pp. 135–165). Hogrefe.

- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S., Fischer, M., Strijbos, J.-W., Heene, M., & Eberle, J. (2014). Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research*, 2(3), 28–45. <https://doi.org/10.14786/flr.v2i2.96>
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286–299. <https://doi.org/10.1037/1040-3590.7.3.286>
- Fraenkel, J. R., Wallen, N. E. and Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.
- Friedlander, M. (1983). A natural science creativity test as a prediction of creative thinking in science. *Creative Child & Adult Quarterly*, 8(4), 211–215.
- Friedman, J.H., Hastie, T., & Tibshirani, R. (2014). Glasso: graphical lasso- estimation of Gaussian graphical models. R package version 1.8. Retrieved from.
- Furnham, A. and Bachtar, V. (2008). Personality and intelligence as predictors of creativity. *Personality and Individual Differences*, 45, 613–617.
- Furnham, A. F. (1999). Personality and creativity. *Perceptual and Motor Skills*, 8, 407-408.
- Gardner, H. (1982). *Art, mind, and brain: A cognitive approach to creativity*. Basic Books.
- Garrido, L. E., Barrada, J. R., Aguasvivas, J. A., Martínez-Molina, A., Arias, V. B., Golino, H. F., Legaz, E., Ferrís, G., & Rojo-Moreno, L. (2020). Is small still beautiful for the strengths and difficulties questionnaire? Novel findings using exploratory structural equation modeling. *Assessment*, 27(6), 1349–1367. <https://doi.org/10.1177/1073191118780461>
- Getzels, J. W., & Csikszentmihalyi, M. (1967). Scientific creativity. *Science Journal*, 3(9), 80–84.
- Gijlers, H., & de Jong, T. (2005). The Relation between Prior Knowledge and Students' Collaborative Discovery Learning Processes. *Journal of Research in Science Teaching*, 42(3), 264–282. <https://doi.org/10.1002/tea.20056>
- Guilford, J. P. (1967). Creativity, yesterday, today and tomorrow. *The Journal of Creative Behavior*, 1(1), 1-14.
- Golino HF, Epskamp S (2017) Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE* 12(6): e0174035. <https://doi.org/10.1371/journal.pone.0174035>
- Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, 25(3), 292–320. <https://doi.org/10.1037/met0000255>
- Golino, H., Lillard, A. S., Becker, I., & Christensen, A. P. (2021). Investigating the structure of the Children's Concentration and Empathy Scale using exploratory

- graph analysis. *Psychological Test Adaptation and Development*. 35-49. <https://doi.org/10.1027/2698-1866/a000008>
- Gooding, D. (1990). *Experiment and the making of meaning*. Nijhoff/Kluwer.
- Goodwin, L. D. (1999). The Role of Factor Analysis in the Estimation of Construct Validity. *Measurement in Physical Education and Exercise Science*, 3(2), 85–100. https://doi.org/10.1207/s15327841mpee0302_2
- Graham, J. M. (2006). Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability: What They Are and How to Use Them. *Educational and Psychological Measurement*, 66(6), 930–944. <https://doi.org/10.1177/0013164406288165>
- Gralewski, J., & Karwowski, M. (2013). Polite girls and creative boys? Students' gender moderates accuracy of teachers' ratings of creativity. *The Journal of Creative Behavior*, 47(4), 290–304. <https://doi.org/10.1002/jocb.36>
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265–275. <https://doi.org/10.1037/0033-2909.103.2.265>
- Guay, F., Morin, A. J. S., Litalien, D., Valois, P., & Vallerand, R. J. (2014). Application of Exploratory Structural Equation Modeling to Evaluate the Academic Motivation Scale. *The Journal of Experimental Education*, 83(1), 51–82. <https://doi.org/10.1080/00220973.2013.876231>
- Guilford, J. P. (1950). Creativity. *American Psychologist*, 5(9), 444–454. <https://doi.org/10.1037/h0063487>
- Guilford, J. P. (1968). *Intelligence, creativity, and their educational implications*. Robert R. Knapp.
- Gupta, S. M. (1988). Distribution of scientific creativity ability. *Indian Journal of Psychometry and Education*, 19(1), 21-24.
- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2), 149-161.
- Gruber, H. E. (1974). *Darwin on man: A psychological study of scientific creativity*. E. P. Dutton.
- Hadzigeorgiou, Y., Fokialis, P., & Kabouropoulou, M. (2012). Thinking about creativity in science education. *Creative Education*, 3(05), 603. <http://dx.doi.org/10.4236/ce.2012.350>
- Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2019) *Multivariate Data Analysis*. 8th Edition, Cengage Learning.
- Han, K. S. (2003). Domain-specificity of creativity in young children: How quantitative and qualitative data support it. *The Journal of Creative Behavior*, 37, 117-142.
- Hart, J. (2013). *The Art of the Storyboard: A filmmaker's introduction*. Routledge.
- Hayes, J. R. (1989). Cognitive processes in creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 135–145). Plenum Press.

- Hayes, A. F., & Coutts, J. J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But.... *Communication Methods and Measures*, 14(1), 1–24. <https://doi.org/10.1080/19312458.2020.1718629>
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238–247. <https://doi.org/10.1037/1040-3590.7.3.238>
- He, W.-j., & Wong, W.-c. (2011). Gender differences in creative thinking revisited: Findings from analysis of variability. *Personality and Individual Differences*, 51(7), 807–811. <https://doi.org/10.1016/j.paid.2011.06.027>
- Hennessey, B. A. and Amabile, T. M. (1999). Consensual assessment. In M. A. Runco, and S. R. Pritzker (Eds.), *Encyclopedia of Creativity* (Vol 1) (pp. 347-359). San Diego, CA: Academic Press.
- Heller, K. A. (2007). Scientific ability and creativity. *High Ability Studies*, 18(2), 209–234.
- Hennessey, B. A. and Amabile, T. M. (1999). Consensual assessment. In M. A. Runco, and S. R. Pritzker (Eds.), *Encyclopedia of Creativity* (Vol 1) (pp. 347-359). Academic Press.
- Hickey, M. (2001). An application of Amabile's consensual assessment technique for rating the creativity of children's musical compositions. *Journal of Research in Music Education*, 49(3), 234–244. <https://doi.org/10.2307/3345709>
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21(5), 967–988. <https://doi.org/10.1177/014920639502100509>
- Hinkin, T. R. (1998). A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organizational Research Methods*, 1(1), 104-121. <https://doi.org/10.1177/109442819800100106>
- Hocevar, D. (1976). Dimensionality of creativity. *Psychological Reports*, 39(3), 869-870.
- Hodson, D., & Reid, D. (1988). Science for all-motives, meaning and implications. *The School Science Review*, 69(249), 653–661.
- Holden, G. W., Gower, T., & Chmielewski, M. (2020). Methodological considerations in ACEs research. In G. J. G. Asmundson & T. O. Afifi (Eds.), *Adverse childhood experiences: Using evidence to advance research, practice, policy, and prevention* (pp. 161–182). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-816065-7.00009-4>
- Hora, S., Badura, K. L., Lemoine, G. J., & Grijalva, E. (2022). A meta-analytic examination of the gender difference in creative performance. *Journal of Applied Psychology*, 107(11), 1926–1950. <https://doi.org/10.1037/apl0000999>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185. <https://doi.org/10.1007/BF02289447>
- Hu, W., and Adey, P. A. (2002). Scientific creativity test for secondary school students. *International Journal of Science Education*, 24(4), 389-403.

- Hu, W., Adey, P., Shen, J., Lin, C. (2004). The comparisons of the development of creativity between English and Chinese adolescents. *Acta Psychologica Sinica*, 36(6), 718–731.
- Hu, W. P. and Han, Q. (2006). A developmental study of creative scientific problem finding in elementary students. *Psychological Science*, 29, 944–946.
- Hu, W., Shi, Q. Z., Han, Q., Wang, X., & Adey, P. (2010). Creative scientific problem finding and its developmental trend. *Creativity Research Journal*, 22(1), 46–52.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hu, W., Wu, B., Jia, X., Yi, X., Duan, C., Meyer, W., & Kaufman, J. C. (2013). Increasing students' scientific creativity: The “learn to think” intervention program. *The journal of creative behavior*, 47(1), 3–21.
- Huang, C. F., & Wang, K. C. (2019). Comparative Analysis of Different Creativity Tests for the Prediction of Students’ Scientific Creativity. *Creativity Research Journal*, 31(4), 443–447. <https://doi.org/10.1080/10400419.2019.1684116>
- Huang, P., Pen, S., Chen, H., Tseng, L., & Hsu, L. (2017). The relative influences of domain knowledge and domain-general diver-gent thinking on scientific creativity and mathematical creativity. *Thinking Skills and Creativity*, 25,1–9. <https://doi.org/10.1016/j.tsc.2017.06.001>
- Humble, S., Dixon, P., & Mpofu, E. (2018). Factor structure of the Torrance Tests of Creative Thinking Figural Form A in Kiswahili speaking children: Multidimensionality and influences on creative behavior. *Thinking Skills and Creativity*, 27, 33–44. <https://doi.org/10.1016/j.tsc.2017.11.005>
- Hunter, J. E., Schmidt, F. L., & Rauschenberger, J. (1984). Methodological, statistical, and ethical issues in the study of bias in psychological tests. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 41–99). Plenum.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking: From childhood to adolescence*. (A. Parsons & S. Milgram, Trans.). Basic Books. <https://doi.org/10.1037/10034-000>
- Jackson, D. L. (2003). Revisiting Sample Size and Number of Parameter Estimates: Some Support for the N:q Hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(1), 128–141. https://doi.org/10.1207/S15328007SEM1001_6
- Jang, Y., and Ko, Y. (2017). Sources of scientific creativity: participant observation of a public research institute in Korea. *Journal of Open Innovation: Technology, Market, and Complexity*, 3(1), 1–16.
- Jirout, J., Zimmerman, C. (2015). Development of Science Process Skills in the Early Childhood Years. In: Cabe Trundle, K., Saçkes, M. (eds) *Research in Early Childhood Science Education*. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9505-0_7

- Jo, S. M. (2009). *A study of Korean students' creativity in science using structural equation modeling* (Doctoral dissertation, The University of Arizona).
- Johnsen, S. K., & Corn, A. L. (2001). *Screening Assessment for Gifted Elementary and Middle School Students: Examiner's Manual*. Pro-Ed.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36(4), 409-426. <https://doi.org/10.1007/BF02291366>
- Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1), 141-151. <https://doi.org/10.1177/001316446002000116>
- Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, 34(1), 111-117. <https://doi.org/10.1177/001316447403400115>
- Kaufman, J. J. and Baer, J. (2004). Sure, I'm creative—but not in mathematics!: Self reported creativity in diverse domains. *Empirical Studies of the arts*, 22(2), 143-155.
- Kaufman, J. C. (2012). Counting the muses: Development of the Kaufman Domains of Creativity Scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*, 6(4), 298–308. doi:10.1037/a002975
- Kaufman, J., Plucker, J. A. and Baer, J. (2008). *Essentials of Creativity Assessment*. Wiley.
- Kaufman, J. C., and Baer, J. (2005). *Creativity across domains: Faces of the muse*. Mahwah, NJ: Lawrence Erlbaum.
- Kaufman, J. C., & Baer, J. (2008). Role of the Domain in Creative Learning in USA. In A. R. Craft, T. Cremin, & P. Burnard (Eds.), *Creative Learning 3-11 and How We Document It* (pp. 27-34). Staffordshire: Trentham Books.
- Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, 20(2), 171–178. <https://doi.org/10.1080/10400410802059929>
- Kaufman, J. C., Niu, W., Sexton, J. D., & Cole, J. C. (2010). In the eye of the beholder: Differences across ethnicity and gender in evaluating creative work. *Journal of Applied Social Psychology*, 40(2), 496–511. <https://doi.org/10.1111/j.1559-1816.2009.00584.x>
- Kaufman, J. C., Baer, J., Cropley, D. H., Reiter-Palmon, R., & Sinnett, S. (2013). Furious activity vs. understanding: How much expertise is needed to evaluate creative work? *Psychology of Aesthetics, Creativity, and the Arts*, 7(4), 332–340. <https://doi.org/10.1037/a0034809>
- Kanlı, E. (2014). *Yaratıcı bilimsel çağrışımlar testinin geliştirilmesi ve psikometrik özelliklerinin incelenmesi*. Yayınlanmamış Doktora Tezi. İstanbul: İstanbul Üniversitesi, Eğitim Bilimleri Enstitüsü
- Kanlı, E. (2017). Üstün yetenekli öğrencilerin bilimsel yaratıcılık düzeyleri, cinsiyet ve bilimsel tutumları arasındaki ilişkilerin incelenmesi. *İlköğretim Online*, 16(4), 1792-1802. <https://doi.org/10.17051/ilkonline.2017.342992>

- Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing: Principles, applications, and issues*. Wadsworth/Thomson Learning.
- Kiehn, M. T. (2003). Development of Music Creativity among Elementary School Students. *Journal of Research in Music Education*, 51(4), 278–288. <https://doi.org/10.2307/3345655>
- Kim, K. H. (2006). Can We Trust Creativity Tests? A Review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal*, 18(1), 3–14. https://doi.org/10.1207/s15326934crj1801_2
- Kim, K. H., Cramond, B., & Bandalos, D. L. (2006). The Latent Structure and Measurement Invariance of Scores on the Torrance Tests of Creative Thinking-Figural. *Educational and Psychological Measurement*, 66(3), 459–477. <https://doi.org/10.1177/0013164405282456>
- Kim, K. H. (2006). Is Creativity Unidimensional or Multidimensional? Analyses of the Torrance Tests of Creative Thinking. *Creativity Research Journal*, 18(3), 251–259. https://doi.org/10.1207/s15326934crj1803_2
- Kim, J., & Michael, W. B. (1995). The relationship of creativity measures to school achievement and to preferred learning and thinking style in a sample of Korean high school students. *Educational and Psychological Measurement*, 55(1), 60–74. <https://doi.org/10.1177/0013164495055001006>
- Kim, H. J., & Pedersen, S. (2011). Advancing young adolescents' hypothesis-development performance in a computer-supported and problem-based learning environment. *Computers & Education*, 57(2), 1780–1789. <https://doi.org/10.1016/j.compedu.2011.03.014>
- Kimbell, R. (2012). Evolving project e-scape for national assessment. *International Journal of Technology and Design Education*, 22, 135–155. <https://doi.org/10.1007/s10798-011-9190-4>
- Kind, P. M. (2013). Establishing assessment scales using a novel disciplinary rationale for scientific reasoning. *Journal of Research in Science Teaching*, 50(5), 530–560. <https://doi.org/10.1002/tea.21086>
- Kind, P. M., & Kind, V. (2007). Creativity in Science Education: Perspectives and Challenges for Developing School Science. *Studies in Science Education*, 43(1), 1–37. <https://doi.org/10.1080/03057260708560225>
- Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. The MIT Press.
- Klahr, D. and Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1-48.
- Klahr, D., & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, 125(5), 524–543. <https://doi.org/10.1037/0033-2909.125.5.524>
- Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive psychology*, 25(1), 111-146.
- Klahr, D., Zimmerman, C., & Matlen, B. J. (2019). Improving students' scientific thinking. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of*

- cognition and education (pp. 67–99). Cambridge University Press. <https://doi.org/10.1017/9781108235631.005>
- Kline, P. (2013). *Handbook of psychological testing*. Routledge.
- Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford publications.
- Knekta, E., Runyon, C., & Eddy, S. (2019). One Size Doesn't Fit All: Using Factor Analysis to Gather Validity Evidence When Using Surveys in Your Research. *CBE life sciences education*, 18(1), rm1. <https://doi.org/10.1187/cbe.18-04-0064>
- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, 15(10), 694-698.
- Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology / Schweizerische Zeitschrift für Psychologie / Revue Suisse de Psychologie*, 64(3), 141–152. <https://doi.org/10.1024/1421-0185.64.3.141>
- Koerber, S., Mayer, D., Osterhaus, C., Schwippert, K., & Sodian, B. (2015). The development of scientific thinking in elementary school: a comprehensive inventory. *Child development*, 86(1), 327–336. <https://doi.org/10.1111/cdev.12298>
- Kogan, N. (1974). Creativity and sex differences. *The Journal of Creative Behavior*, 8(1), 1–14. <https://doi.org/10.1002/j.2162-6057.1974.tb01103.x>
- Kogan, N., & Pankove, E. (1974). Long-term Predictive Validity of Divergent-thinking Tests: Some Negative Evidence. *Journal of Educational Psychology*, 66(6), 802–810. <https://doi.org/10.1037/h0021521>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. The MIT Press.
- Koslowski, B., Marasia, J., Chelenza, M., & Dublin, R. (2008). Information becomes evidence when an explanation can incorporate it into a causal framework. *Cognitive Development*, 23(4), 472–487. <https://doi.org/10.1016/j.cogdev.2008.09.007>
- Kousoulas, F. (2010). The interplay of creative behavior, divergent thinking, and knowledge base in students' creative expression during learning activity. *Creativity Research Journal*, 22(4), 387–396. <https://doi.org/10.1080/10400419.2010.523404>
- Köksal-Tuncer, Ö., & Sodian, B. (2018). The development of scientific reasoning: Hypothesis testing and argumentation from evidence in young children. *Cognitive Development*, 48, 135–145. <https://doi.org/10.1016/j.cogdev.2018.06.011>
- Kruk, M., Pawlak, M., Shirvan, M. E., & Soleimanzadeh, S. (2023). Revisiting boredom in practical English language classes via exploratory structural equation

- modeling. *Research Methods in Applied Linguistics*, 2(1), 100038. <https://doi.org/10.1016/j.rmal.2022.100038>
- Krumm, G., Lemos, V., & Filippetti, V. A. (2014). Factor structure of the Torrance Tests of Creative Thinking Figural Form B in Spanish-speaking children: Measurement invariance across gender. *Creativity Research Journal*, 26(1), 72–81. <https://doi.org/10.1080/10400419.2013.843908>
- Krumm, G., Arán Filippetti, V., Lemos, V., Koval, J., & Balabanian, C. (2016). Construct validity and factorial invariance across sex of the Torrance Test of Creative Thinking—Figural Form A in Spanish-speaking children. *Thinking Skills and Creativity*, 22, 180–189. <https://doi.org/10.1016/j.tsc.2016.10.003>
- Kuang, X., Eysink, T. H. S., & de Jong, T. (2022). Effects of providing domain information on facilitating hypothesis generation in inquiry learning. *The Journal of Educational Research*, 115(5), 285–297. <https://doi.org/10.1080/00220671.2022.2124219>
- Kuang, X., Eysink, T. H. S., & de Jong, T. (2024). Presenting domain information or self-exploration to foster hypothesis generation in simulation-based inquiry learning. *Journal of Research in Science Teaching*, 61(1), 70–102. <https://doi.org/10.1002/tea.21865>
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96, 674–689.
- Kuhn, D. (2005). *Education for thinking*. Harvard University Press.
- Kuhn, D. (2011). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 497–523). Wiley Blackwell.
- Kuhn, D., Amsel, E., & O’Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.
- Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, 60(4), v–128. <https://doi.org/10.2307/1166059>
- Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development*, 1(1), 113–129. https://doi.org/10.1207/S15327647JCD0101N_11
- Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, 18(4), 495–523. https://doi.org/10.1207/S1532690XCI1804_3
- Kuhn, D., & Franklin, S. (2006). The Second Decade: What Develops (and How). In D. Kuhn, R. S. Siegler, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Cognition, perception, and language* (6th ed., pp. 953–993). John Wiley & Sons, Inc..
- Kulkarni, D., and Simon, H. A. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science*, 12(2), 139–175.

- Lau, S., & Cheung, P. C. (2010). Developmental trends of creativity: What twists of turn do boys and girls take at different grades? *Creativity Research Journal*, 22(3), 329–336. <https://doi.org/10.1080/10400419.2010.503543>
- Lau, S., & Cheung, P. C. (2015). A gender-fair look at variability in creativity: Growth in variability over a period versus gender comparison at a time point. *Creativity Research Journal*, 27(1), 87–95. <https://doi.org/10.1080/10400419.2015.992685>
- Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, 86(3), 681–718. <https://doi.org/10.3102/0034654315627366>
- Leandre, R., Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press.
- Ledesma, R. D. & Valero-Mora, P., (2007) Determining the Number of Factors to Retain in EFA: An easy-to-use computer program for carrying out Parallel Analysis. *Practical Assessment, Research, and Evaluation*, 12(1): 2. <https://doi.org/10.7275/wjnc-nm63>
- Lee, J.-E., Day, J. D., Meara, N. M., & Maxwell, S. (2002). Discrimination of social knowledge and its flexible application from creativity: A multitrait-multimethod approach. *Personality and Individual Differences*, 32(5), 913–928. [https://doi.org/10.1016/S0191-8869\(01\)00099-X](https://doi.org/10.1016/S0191-8869(01)00099-X)
- Leu, Y. C., & Chiu, M. S. (2015). Creative behaviours in mathematics: Relationships with abilities, demographics, affects and gifted behaviours. *Thinking Skills and Creativity*, 16, 40-50. <https://doi.org/10.1016/j.tsc.2015.01.001>
- Liang, J. C. (2002). *Exploring scientific creativity of eleventh-grade students in Taiwan*. Unpublished Doctoral Dissertation. Austin, TX: The University of Texas. Retrieved from ProQuest (PQ number: 3099482).
- Lim Y. J. (2022). Exploratory Structural Equation Modeling Analysis of the Social Network Site Use Motives Scale. *Psychiatry investigation*, 19(2), 146–153. <https://doi.org/10.30773/pi.2021.0092>
- Lin, S., & Wong, C. S. (2014). The creativity level of Taiwan hospitality undergraduate students. *Procedia – Social and Behavioral Sciences*, 144, 54-59. <http://doi.org/10.1016/j.sbspro.2014.07.273>
- Lin, C. D., Hu, W. P., Adey, P., & Shen, J. L. (2003). The influence of CASE on scientific creativity. *Research In Science Education*, 33(2), 143-162. <https://doi.org/10.1023/A:1025078600616>
- Liu, C. W., & Wang, W. C. (2016). A comparison of methods for dimensionality assessment of categorical item responses. In *Pacific Rim Objective Measurement Symposium (PROMS) 2015 Conference Proceedings* (pp. 395-410). Springer Singapore.
- Loevinger, J. (1957). Objective Tests as Instruments of Psychological Theory. *Psychological Reports*, 3(3), 635-694. <https://doi.org/10.2466/pr0.1957.3.3.635>
- Long, H. (2014). More than appropriateness and novelty: Judges' criteria of assessing creative products in science tasks. *Thinking Skills and Creativity*, 13, 183–194. <https://doi.org/10.1016/j.tsc.2014.05.002>

- Lubart, T., & Guignard, J.-H. (2004). The Generality-Specificity of Creativity: A Multivariate Approach. In R. J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity: From potential to realization* (pp. 43–56). American Psychological Association. <https://doi.org/10.1037/10692-004>
- Lubart, T.I., Besanc_on, M., & Barbot, B. (2011). Evaluation du Potentiel Cr_eatif (EPoC). Paris: Editions Hogrefe.
- Luo, M., Sun, D., Zhu, L., & Yang, Y. (2021). Evaluating scientific reasoning ability: Student performance and the interaction effects between grade level, gender, and academic achievement level. *Thinking Skills and Creativity*, 41, 100899.
- MacKinnon, D. W. (1975). IPAR's Contribution to the Conceptualization and Study of Creativity. In I. Taylor, & J. Getzels (Eds.), *Perspectives in Creativity* (pp. 62-80). Chicago, IL: Aldine.
- Madjar, N., Greenberg, E., & Chen, Z. (2011). Factors for radical creativity, incremental creativity, and routine, noncreative performance. *Journal of Applied Psychology*, 96(4), 730–743. <https://doi.org/10.1037/a0022416>
- Majumdar, S. K. (1975). A systems approach to identification and nurture of scientific creativity [Abstract]. *Journal of Indian Education*, 1(2), 17-23.
- Mansfield, R.S., & Busse, T.V. (1981). *The Psychology of Creativity and Discovery: Scientists and Their Work*. Nellson Hall Inc.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the retrospective study of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519-530. <https://doi.org/10.1093/biomet/57.3.519>
- Markos, A., & Tsigilis, N. (2024). Dimensionality assessment in ordinal data: a comparison between parallel analysis and exploratory graph analysis. *Frontiers in Psychology*, 15, 1359111. <https://doi.org/10.3389/fpsyg.2024.1359111>
- Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual review of clinical psychology*, 10(1), 85-110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 439–476. <https://doi.org/10.1080/10705510903008220>
- Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2020). Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and set- esem: optimal balance between goodness of fit and parsimony. *Multivariate behavioral research*, 55(1), 102–119. <https://doi.org/10.1080/00273171.2019.1602503>
- Marsh, H., & Alamer, A. (2024). When and how to use set-exploratory structural equation modelling to test structural models: A tutorial using the R package lavaan. *British*

- Journal of Mathematical and Statistical Psychology*, 77. 459-476.
<https://doi.org/10.1111/bmsp.12336>
- Masnick, A. M., & Morris, B. J. (2008). Investigating the development of data evaluation: The role of data characteristics. *Child Development*, 79(4), 1032–1048. <https://doi.org/10.1111/j.1467-8624.2008.01174.x>
- Masnick, A. M., Klahr, D., & Knowles, E. R. (2016). Data-Driven Belief Revision in Children and Adults. *Journal of Cognition and Development*, 18(1), 87–109. <https://doi.org/10.1080/15248372.2016.1168824>
- Matud, M. P., Rodríguez, C., & Grande, J. (2007). Gender differences in creative thinking. *Personality and Individual Differences*, 43(5), 1137–1147. <https://doi.org/10.1016/j.paid.2007.03.006>
- Mayer, R. E. (1999). Fifty years of creativity research. In R. J. Sternberg (Ed.), *Handbook of Creativity* (pp. 449-460). Cambridge: Cambridge University Press.
- Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction*, 29, 43–55. <https://doi.org/10.1016/j.learninstruc.2013.07.005>
- McDonald, R.P. (1985). *Factor Analysis and Related Methods* (1st ed.). Psychology Press. <https://doi.org/10.4324/9781315802510>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum.
- McNeish, D. (2017). Exploratory factor analysis with small samples and missing data. *Journal of Personality Assessment*, 99(6), 637–652. <https://doi.org/10.1080/00223891.2016.1252382>
- Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69(3), 220–232. <https://doi.org/10.1037/h0048850>
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International journal of educational research*, 13(2), 127-143. [https://doi.org/10.1016/0883-0355\(89\)90002-5](https://doi.org/10.1016/0883-0355(89)90002-5)
- Meng, X.-l., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111(1), 172–175. <https://doi.org/10.1037/0033-2909.111.1.172>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543. <https://doi.org/10.1007/BF02294825>
- Millsap, R. E., & Everson, H. T. (1993). Methodology Review: Statistical Approaches for Assessing Measurement Bias. *Applied Psychological Measurement*, 17(4), 297-334. <https://doi.org/10.1177/014662169301700401>
- Miroshnik, K. G., Shcherbakova, O. V., & Kaufman, J. C. (2022). Kaufman Domains of Creativity Scale: Relationship to occupation and measurement invariance across gender. *Creativity Research Journal*, 34(2), 159–177. <https://doi.org/10.1080/10400419.2021.1953823>
- Misra, K.S.(1986). Effect of Home and School Environment On Scientific Creativity, Sangyanalaya, Kanpur, India.

- Mohamed, A. H. (2006). *Investigating the scientific creativity of fifth-grade students*. Unpublished Doctoral Dissertation. Tucson, AZ: University of Arizona. Retrieved from ProQuest (PQ number: 3207635).
- Morais, M. F., & Almeida, L. (2019). "I would be more creative if...": Are there perceived barriers to college students' creative expression according to gender? ["Eu seria mais criativo se...": Barreiras à expressão criativa percebidas segundo o gênero dos universitários]. *Estudos de Psicologia*, 36, Article e180011. <https://doi.org/10.1590/1982-0275201936e180011>
- Moran III, J. D., Sawyers, J. K., & Moore, A. J. (1988). Effects of structure in instruction and materials on preschoolers' creativity. *Home economics research journal*, 17(2), 148-152.
- Moravcsik, M. J. (1981). Creativity in science education. *Science Education*, 65, 221-227. doi:10.1002/sce.3730650212
- Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2015). A Bifactor Exploratory Structural Equation Modeling Framework for the Identification of Distinct Sources of Construct-Relevant Psychometric Multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 116–139. <https://doi.org/10.1080/10705511.2014.961800>
- Morin, A. J. S., Myers, N. D., & Lee, S. M. (2020). Modern factor analytic techniques: Bifactors models, exploratory structural equation modeling (ESEM), and bifactor-ESEM. In G. Tenenbaum, & R. C. Eklund (Eds.), *Handbook of sport psychology* (fourth ed., pp. 1044–1073). Wiley. <https://doi.org/10.1002/9781119568124.ch51>.
- Morin, A. J., Marsh, H., & Nagengast, B. (2013). Exploratory structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 395–436). Information Age Publishing
- Mukhopadhyay, R. (2013). Measurement of Creativity in Physics - A Brief Review on Related Tools. *IOSR Journal of Humanities and Social Science*, 6, 45-50.
- Mumford, M. D., Mobley, M. I., Reiter-Palmon, R., Uhlman, C. E., & Doares, L. M. (1991). Process analytic models of creative capacities. *Creativity Research Journal*, 4(2), 91–122. <https://doi.org/10.1080/10400419109534380>
- Muthén, L. K., & Muthén, B. O. (2021). *Mplus (Version 8.6)* [Statistical software].
- Muthén, B., & Muthén, B. O. (2009). *Statistical analysis with latent variables* (Vol. 123, No. 6). Wiley.
- Nakano, T. d. C., Oliveira, K. d. S., & Zaia, P. (2021). Gender differences in creativity: A systematic literature review. *Psicologia: Teoria e Pesquisa*, 37, Article e372116. <https://doi.org/10.1590/0102.3772e372116>
- Netemeyer, R. G. (2003). *Scaling procedures: Issues and applications*. Sage Publications.
- National Research Council (NRC) (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. The National Academies Press.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175-220.
- Nilada, N., Payoungkiattikun, W., & Thongsuk, T. (2024). The study of scientific creativity using a project-based learning management model. *International Journal on*

- Social and Education Sciences (IJonSES)*, 6(2), 253-263.
<https://doi.org/10.46328/ijonSES.667>
- Njoo, M., & de Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching*, 30(8), 821–844. <https://doi.org/10.1002/tea.3660300803>
- Nunnally, J.C. and Bernstein, I.H. (1994) The Assessment of Reliability. *Psychometric Theory*, 3, 248-292.
- O’connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer’s MAP test. *Behavior research methods, instruments, & computers*, 32(3), 396-402.
<https://doi.org/10.3758/BF03200807>
- Opitz, A., Heene, M., & Fischer, F. (2017). Measuring scientific reasoning – a review of test instruments. *Educational Research and Evaluation*, 23(3–4), 78–101.
<https://doi.org/10.1080/13803611.2017.1338586>
- Ozdemir, G., & Dikici, A. (2017). Relationships between scientific process skills and scientific creativity: Mediating role of nature of science knowledge. *Journal of Education in Science Environment and Health*, 3(1), 52-68.
- Özdemir, N. N., & Sak, U., (2013). A Componential Analysis of Gender Differences in Scientific Creativity. *Turkish Journal of Giftedness and Education* , 3, 53-65.
- Paek, S. H. and Runco, M. A. (2017). Dealing with the criterion problem by measuring the quality and quantity of creative activity and accomplishment. *Creativity Research Journal*, 29(2), 167–173.
- Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., ... & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational research review*, 14, 47-61.
<https://doi.org/10.1016/j.edurev.2015.02.003>
- Penner, D. E., & Klahr, D. (1996). The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects. *Child Development*, 67(6), 2709–2727. <https://doi.org/10.2307/1131748>
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, 21(2), 381–391. <https://doi.org/10.1086/209405>
- Piaget, J. (1929). *The child’s conception of the world*. Kegan Paul.
- Pickering, A. (1992). *Science as practice and culture*. University of Chicago Press.
- Piekny, J., & Maehler, C. (2012). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology*, First article. doi:10.1111/j.2044835X.2012.02082.x
- Piekny, J., Grube, D., & Maehler, C. (2013). The Development of Experimentation and Evidence Evaluation Skills at Preschool Age. *International Journal of Science Education*, 36(2), 334–354. <https://doi.org/10.1080/09500693.2013.776192>
- Plucker, J. A. (1998). Beware of simple conclusions: The case for content generality of creativity. *Creativity Research Journal*, 11(2), 179-182.

- Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, 39(2), 83–96. https://doi.org/10.1207/s15326985ep3902_1
- Plucker, J. A., Holden, J., & Neustadter, D. (2008). The criterion problem and creativity in film: Psychometric characteristics of various measures. *Psychology of Aesthetics, Creativity, and the Arts*, 2(4), 190–196. <https://doi.org/10.1037/a0012839>
- Plucker, J., & Zabelina, D. (2009). Creativity and interdisciplinarity: One creativity or many creativities?. *Zdm*, 41, 5-11. <https://doi.org/10.1007/s11858-008-0155-3>
- Popper, K. R. (1959). *The logic of scientific discovery*. Basic Books.
- Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics*, 2(1), 13–43. https://doi.org/10.1207/S15328031US0201_02
- Preist, T. (2006). Self-evaluation, creativity, and musical achievement. *Psychology of Music*, 34(1), 47-61.
- Prokofieva, M., Zarate, D., Parker, A., & Stavropoulos, V. (2023). Exploratory structural equation modeling: A streamlined step-by-step approach using the R Project software. *BMC Psychiatry*, 23, 546. <https://doi.org/10.1186/s12888-023-05028-9>
- Purc, E., & Laguna, M. (2019). Factorial structure and measurement invariance of the Innovative Behavior Questionnaire. *The Journal of Creative Behavior*, 53(3), 404–410. <https://doi.org/10.1002/jocb.215>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review*, 41, 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Raina, T. N. (1980). Sex differences in creativity in India: A second look [Abstract]. *Journal of Creative Behavior*, 14, 218
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87(3), 517–529. <https://doi.org/10.1037/0021-9010.87.3.517>
- Raju, N. S., & Ellis, B. B. (2002). Differential item and test functioning. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 156–188). Jossey-Bass/Wiley.
- Ramly, S. N. F., Ahmad, N. J., & Yakob, N. (2022). Development, validity, and reliability of chemistry scientific creativity test for pre-university students. *International Journal of Science Education*, 44(14), 1–16. <https://doi.org/10.1080/09500693.2022.2116298>
- Raubenheimer, J. (2004). An item selection procedure to maximize scale reliability and validity. *SA Journal of Industrial Psychology*, 30(4), 59-64.
- Reis, S. M. and Renzulli, J. S. (1991). The assessment of creative products in programs for gifted and talented students. *Gifted Child Quarterly*, 35, 128-134.

- Reiter-Palmon, R., & Arreola, N. J. (2015). Does generating multiple ideas lead to increased creativity? A comparison of generating one idea vs. many. *Creativity Research Journal*, 27(4), 369–374. <https://doi.org/10.1080/10400419.2015.1087274>
- Reynolds, C. R. (2000). Why is psychometric research on bias in mental testing so often ignored? *Psychology, Public Policy, and Law*, 6(1), 144–150. <https://doi.org/10.1037/1076-8971.6.1.144>
- Rhodes, M. (1961). An Analysis of Creativity. *The Phi Delta Kappan*, 42(7), 305-310.
- Rocke, A. J. (2010). *Image and reality: Kekulé, Kopp, and the scientific imagination*. Chicago: University of Chicago Press
- Romo, M. (2018). Tiene género la creatividad? Obstáculos a la excelencia en mujeres [Does creativity have a gender? Obstacles to excellence in women]. *Estudios de Psicología*, 35(3), 247–258. <https://doi.org/10.1590/1982-02752018000300003>
- Rosenblatt, E., & Winner, E. (1988). The art of children's drawing. *Journal of Aesthetic Education*, 22(1), 3–15. <https://doi.org/10.2307/3332960>
- Rothenberg, A. (1995). Creative cognitive processes in Kekule's discovery of the structure of the benzene molecule. *The American Journal of Psychology*, 108(3), 419-438.
- Runco, M. A. (1984). Teachers' judgments of creativity and social validation of divergent thinking tests. *Perceptual and Motor Skills*, 59(3), 711–717. <https://doi.org/10.2466/pms.1984.59.3.711>
- Runco, M. A. (1989). The creativity of children's art. *Child Study Journal*, 19(3), 177–189.
- Runco, M. A. (1996). Personal creativity: Definition and developmental issues. *New Directions for Child and Adolescent Development*, 1996(72), 3-30. <https://doi.org/10.1002/cd.23219967203>
- Runco, M. (1999a). Divergent Thinking. In M. Runco, and S. R. Pritzker (Eds.), *Encyclopedia of creativity* (Vol 1). (pp. 577-582). San Diego: Academic Press.
- Runco, M. (1999b). Tension, adaptability, and creativity. In S. Russ (Ed.), *Affect, creative experience, and psychological adjustment* (pp. 165–194). Brunner/Mazel.
- Runco, M. (2007). *Creativity*. Elsevier.
- Runco, M. A. (2008). Commentary: Divergent thinking is not synonymous with creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 2(2), 93–96. <https://doi.org/10.1037/1931-3896.2.2.93>
- Runco, M. A. (1987). The generality of creative performance in gifted and nongifted Children. *Gifted Child Quarterly*, 31(3), 121-125.
- Runco, M. A. (2020). Creativity Testing Services. Retrieved from <https://www.creativitytestingservices.com>
- Runco, M. A., & Charles, R. E. (1993). Judgments of originality and appropriateness as predictors of creativity. *Personality and Individual Differences*, 15(5), 537–546. [https://doi.org/10.1016/0191-8869\(93\)90337-3](https://doi.org/10.1016/0191-8869(93)90337-3)
- Runco, M. A., Millar, G., Acar, S., & Cramond, B. (2010). Torrance tests of creative thinking as predictors of personal and public achievement: A fifty-year follow-

- up. *Creativity Research Journal*, 22(4), 361–368. <https://doi.org/10.1080/10400419.2010.523393>
- Runco M. A., Cramond B., Pagnani A. R. (2010). Gender and Creativity. In J. Chrisler & D. McCreary (Eds.), *Handbook of Gender Research in Psychology* (pp. 343-357). Springer. https://doi.org/10.1007/978-1-4419-1465-1_17
- Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, 24(1), 66–75. <https://doi.org/10.1080/10400419.2012.652929>
- Ruvman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). ReXecting on scientiWc thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, 64, 1617–1636.
- Ryser, G. R. (2007). *Profiles of creative abilities: Examiner's manual*. Austin: Pro-Ed.
- Pallant, J. (2020). *SPSS Survival Manual: A Step by Step Guide to Data Analysis using IBM SPSS* (7th ed.). Open University Press.
- Runco, M. (1999). Divergent thinking. In M. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity* (pp. 577–582). Academic Press.
- Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, 24(2), 282–292. <https://doi.org/10.1037/a0025697>
- Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., Van den Noortgate, W., & Barbot, B. (2021). Does the fourth-grade slump in creativity actually exist? A meta-analysis of the development of divergent thinking in school-age children and adolescents. *Educational Psychology Review*, 33(1), 275–298. <https://doi.org/10.1007/s10648-020-09547-9>
- Said-Metwaly, S., Van den Noortgate, W., & Barbot, B. (2021). Torrance Test of Creative Thinking-Verbal, Arabic version: Measurement invariance and latent mean differences across gender, year of study, and academic major. *Thinking Skills and Creativity*, 39, Article 100768. <https://doi.org/10.1016/j.tsc.2020.100768>
- Saggar, M., Xie, H., Beaty, R. E., Stankov, A. D., Schreier, M., & Reiss, A. L. (2019). Creativity slumps and bumps: Examining the neurobehavioral basis of creativity development during middle childhood. *NeuroImage*, 196, 94–101. <https://doi.org/10.1016/j.neuroimage.2019.03.080>
- Sak, U., & Maker, C. J. (2006). Developmental Variation in Children's Creative Mathematical Thinking as a Function of Schooling, Age, and Knowledge. *Creativity Research Journal*, 18(3), 279–291. https://doi.org/10.1207/s15326934crj1803_5
- Sak, U., & Ayas, M. B. (2013). Creative Scientific Ability Test (C-SAT): A new measure of scientific creativity. *Psychological Test and Assessment Modeling*, 55(3), 316–329.
- Sah, G. (2019). Examining the development of creativity in adolescents in 9th and 12th grades: a four-year longitudinal study. *Creativity studies*, 12(2), 341–360. <https://doi.org/10.3846/cs.2019.10260>

- Santiago, P. H. R., Manzini, D., Haag, D., Roberts, R., Smithers, L. G., & Jamieson, L. (2022). Exploratory Graph Analysis of the Strengths and Difficulties Questionnaire in the Longitudinal Study of Australian Children. *Assessment*, 29(8), 1622-1640. <https://doi.org/10.1177/10731911211024338>
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514. <https://doi.org/10.1007/BF02296192>
- Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing children's early competencies to engage with science. *Educational Psychologist*, 49(2), 139–152. <https://doi.org/10.1080/00461520.2014.917589>
- Sawyer, R. K. (2012). *Explaining creativity: The science of human innovation* (2nd ed.). Oxford University Press.
- Sawyer, R. K., & Henriksen, D. (2024). *Explaining creativity: The science of human innovation*. Oxford university press.
- Shapiro, R. J. (1970). The criterion problem. In P. E. Vernon (Ed.), *Creativity* (pp. 257–269). New York: Penguin.
- Sharma, N. (2015). Scientific creativity in relation to cognitive style and achievement in Science of secondary school students. *Educational Quest*, 6(1), 25-29.
- Shukla, J. P., and Sharma, V. P. (1986). Sex differences in scientific creativity [Abstract]. *Indian Psychological Review*, 30(3), 32-35.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Schalk, L., Edelsbrunner, P. A., Deiglmayr, A., Schumacher, R., & Stern, E. (2019). Improved application of the control-of-variables strategy as a collateral benefit of inquiry-based physics education in elementary school. *Learning and Instruction*, 59, 34–45. <https://doi.org/10.1016/j.learninstruc.2018.09.006>
- Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal Models and Experimentation Strategies in Scientific Reasoning. *Journal of the Learning Sciences*, 1(2), 201–238. https://doi.org/10.1207/s15327809jls0102_3
- Schlatter, E., Molenaar, I., & Lazonder, A. W. (2020). Individual differences in children's development of scientific reasoning through inquiry-based instruction: Who needs additional guidance? *Frontiers in Psychology*, 11, Article 904. <https://doi.org/10.3389/fpsyg.2020.00904>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Shadish, W. R., & Fuller, S. (1994). *Social psychology of science*. Guilford Press.
- Siegler, R. S., & Liebert, R. M. (1975). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Designing a factorial experiment. *Developmental Psychology*, 11(3), 401–402. <https://doi.org/10.1037/h0076579>

- Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology*, 36(3), 273–310. <https://doi.org/10.1006/cogp.1998.0686>
- Siew, N. M., Chong, C. L., and Chin, K. O. (2014). Developing a scientific creativity test for fifth graders. *Problems of Education in the 21st Century*, 62, 109-123.
- Siler, S., Klahr, D., & Matlen, B. (2013). Conceptual change in experimental design: From engineering goal to science goals. In S. Vosniadou (ed.), *Handbook of research on conceptual change*, 2nd edn. New York: Routledge
- Silvia, P. J., Kaufman, J. C., & Pretz, J. E. (2009). Is creativity domain-specific? Latent class models of creative accomplishments and creative self-descriptions. *Psychology of Aesthetics, Creativity, and the Arts*, 3(3), 139–148. <https://doi.org/10.1037/a0014940>
- Smith, G. J. W., & Carlsson, I. (1983). Creativity in early and middle school years. *International Journal of Behavioral Development*, 6(2), 167-195. <https://doi.org/10.1177/016502548300600204>
- Smith, G. J., & Carlsson, I. (1985). Creativity in middle and late school years. *International Journal of Behavioral Development*, 8(3), 329–343. <https://doi.org/10.1177/016502548500800307>
- Smith, G.J.W. and Carlsson, L.M. (1990) *The Creative Process: A Functional Model Based on Empirical Studies from Childhood to Middle Age*. International Universities Press.
- Simonton, D. K. (1984). *Genius, creativity, and leadership*. Cambridge University Press
- Simonton, D. K. (1988). Age and outstanding achievement: What do we know after a century of research? *Psychological Bulletin*, 104(2), 251–267. <https://doi.org/10.1037/0033-2909.104.2.251>
- Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review*, 104(1), 66–89. <https://doi.org/10.1037/0033-295X.104.1.66>
- Simonton, D. K. (1999). *Origins of Genius: Darwinian Perspectives on Creativity*. Oxford University Press USA.
- Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist*. Cambridge Univ Pr.
- Simonton, D. K. (2021). Scientific creativity: Discovery and invention as combinatorial. *Frontiers in psychology*, 12, 721104.
- Sinha, A. K., and Singh, C. (1987). Measurement of scientific creativity [Abstract]. *Indian Journal of Psychometry & Education*, 18(1), 1–13.
- Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, 21, 177-237. doi:10.1016/0010-0277(85)90025-3
- Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child development*, 58(2), 424-433.

- Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child development*, 62(4), 753-766.
- Soldz, S. and Vaillant, G. E. (1999). The big five personality traits and the life course: A 45 year longitudinal study. *Journal of Research in Personality*, 33, 208-232.
- Somerville, S. C., Hadkinson, B. A., & Greenberg, C. (1979). Two levels of inferential behavior in young children. *Child Development*, 50(1), 119-131. <https://doi.org/10.2307/1129048>
- Soper, D. S. (2020, July). *A-priori sample size calculator for structural equation models [Software]*.
- Starko, A. J. (2017). *Creativity in the Classroom: Schools of Curious Delight*. Routledge. <https://doi.org/10.4324/9781315391625>
- Stein, M. I. (1953). Creativity and Culture. *The Journal of Psychology*, 36(2), 311-322. <https://doi.org/10.1080/00223980.1953.9712897>
- Sternberg, R. J. (1999). *Handbook of Creativity*. Cambridge University Press.
- Sternberg, R. J. (2006). Creating a vision of creativity: The first 25 years. *Psychology of Aesthetics, Creativity, and the Arts*, 5(1), 2-12. <https://doi.org/10.1037/1931-3896.5.1.2>
- Sternberg, R. J., & Lubart, T. I. (1993). Investing in creativity. *Psychological Inquiry*, 4(3), 229-232. https://doi.org/10.1207/s15327965pli0403_16
- Sternberg, R. J., & Lubart, T. I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity*. Free Press.
- Sternberg, R. J., Kaufman, J. C. and Pretz, J. E. (2001). The propulsion model of creative contributions applied to the arts and letters. *Journal of Creative Behavior*, 35(2), 75-101.
- Sternberg, R. J., & Grigorenko, E. L. (2001). Guilford's structure of intellect model and model of creativity: Contributions and limitations. *Creativity Research Journal*, 13(3-4), 309-316. https://doi.org/10.1207/S15326934CRJ1334_08
- Sternberg, R. J. and Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In R. J. Sternberg (Ed.), *Handbook of Creativity* (pp. 3-15). Cambridge University Press.
- Sternberg, R. J., Todhunter, R. J., Litvak, A., & Sternberg, K. (2020). The relation of scientific creativity and evaluation of scientific impact to scientific reasoning and general intelligence. *Journal of Intelligence*, 8(2), 17. <https://doi.org/10.3390/jintelligence8020017>
- Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (Vol. 4). Lawrence Erlbaum Associates.
- Stoltzfus, G., Nibbelink, B. L., Vredenburg, D. and Hyrum, E. (2011). Gender, Gender Role, and Creativity. *Social Behavior and Personality: An International Journal*, 39, 425-432. <http://dx.doi.org/10.2224/sbp.2011.39.3.42>
- Strauss, S., & Stavy, R. (1982). U-shaped behavioral growth: Implications for theories of development. In W. W. Hartup (Ed.), *Review of child development research* (vol. 6) (pp. 547-599). University of Chicago Press

- Streiner, D. L. (1994). Figuring out factors: the use and misuse of factor analysis. *The Canadian Journal of Psychiatry*, 39(3), 135-140. <https://doi.org/10.1177/070674379403900303>
- Sundre, D. L. (2008). *The Scientific Reasoning Test, Version 9 (SR-9) test manual*. Center for Assessment and Research Studies.
- Swami, V., Maïano, C., & Morin, A. J. (2023). A guide to exploratory structural equation modeling (ESEM) and bifactor-ESEM in body image research. *Body Image*, 47, 101641. <https://doi.org/10.1016/j.bodyim.2023.101641>
- Székel, G. J., & Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, 93(1), 58-80. <https://doi.org/10.1016/j.jmva.2003.12.002>
- Tamhane, A. C. (1979). A Comparison of Procedures for Multiple Comparisons of Means with Unequal Variances. *Journal of the American Statistical Association*, 74(366a), 471-480. <https://doi.org/10.1080/01621459.1979.10482541>
- Taylor, C. L., Said-Metwaly, S., Camarda, A., & Barbot, B. (2024). Gender differences and variability in creative ability: A systematic review and meta-analysis of the greater male variability hypothesis in creativity. *Journal of Personality and Social Psychology*, 126(6), 1161-1179. <https://doi.org/10.1037/pspp0000484>
- Tezbaşaran A. Likert Tipi Ölçek Hazırlama Kılavuzu. 2008; [Elektronik sürüm]. URL: http://www.academia.edu/1288035/Likert_Tipi_Ölçek_Hazırlama_Kılavuz, 20 Mart 2017.
- Thompson, T. L. (2016). The Mothers and Fathers of Invention: A Meta-Analysis of Gender Differences in Creativity. Retrieved from http://purl.flvc.org/fsu/fd/FSU_2016SP_Thompson_fsu_0071E_13059
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209-220. <https://doi.org/10.1037/a0023353>
- Torrance, E. P. (1968). A longitudinal examination of the fourth grade slump in creativity. *Gifted Child Quarterly*, 12(4), 195-199.
- Torrance, E. P. (1971). Are the Torrance Tests of Creative Thinking Biased Against or in Favor of "Disadvantaged" Groups? *Gifted Child Quarterly*, 15(2), 75-80. <https://doi.org/10.1177/001698627101500201>
- Torrance, E. P. (1972). Predictive Validity of the Torrance Tests of Creative Thinking. *The Journal of Creative Behavior*, 6(4), 236-252.
- Tóth-Király, I., Bőthe, B., Rigó, A., & Orosz, G. (2017). An Illustration of the Exploratory Structural Equation Modeling (ESEM) Framework on the Passion Scale. *Frontiers in psychology*, 8, 1968. <https://doi.org/10.3389/fpsyg.2017.01968>
- Tourangeau, R. (2003). Cognitive aspects of survey measurement and mismeasurement. *International Journal of Public Opinion Research*, 15(1), 3-7. <https://doi.org/10.1093/ijpor/15.1.3>
- Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, 51(1), 1-10. <https://doi.org/10.2307/1129583>

- Urban, K. K. (1991). On the development of creativity in children. *Creativity Research Journal*, 4(2), 177–191. <https://doi.org/10.1080/10400419109534384>
- Urban, K. K. (2003). Toward a componential model of creativity. In D. Ambrose, L. M. Cohen, & A. J. Tannenbaum (Eds.), *Creative intelligence: Toward theoretic integration* (pp. 81–112). Hampton Press
- Usta, E., and Akkanat, Ç. (2015). Investigating scientific creativity level of seventh grade students. *Procedia-Social and Behavioral Sciences*, 191, 1408-1415.
- Van der Graaf, J., Segers, E., & Verhoeven, L. (2015). Scientific reasoning abilities in kindergarten: Dynamic assessment of the control of variables strategy. *Instructional Science*, 43, 381–400. <http://dx.doi.org/10.1007/s11251-015-9344-y>.
- Van der Graaf, J., Segers, E., & Verhoeven, L. (2016). Scientific reasoning in kindergarten: Cognitive factors in experimentation and evidence evaluation. *Learning and Individual Differences*, 49, 190–200. <https://doi.org/10.1016/j.lindif.2016.06.006>
- Van der Graaf, J., Segers, E., & Verhoeven, L. (2018). Individual differences in the development of scientific thinking in kindergarten. *Learning and Instruction*, 56, 1–9. <https://doi.org/10.1016/j.learninstruc.2018.03.005>
- Van Zyl, L. E., & Ten Klooster, P. M. (2022). Exploratory structural equation modeling: Practical guidelines and tutorial with a convenient online tool for Mplus. *Frontiers in Psychiatry*, 12, 795672. <https://doi.org/10.3389/fpsyt.2021.795672>
- Velicer, W. F., & Jackson, D. N. (1990). Component Analysis versus Common Factor Analysis: Some issues in Selecting an Appropriate Procedure. *Multivariate Behavioral Research*, 25(1), 1–28. https://doi.org/10.1207/s15327906mbr2501_1
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327. <https://doi.org/10.1007/BF02293557>
- Vernon, P. E. (1989). The nature-nurture problem in creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 93–110). Plenum Press.
- Vries, H. B., & Lubart, T. I. (2019). Scientific creativity: Divergent and convergent thinking and the impact of culture. *The Journal of Creative Behavior*, 53(2), 145–155. <https://doi.org/10.1002/jocb.184>
- Yang, Y., & Green, S. B. (2011). Coefficient Alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, 29(4), 377–392. <https://doi.org/10.1177/0734282911406668>
- Yang, K. K., Lin, S. F., Hong, Z. R., & Lin, H. shyang. (2016). Exploring the Assessment of and Relationship Between Elementary Students' Scientific Creativity and Science Inquiry. *Creativity Research Journal*, 28(1), 16–23. <https://doi.org/10.1080/10400419.2016.1125270>
- Yang, K. K., Hong, Z. R., Lee, L., and Lin, H. S. (2019). Exploring the significant predictors of convergent and divergent scientific creativities. *Thinking Skills and Creativity*, 31, 252-261.

- Yi, X., Hu, W., Plucker, J. A., & McWilliams, J. (2013). Is there a developmental slump in creativity in China? The relationship between organizational climate and creativity development in Chinese adolescents. *The Journal of Creative Behavior*, 47(1), 22–40. <https://doi.org/10.1002/jocb.21>
- Yildiz, C., & Yildiz, T. G. (2021). Exploring the relationship between creative thinking and scientific process skills of preschool children. *Thinking Skills and Creativity*, 39, 100795. <https://doi.org/10.1016/j.tsc.2021.100795>
- Wallas, G. (1926). *The art of thought*. New York, NY: Harcourt, Brace and Company.
- Warne, R. T., Yoon, M., & Price, C. J. (2014). Exploring the various interpretations of “test bias”. *Cultural Diversity & Ethnic Minority Psychology*, 20(4), 570–582. <https://doi.org/10.1037/a0036503>
- Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, 44(3), 219–246. <https://doi.org/10.1177/0095798418771807>
- Weiner, I. B., Nelson, R. J., & Mizumori, S. (2012). *Handbook of Psychology, Behavioral Neuroscience* (Vol. 3). John Wiley & Sons.
- Weisberg, R. W. (1999). Creativity and knowledge: A challenge to theories. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 226–250). Cambridge University Press.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). American Psychological Association. <https://doi.org/10.1037/10222-009>
- Widaman, K. F., & Grimm, K. J. (2014). Advanced psychometrics: Confirmatory factor analysis, item response theory, and the study of measurement invariance. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 534–570). Cambridge University Press.
- Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage publications.
- Wilkening, F., & Huber, S. (2004). Children's intuitive physics. In U. Goswami (Ed.), *The Blackwell handbook of childhood cognitive development*. Blackwell Reference Online. <http://dx.doi.org/10.1111/b.9780631218418.2004.00019.x>.
- Winder, C., & Dowlatabadi, Z. (2013). *Producing animation*. Routledge.
- Winner, E. (1982) *Invented worlds: The psychology of the arts*. Harvard University Press.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, 73(6), 913–934. <https://doi.org/10.1177/0013164413495237>
- Ximénez, C., Revuelta, J., & Castañeda, R. (2022). What are the consequences of ignoring cross-loadings in bifactor models? A simulation study assessing parameter recovery and sensitivity of goodness-of-fit indices. *Frontiers in psychology*, 13, 923877. <https://doi.org/10.3389/fpsyg.2022.923877>

- Xiong, Y., Li, Y., Chen, Y., Yuan, P., Fan, Y., & Jiang, W. (2015). The creative investigation of brain activity with eeg for gender and left/right-handed differences. *Journal of Mechanics in Medicine and Biology*, 15(04), 1550054. <https://doi.org/10.1142/s0219519415500542>
- Xu, X. (2013). *The construct and type analysis on scientific creativity of High-school students* [Unpublished master dissertation]. Beijing Normal University. (in Chinese).
- Xu, S., Reiss, M.J. & Lodge, W. Comprehensive Scientific Creativity Assessment (C-SCA): A New Approach for Measuring Scientific Creativity in Secondary School Students. *Int J of Sci and Math Educ* **23**, 293–319 (2025). <https://doi.org/10.1007/s10763-024-10469-z>
- Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creativity Research Journal*, 23(1), 24–37. <https://doi.org/10.1080/10400419.2011.545713>
- Zhu, W., Shang, S., Jiang, W., Pei, M., & Su, Y. (2019). Convergent Thinking Moderates the Relationship between Divergent Thinking and Scientific Creativity. *Creativity Research Journal*, 31(3), 320–328. <https://doi.org/10.1080/10400419.2019.1641685>
- Zinbarg, R.E., Revelle, W., Yovel, I. and Li, W. (2005). Cronbach's a, Revelle's b, and McDonald's vH: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, (70)1, 123-133, doi: 10.1007/s11336-003-0974-7.
- Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, 20(1), 99–149. <https://doi.org/10.1006/drev.1999.0497>
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27, 172-223.
- Zimmerman, C., & Glaser, R. (2003). A Follow-Up Investigation of the Role of Cover Story on the Assessment of Experimental Design Skills. CSE Technical Report.
- Zimmerman, C., & Croker, S. (2014). A prospective cognition analysis of scientific thinking and the implications for teaching and learning science. *Journal of Cognitive Education and Psychology*, 13(2), 245–257. <https://doi.org/10.1891/1945-8959.13.2.245>
- Zimmerman, C., & Klahr, D. (2018). Development of scientific thinking. *Stevens' handbook of experimental psychology and cognitive neuroscience*, 4, 1-25. <https://doi.org/10.1002/9781119170174.epcn407>
- Zwack, W.R. and Velicer, W.F. (1986) Comparison of Five Rules for Determining the Number of Components to Retain. *Psychology Bulletin*, 99, 432-442. <https://doi.org/10.1037/0033-2909.99.3.432>

EKLER

EK-1. Madde Geliştirme Taslak Formu

Maddeyi geliştiren uzmanın adı soyadı	
Maddenin ait olduğu disiplin	
Hipotez	
Değişkenler	
	1.
	2.
	3.
	4.
	5.
	...
Kanıtlar	
	1.
	2.
	3.
	4.
	5.
	...

EK-2. Alan Uzmanları Madde Değerlendirme Kriterleri

Sayın Değerlendirici,

Bu değerlendirme formunda ana sınıfı, birinci sınıf ve ikinci sınıf öğrencilerinin bilimsel yaratıcılıklarını değerlendirmek amacıyla geliştirilen bilimsel yaratıcılık testinin 18 maddesi yer almaktadır. Testteki maddeler görseller ve olası yanıtlardan oluşmaktadır. Test maddeleri öğrencilerin kanıt değerlendirme becerisini ölçmeyi hedeflemektedir. Lütfen testte yer alan 18 görseli inceleyiniz ve yönergeyi dikkatlice okuyunuz. Size yöneltilen soruları cevaplayarak görüş ve önerilerinizi yazınız.

Araştırmaya sağlayacağınız katkı için teşekkür ederim.

Araştırmacı

Öğr. Gör. M. Arda ATAKAYA

Kanıt Değerlendirme: Hipotezin onaylanmasına ya da reddedilmesine hizmet eden bilgi olarak tanımlanmaktadır. Deneylerden elde edilen veriler ile mevcut hipotezin karşılaştırılması ve değerlendirilmesi sürecini kapsamaktadır.

Kanıt değerlendirme görevinde öğrenciye bir hipotez verilir. Farklı değişkenlerin yer aldığı hipoteze ilişkin kanıtlar sunulur. Öğrenciden kanıtları değerlendirerek hipotezi destekleyen değişkenlerin bulunması istenir.

Madde Nu.	...		
Modele uygunluk	Uygun ☐	Uygun Değil ☐	Düzeltilmeli ☐
Öneriler			
Bilimsellik	Doğru ☐	Doğru Değil ☐	Düzeltilmeli ☐
Öneriler			
Alan Bilgisi	Yeterli ☐	Yeterli Değil ☐	Düzeltilmeli ☐
Öneriler			
Anasınıfı, 1.ve 2. Sınıflara uygunluk	Uygun ☐	Uygun Değil ☐	Düzeltilmeli ☐
Öneriler			

EK-3. Araştırma Veli İzin Formu

Bu çalışma, “Çocuklar İçin Animasyonlu Bilimsel Yaratıcılık Testi Kanıt Değerlendirme Bileşeninin Geliştirilmesi ve Psikometrik Niteliklerinin İncelenmesi” başlıklı bir araştırma çalışması olup *bilimsel yaratıcılığı ölçmek için “Bilimsel Üretkenlik Testinin” geliştirilmesi* amacını taşımaktadır. Çalışma, Öğr. Gör. M. Arda ATAKAYA tarafından yürütülmektedir. Çalışmanın sonucunda *bilimsel yaratıcılığı (fen bilimleri alanını) değerlendiren bir ölçek* ortaya konacaktır.

- Araştırma kapsamında toplanan veriler, sadece bilimsel amaçlar doğrultusunda kullanılacak, araştırmanın amacı dışında ya da bir başka araştırmada kullanılmayacak ve başkalarıyla paylaşılmayacaktır.
- Çalışmada tablet üzerinden kısa animasyonlar ile çocuğunuzun fen bilimleri alanındaki yanıtları incelenecektir.
- Bu uygulamanın amacı kesinlikle çocuğunuzun değerlendirilmesi değil, sadece değerlendirme aracının niteliğini incelemektir.
- Ölçek geliştirilme aşamaları tamamlandığında gerekli görüldüğü takdirde çocuğunuzun fen bilimleri alanındaki yeterliği ile ilgili bilgi verilebilir.
- Çocuğunuzdan alınan veriler araştırmacı tarafından korunacak ve araştırma bitiminde arşivlenecek veya imha edilecektir.
- Çocuğunuz istediği zaman araştırmadan kendi isteği doğrultusunda ayrılabilir. Çocuğunuz çalışmadan ayrılması durumunda toplanan verileri çalışmadan çıkarılacak ve imha edilecektir.

Gönüllü katılım formunu okumak ve değerlendirmek üzere ayırdığınız zaman için teşekkür ederim. Çalışma hakkındaki sorularınızı Anadolu Üniversitesi Özel Eğitim bölümünden..... e-posta adresine yöneltebilirsiniz.

Araştırmacı : M. Arda ATAKAYA

İletişim No :

Cep No :

Danışman : Prof. Dr. Uğur SAK

Bu çalışmaya çocuğumun katılmasını, çocuğum istediği takdirde çalışmadan ayrılabilceğini bilerek çalışmadan elde edilecek bilgilerin bilimsel amaçlarla kullanılmasını:

Kabul ediyorum. ☐

Kabul etmiyorum. ☐

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

Veli Ad ve Soyadı:

İmza

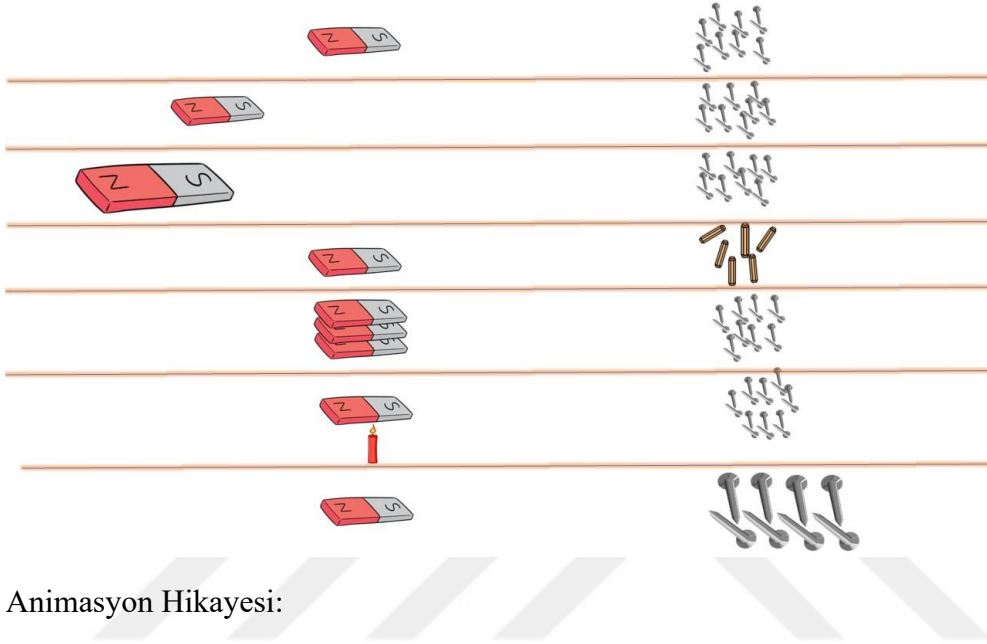
EK-4. ÇABİYAT-R Madde Havuzunda Yer Alan Örnek Madde ve Animasyon Hikayesi

Madde Adı: Mıknatıs

Ölçülen Beceri: Kanıt Değerlendirme

Maddenin Ait Olduğu Disiplin: Fizik

Madde Görseli:



Animasyon Hikayesi:

Ahmet merakla, "Anne, bak arabamın üzerinde duran bir nesne var, bu ne?"

Annesi gülerek, "Ahmet, o bir mıknatıs. Mıknatıs, bazı nesneleri kendisine doğru çekebilir."(hipotez)

Ahmet merakla, "Gerçekten mi? Nasıl yapar?"

Annesi, "Mıknatısın içinde minik manyetik alanlar vardır. Bazı nesneler bu alanlara çekilir. İstersen, bir deney yapalım!" dedi.

Ahmet sevinçle, "Vay, gerçekten işe yarıyor!"

Annesi gülümseyerek, "Evet, mıknatısın bu özelliği gerçekten de çok ilginç. Şimdi seninle farklı deneyler yapalım."

Annesi, "Ahmet senden nesnelerin hareketlerini dikkatlice izlemeni istiyorum."

Ahmet, "Anne, neden mıknatıs bazılarını çok çekerken, bazılarını az çekiyor ya da hiç çekmiyor?"

Annesi, "Mıknatısın nesneleri çekmesi bazı nedenlere bağlı Ahmet. Deney sonuçlarına dikkatlice bak ve bu nedenleri bulmaya çalış. (hipotezle uyumlu kanıtlar üretme)

Dış ses: "Sence mıknatıs, nesneleri neden farklı miktarlarda kendisine çekiyor olabilir? Unutma! Bu farklılıkların birden fazla nedeni var.

EK-5. Etik Kurul Kararı

	
ANADOLU ÜNİVERSİTESİ	
SOSYAL VE BEŞERÎ BİLİMLER BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU	
KARAR BELGESİ	
ÇALIŞMANIN TÜRÜ:	BAP Projesi-Doktora Tez Çalışması
KONU:	Eğitim Bilimleri
BAŞLIK:	ÇABIYAT Bilimsel Yaratıcılık Testi İçin Kanıt Değerlendirme Bileşeninin Geliştirilmesi ve Psikometrik Niteliklerinin İncelenmesi
PROJE/TEZ YÜRÜTÜCÜSÜ:	Prof. Dr. Uğur SAK
TEZ YAZARI:	Mehmet Arda ATAKEYA
ALT KOMİSYON GÖRÜŞÜ:	-
KARAR:	Olumlu

ÖZGEÇMİŞ

Adı Soyadı : Mehmet Arda ATAKAYA

Yabancı Dil : İngilizce

Eğitim Bilgileri:

- 2009 – 2013 İstanbul Üniversitesi, Hasan Ali Yücel Eğitim Fakültesi, Özel Eğitim Bölümü Lisans Programı
- 2015-2018 Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü, Üstün Zekalılar Öğretmenliği (YI) Tezli Yüksek Lisans Programı

Mesleki Deneyimi:

- 2015- devam ediyor Öğretim Görevlisi, Anadolu Üniversitesi ÜYEP Merkezi

Yayınları:

- Atakaya, M. A., Sak, U., & Ayas, M. B. (2022). A study on psychometric properties of creativity indices. *Creativity Research Journal*. Advance online publication. <https://doi.org/10.1080/10400419.2022.2134550>.
- Atakaya, M. A. ve Köprü, F. (2023). Yaratıcı düşünme becerilerinin geliştirilmesine yönelik stratejiler. B. B. Sezerel ve N. N. Ateşgöz (Ed.), *Üst düzey düşünme becerilerini geliştirme stratejileri* içinde (s. 51-86). Anı Yayıncılık.

Bilimsel Faaliyetleri (Kongre, Sempozyum, Çalıştay vb.):

- Atakaya, M. A. and Ayas, M. B. (2017). Perfectionism levels of gifted and talented students. 15th Internaitonal Conference on Excellence in Education (ICIE), oral presantation, 3-5 July, Lisboa, Portugual.
- Atakaya, M. A. and Ayas, M. B. (2018). Comparison of different creativity indexes obtained from Creativity Scientific Ability Test (C-SAT) and expert judges. The 15th Asia Pacific Conference on Giftedness (APCG), oral presantation, 20 -24 August, Bangkok, Tayland.
- Atakaya, M. A. (2019). Bilimsel Üretkenlik Testi (BÜT) ve uzman değerlendirmelerinden elde edilen farklı yaratıcılık endekslerinin karşılaştırılması. VI. Ulusal Üstün Yeteneklilerin Eğitimi Kongresi, sözlü bildiri, 10- 12 Ekim, İstanbul, Türkiye.

- Atakaya, M. A. ve Ayas, M. B. (2023). Yaratıcılığın ölçülmesinde altın standart: konsensüs değerlendirme tekniği. 8. Ulusal Üstün Yeteneklilerin Eğitimi Kongresi, sözlü bildiri, 2-4 Kasım, Ankara, Türkiye.
- Atakaya, M. A. (2023). ÖZEL yetenekli öğrenciler ve mentorluk: sistematik bir derleme. 8. Ulusal Üstün Yeteneklilerin Eğitimi Kongresi, sözlü bildiri, 2-4 Kasım, Ankara, Türkiye.
- Bal-Sezerel, B., Ateşgöz, N. N. ve Atakaya, M. A. (2024). Yaratıcılığın seri-sıra etkisi alana özgü yaratıcılıkta da işliyor mu? IX. Ulusal Üstün Yeteneklilerin Eğitimi Kongresi, sözlü bildiri, 6-8 Mayıs, Eskişehir, Türkiye.
- Atakaya, M. A. (2024). Yaratıcılığın ölçülmesinde kullanılan endekslerin karşılaştırılması. 9. Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (EPOC), sözlü bildiri, 3-6 Eylül, Eskişehir, Türkiye.
- Atakaya, M. A. (2025). Bilişsel süreç becerileri ile görsel ve sözel analogik muhakeme becerileri arasındaki ilişkinin incelenmesi. 10. Ulusal Üstün Yeteneklilerin Eğitimi Kongresi, sözlü bildiri, 2-4 Mayıs, İzmir, Türkiye.

Görev Aldığı Projeler:

- Atakaya, M. A. (Araştırmacı). Proje ASİS 2: Türkiye'nin ilk yerli zekâ ölçeği ASİS'in ikinci sürümünün standardizasyon çalışması. SBA-2024-30 nolu Anadolu Üniversitesi Bilimsel Araştırma Projesi, 2024-2026.