

T.C.
GAZİANTEP ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

SINIF İÇİ DEĞERLENDİRME YAKLAŞIMLARI
ENVANTERİNİN TÜRKÇE'YE UYARLANMASI VE
ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

LÜTFULLAH DİNÇER

GAZİANTEP
MAYIS 2023

T.C.
GAZİANTEP ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

SINIF İÇİ DEĞERLENDİRME YAKLAŞIMLARI
ENVANTERİNİN TÜRKÇE'YE UYARLANMASI VE
ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

LÜTFULLAH DİNÇER

Tez Danışmanı: Doç. Dr. Yeşim ÖZER ÖZKAN

GAZİANTEP
MAYIS 2023

TEZ ONAY SAYFASI**Öğrencinin Adı ve Soyadı:** Lütfullah DİNÇER**Üniversite:** Gaziantep Üniversitesi**Enstitü:** Eğitim Bilimleri Enstitüsü**Ana Bilim Dalı ve Bilim Dalı:** Eğitim Bilimleri Ana Bilim Dalı / Eğitimde Ölçme ve Değerlendirme Bilim Dalı**Tezin Başlığı:** Sınıf İçi Değerlendirme Yaklaşımları Envanterinin Türkçe'ye Uyarlanması ve Ölçme Değişmezliğinin İncelenmesi**Tezin Savunma Tarihi:** 02/05/2023

Bu tezin yüksek lisans tezi olarak gerekli şartları sağladığımı onaylıyorum.

Prof. Dr. Mehmet MURAT
Enstitü ABD Başkanı

Bu tez tarafımda okunmuş, kapsamı ve niteliği açısından bir yüksek lisans tezi olarak kabul edilmiştir.

Doç. Dr. Yeşim ÖZER ÖZKAN
Tez Danışmanı

Bu tez tarafımızca okunmuş, kapsam ve niteliği açısından bir yüksek lisans tezi olarak kabul edilmiştir.

Jüri Üyeleri

İmzası

Doç. Dr. Yeşim ÖZER ÖZKAN

Doç. Dr. İbrahim YILDIRIM

Doç. Dr. Meltem ACAR GÜVENDİR

Eğitim Bilimleri Enstitüsü Onayı

Doç. Dr. Ahmet İhsan KAYA
Eğitim Bilimleri Enstitüsü Müdürü

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde, bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu kabul ederim. Bu bilgiler doğrultusunda tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edilmesi halinde doğabilecek her türlü hukuki sorumluluğu onayladığımı beyan ederim.

İmza:

Adı ve Soyadı: Lütfullah DİNÇER
Öğrenci Numarası: 202617111001
Tezin Savunma Tarihi: 02/05/2023

ÖNSÖZ

Bu arařtırmada Sınıf İi Deęerlendirme Yaklařımları Envanteri Trke'ye uyarlanmıř ve envanterin cinsiyet gruplarında lme deęiřmezlięi incelenmiřtir. Elde edilen bulgular sonucunda envanter kullanılarak cinsiyet gruplarında geerli ve gvenilir lmler yapılabileceęi sonucuna varılmıřtır. Bu arařtırmanın gerekleřtirilmesinde doęrudan ve dolaylı katkıları olan kiřilere teřekkr ederim.

Yksek lisans eęitimim boyunca desteklelerini esirgemeyen, bilgilendirmeleriyle alana olan ilgi ve sevgimi artıran bařta deęerli hocam ve danıřmanım Do. Dr. Yeřim ZER ZKAN'a;

Sorduęum sorulara sabırla cevap veren, akademik olarak geliřmeme katkı sunan saygıdeęer hocalarım Do. Dr. İbrahim YILDIRIM ve Prof. Dr. Bayram ETİN'e;

Tez yazım srecinde desteklerini esirgemeyen, Sınıf İi Deęerlendirme Yaklařımları Envanteri'nin geliřtiricilerinden Prof. Dr. Christopher DELUCA ve Dr. Andrew COOMBS'a;

Tez uyarlama ařamasında grřleriyle tezimin geliřimine katkı sunan deęerli hocalarım Do. Dr. Funda NALBANTOęLU YILMAZ, Do. Dr. Hatice KUMANDAŐ ZTRK, Do. Dr. Meltem ACAR GVENDİR, Prof. Dr. Adnan ERKUŐ ve Prof. Dr. Selahattin TURAN'a;

Sabırla sorularıma cevap veren ve zm bulan deęerli yksek lisans sınıf arkadařlarıma ve Arř. Gr. Tuęberk MELEN'e;

Yksek lisans eęitimime ve tezime katkı sunan kıymetli ęretmen arkadařlarıma ve saygıdeęer okul mdrm Emre EMİROęLU'na;

Hayatıma anlam katan biricik aileme;

sonsuz sevgi ve saygılarımı sunuyorum.

Tezimi 6 Őubat Kahramanmarař depremlerinde hayatlarını kaybeden yeęenlerim Gizem ve Beyza'ya atfediyorum.

*Gizem ve Beyza'ya
Huzur iinde uyuyun kuzularım.*

Ltfullah DİNER
Mayıs, 2023

ÖZET

SINIF İÇİ DEĞERLENDİRME YAKLAŞIMLARI ENVANTERİNİN TÜRKÇE'YE UYARLANMASI VE ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ

DİNÇER, Lütfullah

Yüksek Lisans Tezi,

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Tez Danışmanı: Doç. Dr. Yeşim ÖZER ÖZKAN

Mayıs-2023, 145 sayfa

Bu araştırmada Sınıf İçi Değerlendirme Yaklaşımları Envanteri (*Approaches to Classroom Assessment Inventory, ACAI*) Türkçe'ye uyarlanmış ve envanterin cinsiyete göre ölçme değişmezliği incelemiştir. Araştırmanın çalışma grubunu 154'ü erkek ve 251'i kadın olmak üzere toplam 405 öğretmen oluşturmaktadır. Uyarlama için öncelikle envanteri geliştiren araştırmacılardan izin alınmıştır. Envanterin çevirisi her iki dile hakim iki çevirmen ve iki ölçme değerlendirme uzmanı tarafından yapılmıştır. Dil eşdeğerliğinin sağlanması amacıyla uzman görüşleri alınmış, envanterin Türkçe ve İngilizce formları 15 İngilizce öğretmenine iki hafta ara ile uygulanmıştır. Dil eşdeğerliğinin sağlanmasının ardından ise envanterin deneme uygulaması ve son uygulaması yapılmıştır. Envanterin güvenilirlik analizlerine ilişkin Cranbach α ve McDonald's ω katsayıları hesaplanmıştır. Son uygulamadan elde edilen verilerle sırasıyla varsayımların kontrolü, madde analizleri, ki-kare analizi ve doğrulayıcı faktör analizleri gerçekleştirilmiştir. Uyarlama sürecinden sonra ise envanterin cinsiyete göre ölçme değişmezliği ÇGDFA ile incelenmiştir. Analiz sonuçlarına göre envantere yer alan ölçeklerin tüm ölçme değişmezliği aşamalarını sağladıkları görülmüştür. Sonuç olarak Sınıf İçi Değerlendirme Yaklaşımları Envanteri kullanılarak cinsiyet gruplarında geçerli ve güvenilir ölçümler ve karşılaştırmalar yapılabileceği ifade edilebilir.

Anahtar Sözcükler: Öğretmen Değerlendirme Okuryazarlığı, Sınıf İçi Değerlendirme, Öğretmen Yeterliliği, Ölçek Uyarlama, Ölçme Değişmezliği

ABSTRACT**ADAPTATION OF APPROACHES TO CLASSROOM ASSESSMENT
INVENTORY INTO TURKISH AND ANALYZING MEASUREMENT
INVARIANCE**

DİNÇER, Lütfullah

MA Thesis,

Department of Educational Sciences

Discipline of Measurement and Evaluation in Education

Supervisor: Assoc. Prof. Dr. Yeşim ÖZER ÖZKAN

May-2023, 145 pages

In this study, the Approaches to Classroom Assessment Inventory (ACAI) was adapted into Turkish language and examined its measurement invariance across the gender groups. The study group of the research consists of a total of 405 teachers, 154 male and 251 female. Initially, permission for the adaptation was obtained and the translation of the inventory was carried out by two translators fluent in both languages and two assessment and evaluation experts. For language equivalence; expert opinions were taken, and the Turkish and English forms of the inventory were administered to 15 English teachers with a two-week interval. After the linguistic equivalence was achieved, the pilot study and final application of the inventory were conducted. Cronbach's α and McDonald's ω coefficients were calculated for the reliability analysis of the inventory. Control of assumptions, item analysis, chi-square analysis, CFA were performed with the data. After the adaptation process, measurement invariance of the inventory across the gender groups was analyzed with MGCFA. According to the results, the measurement invariance was established for the scales in the inventory across the gender groups. Consequently, it can be concluded that valid and reliable measurements and comparisons for gender groups can be carried out by using the Approaches to Classroom Assessment Inventory (ACAI).

Keywords: Teacher Assessment Literacy, Classroom Assessment, Teacher Competence, Test Adaptation, Measurement Invariance

İÇİNDEKİLER

	Sayfa
TEZ ONAY SAYFASI	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
ÖNSÖZ	iii
ÖZET.....	iv
ABSTRACT	v
İÇİNDEKİLER	vi
TABLolar LİSTESİ	viii
ŞEKİLLER LİSTESİ	ix
EKLER LİSTESİ	x
KISALTMALAR	xi

BÖLÜM I

GİRİŞ

1.1. Problem Durumu	1
1.2. Araştırmanın Amacı	4
1.3. Araştırmanın Önemi	4
1.4. Sayıltılar	5
1.5. Sınırlılıklar	5
1.6. Tanımlar	5

BÖLÜM II

KAVRAMSAL ÇERÇEVE

2.1. Eğitimde Ölçme ve Değerlendirme.....	7
2.1.1. Eğitimde Ölçme	7
2.1.2. Eğitimde Değerlendirme	9
2.2. Değerlendirme Okuryazarlığı	12
2.2.1. Değerlendirme Okuryazarı Olmanın Önemi	15
2.2.2. Ölçme Değerlendirme Standartları	17
2.2.3. Değerlendirme Okuryazarlığı Modelleri	21
2.2.4. Değerlendirme Okuryazarlığı ile İlgili Geliştirilen Ölçme Araçları	26
2.3. Konu İle İlgili Ulusal ve Uluslararası Araştırmalar ve Genel Değerlendirme	29
2.3.1. Ulusal Düzeyde Yapılmış Araştırmalar	29
2.3.2. Uluslararası Düzeyde Yapılmış Araştırmalar	38
2.4. Psikolojik Ölçme Araçlarının Uyarlanması	46
2.4.1. Ölçek Uyarlama Adımları	49
2.4.2. Uluslararası Test Komisyonu (ITC).....	53
2.5. Yapısal Eşitlik Modellemesi (YEM).....	56
2.5.1. Uyum İndeksleri.....	58
2.5.2. Varsayımlar	62
2.6. Ölçme Değişmezliği.....	65
2.6.1. Doğrulayıcı Faktör Analizi (DFA).....	66
2.6.2. Çok Gruplu Doğrulayıcı Faktör Analizi (ÇGDFA)	67

2.6.3. Ölçme Değişmezliği Aşamaları	vii 68
---	-----------

BÖLÜM III YÖNTEM

3.1. Araştırma Modeli	71
3.2. Çalışma Grubu	71
3.3. Veri Toplama Aracı.....	72
3.3.1. Envanterin Uyarlanma Adımları.....	76
3.4. Verilerin Toplanması	77
3.5. Verilerin Analizi	77
3.5.1. Varsayımların Test Edilmesi.....	77

BÖLÜM IV BULGULAR

4.1. Dil Eşdeğerliği ve Deneme Uygulaması	81
4.2. Madde Analizleri.....	84
4.3. Güvenirlilik Analizi.....	85
4.4. Ki Kare Analizi	86
4.5. DFA Bulguları.....	87
4.6. Ölçme Değişmezliği Bulguları.....	91

BÖLÜM V TARTIŞMA

5.1. Araştırma Bulgularının Tartışılması	94
--	----

BÖLÜM VI SONUÇ VE ÖNERİLER

6.1. Sonuçlar.....	100
6.2. Öneriler	102

KAYNAKLAR.....	103
EKLER.....	113
ÖZGEÇMİŞ.....	145
VITAE	145

TABLULAR LİSTESİ

	Sayfa
Tablo 1. Sınıf içi ve geniş ölçekli değerlendirme	11
Tablo 2. Değerlendirme türleri	12
Tablo 3. Ölçme değerlendirme standartları	18
Tablo 4. Sınıf içi değerlendirme standartları	20
Tablo 5. Değerlendirme okuryazarlığı ile ilgili geliştirilen ölçme araçları	27
Tablo 6. YEM’de kullanılan şekil ve semboller	57
Tablo 7. Uyum indekslerinin sınıflandırılması.....	59
Tablo 8. Uyum ölçütleri.....	61
Tablo 9. Ölçme değişmezliği aşamaları.....	69
Tablo 10. Değerlendirme temaları ve alt boyutlar	72
Tablo 11. Değerlendirme yaklaşımlarının temasal dağılımı.....	74
Tablo 12. Tek değişkenli normallik analizi sonuçları	79
Tablo 13. Çok değişkenli normallik analizi sonuçları	79
Tablo 14. Uzman görüşleri arasındaki uyum indeksleri	82
Tablo 15. Uzman görüşleri arasındaki uyum indeksleri	82
Tablo 16. Dil eşdeğerliği korelasyon değerleri.....	83
Tablo 17. Madde-toplam test korelasyonları	84
Tablo 18. Güvenirlik analizi sonuçları.....	85
Tablo 19. Ki kare analizi sonuçları	86
Tablo 20. Cinsiyete göre ölçme değişmezliği bulguları	92

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 1. Ölçme.....	8
Şekil 2. Değerlendirme	10
Şekil 3. Fen bilimleri öğretmenleri için değerlendirme okuryazarlığı modeli	21
Şekil 4. Uygulamada değerlendirme okuryazarlığı modeli.....	22
Şekil 5. Üç boyutlu değerlendirme okuryazarlığı modeli	24
Şekil 6. Bütüncül yeterliliklerde öğretmen değerlendirme okuryazarlığı modeli	25
Şekil 7. YEM için geleneksel yaklaşım	58
Şekil 8. DFA modeli	67
Şekil 9. Sınıf içi değerlendirmede güven ölçeği DFA modeli	87
Şekil 10. Değerlendirmeye ilişkin mesleki öğrenme öncelikleri ölçeği	89
Şekil 11. Değerlendirmeye ilişkin mesleki öğrenme tercihleri ölçeği	90

EKLER

	Sayfa
Ek 1. Uyarlama İçin Alınan İzin Örneği	113
Ek 2. Çalışma Grubu.....	114
Ek 3. Ölçek Maddelerine Ait VIF ve Tolerans Değerleri	116
Ek 4. Dil Eşdeğerliği Verilerine Ait Normallik Analizi Sonuçları	117
Ek 5. Deneme Uygulamasına Ait Betimsel İstatistikler.....	118
Ek 6. Uygulamaya Ait Betimsel İstatistikler.....	119
Ek 7. Ölçek Maddelerine Ait Korelasyon Değerleri.....	120
Ek 8. Analizlerde Kullanılan MPlus Syntaxları.....	121
Ek 9. Envanterin Orjinal Formu.....	124
Ek 10. Envanterin Türkçe'ye Uyarlanmış Formu.....	131
Ek 11. Envanter Kullanıcı Rehberi	138
Ek 12. Etik Kurul Onay Yazısı	144

KISALTMALAR

ACAI	: Approaches to Classroom Assessment Inventory
AFT	: American Federation of Teachers
AGFI	: Adjusted Goodness of Fit Index
ALI	: Assessment Literacy Inventory
API	: Assessment Practices Inventory
AVCQ	: Assessment in Vocational Classroom Questionnaire
CAK	: Classroom Assessment Knowledge Test
CALI	: Classroom Assessment Literacy Inventory
CFI	: Comparative Fit Index
ÇG DFA	: Çok Gruplu Doğrulayıcı Faktör Analizi
DFA	: Doğrulayıcı Faktör Analizi
DTUG	: Değerlendirme Tasarımı, Uygulama ve Geribildirim
DUBİ	: Değerlendirme Uygulamalarının Bütünleştirilmesi/İletimi
GLS	: Generalized Least Squares
GFI	: Goodness of Fit Index
ITC	: International Test Commission
İADSI	: İzleme, Analiz Etme ve Değerlendirme Sonuçlarını İletme
MDTİU	: Mevcut Değerlendirme Teorisi, İlkeleri ve Uygulamalarla Uyum
MKS	: Measurement Knowledge Scale
ML	: Maximum Likelihood
MLM	: Maximum Likelihood parameter estimates with standart errors and a mean-adjusted chi-square test statistic
MSS	: Measurement Self Efficacy Scale
NCME	: The National Council on Measurement in Education
NEA	: National Education Association
NFI	: Normed Fit Index
NNFI	: Non-normed Fit Index
PGFI	: Parsonomy Goodness of Fit Index
RMSEA	: Root Mean Square Error Approximation
RMR	: Root Mean Square Residual
TALQ	: Teacher Assessment Literacy Questionnaire
TEDP	: Temel Eğitime Destek Programı
TLI	: Tucker-Lewis Index
ULS	: Unweighted Least Squares
WLS	: Weighted Least Squares
YEM	: Yapısal Eşitlik Modellemesi

BÖLÜM I

GİRİŞ

1.1. Problem Durumu

Eğitim, insan hayatına yön veren ve böylece toplumu biçimlendiren dinamik bir yapıdır. Bu yapının ana hatlarını girdiler, süreç ve çıktılar oluşturur. Bu öğelerin denetlenmesi ise eğitimde ölçme ve değerlendirme alanının kapsamındadır. Ölçme ve değerlendirmeyeyle eğitim paydaşları eğitim programının verimini, eğitimde kullanılan öğretim yöntem ve tekniklerin etkinliğini, öğrenci başarısını ve öğrencilerin eksik öğrenmelerini analiz eder ve ihtiyaç duyulan düzenlemeleri yapma fırsatı bulur (Şata, 2016). Dolayısıyla eğitim sürecinin sürekli olarak kendini yenilemesinde ölçme ve değerlendirme önemli bir rol üstlenmektedir. Ölçme ve değerlendirmenin bu önemli rolü düşünüldüğünde kullanılan ölçme aracı ve bu aracı kullanan kişinin yetkinliği önem kazanmaktadır. Bu yetkinlik durumu ise alanyazında *değerlendirme okuryazarlığı* olarak isimlendirilmiştir.

Stiggins (1991) değerlendirme okuryazarlığını bir öğretmenin yüksek ve düşük kaliteli ölçme ve değerlendirmeleri kavrayabilme ve bu bilgiyi öğrenci başarısını artırmak için kullanabilme yeteneği olarak tanımlamıştır. Benzer şekilde Webb (2002) de değerlendirme okuryazarlığını; öğrencilerin ne bildiği, bunun nasıl değerlendirileceği ve hem öğrencilerin başarısını hem de müfredatı geliştirme amacıyla sonuçların nasıl analiz edileceği bilgisi olarak açıklamıştır. Sonuç olarak bir test sonucunun istatistiksel anlamını bilmenin değerlendirme okuryazarı olabilmek için yeterli olmadığı; ayrıca elde edilen ölçme sonuçlarının anlamlı bir şekilde nasıl yorumlanacağını ve bu durumun eğitim ve öğretime nasıl katkı sunacağını bilmesi gerektiği anlaşılmaktadır.

Popham (2004) ise değerlendirme okuryazarlığı kavramına büyük önem vermiş ve değerlendirme okuryazarı olmama durumunu profesyonel intihar olarak tanımlamıştır. Araştırmacılar Popham'ın bu iddiasını göz önünde bulundurarak, öğretmenlerin değerlendirme okuryazarlık düzeylerini ve bu düzeyleri etkileyen değişkenleri ölçmek amacıyla geçerliği kanıtlanmış çeşitli ölçme araçları geliştirmiştir (Mertler, 2003; Mertler & Campbell, 2005; Plake, Impara & Fager, 1993; Tao, 2014). Geliştirilen bu ölçme araçları 1990 yılında belirlenen ölçme değerlendirme standartlarına göre yapılandırılmıştır (American Federation of Teachers, National Council on Measurement in Education & National Education Association, 1990).

Brookhart (2011) ise 1990 yılında yayımlanan ölçme değerlendirme standartlarını modası geçmiş olarak nitelendirmiş ve güncellenmesi gerektiğini belirterek eleştirmiştir. Yazara göre bu standartlar, öğretmenlerin son yirmi yılda yeniden şekillenen ve öğrenme için değerlendirme anlamına gelen biçimlendirici değerlendirme ihtiyaçlarını yansıtmamaktadır. Brookhart'ın (2011) vurguladığı bir diğer nokta ise, 1990 standartlarının hesap verebilirlik olgusundaki ve standart tabanlı testlerde yaşanan değişimlere uygun olmamasıdır. Hem Brookhart (2011) hem de Gotch ve French (2014), 1990 standartlarının öğretmenlerin mevcut değerlendirme bilgilerinin çok gerisinde olduğunu belirtmektedir. Ayrıca Gotch ve French (2014) araştırmalarında, 1991'den 2012'ye kadar değerlendirme okuryazarlığı ile ilgili yapılan araştırmaları analiz etmişler ve bu araştırmalarda kullanılan ölçme araçlarının, değerlendirme bilgisindeki dönüşüme uygunluğunun (hesap verebilirlik sistemleri, biçimlendirici değerlendirme kavramları vb.) ve temsiliyetinin düşük düzeyde olması nedeniyle zayıf psikometrik özellikler gösterdiklerini iddia etmişlerdir (s. 17).

Türkiye'de değerlendirme okuryazarlığı ile ilgili yapılan araştırmalarda (Acar Güvendir ve Özer Özkan, 2021; Azrak ve Yalçinkaya, 2019; Bütüner, Çimer ve Yiğit, 2010; Ergül, 2019; Gül, 2011; Karaman & Şahin, 2014) çoğunlukla 1990 standartlarına göre geliştirilen ve dilimize uyarlanan ölçme araçlarının kullanıldığı görülmektedir. 1990 standartlarına yöneltilen eleştiriler ve bu standartlara göre geliştirilen ölçme araçlarının zayıf psikometrik özellikler gösterdiğinin öne sürülmesi nedeniyle, değerlendirme okuryazarlığının ölçülmesinde kullanılacak yeni bir ölçme aracına ihtiyaç duyulduğu görülmektedir. Sınıf İçi Değerlendirme Envanteri'nin (*Approaches to Classroom Assessment Inventory, ACAI*) yazarları Deluca, LaPointe-

McEwan ve Luhanga (2016a) bu ihtiyaç doğrultusunda ölçme araçlarını Sınıf İçi Değerlendirme Standartları'nı (Klinger vd., 2015) temel alarak ve altı farklı bölgede (Amerika Birleşik Devletleri, Avrupa, Avusturalya, Birleşik Krallık, Kanada ve Yeni Zelanda) tanımlanan ölçme değerlendirme standartlarını inceleyerek geliştirmiştir.

Sınıf İçi Değerlendirme Standartları'nda (Klinger vd., 2015) öğretmenler tarafından öğrencilerin etkili bir şekilde değerlendirilmesi için ilke ve yönergeler sunmak amaçlanmaktadır. 1990 standartlarıyla karşılaştırıldığında Sınıf İçi Değerlendirme Standartları öğretmenlerin ölçme ve değerlendirme yeterliliklerini daha kapsamlı bir şekilde tanımlamaktadır. Bu nedenle bu araştırmada DeLuca, LaPointe-McEwan ve Luhanga (2016b) tarafından geliştirilen "Approaches to Classroom Assessment Inventory (ACAI)" Türkçe'ye "Sınıf İçi Değerlendirme Yaklaşımları Envanteri" ismi ile uyarlanmış ve ölçme aracının cinsiyete göre ölçme değişmezliği incelenmiştir. Araştırmayla ilgili belirlenen problemler ve ilgili alt problemler sırasıyla sunulmuştur.

Araştırma problemleri:

1. Türkçe'ye uyarlaması yapılan Sınıf İçi Değerlendirme Yaklaşımları Envanteri'nin güvenilirlik kanıtları nelerdir?
2. Türkçe'ye uyarlaması yapılan Sınıf İçi Değerlendirme Yaklaşımları Envanteri'nin geçerlilik kanıtları nelerdir?
 - a. Sınıf İçi Değerlendirme Yaklaşımları Envanterinden elde edilen puanlar cinsiyete göre ölçme değişmezliği aşamalarından yapısal değişmezliği sağlamakta mıdır?
 - b. Sınıf İçi Değerlendirme Yaklaşımları Envanterinden elde edilen puanlar cinsiyete göre ölçme değişmezliği aşamalarından metrik değişmezliği sağlamakta mıdır?
 - c. Sınıf İçi Değerlendirme Yaklaşımları Envanterinden elde edilen puanlar cinsiyete göre ölçme değişmezliği aşamalarından ölçek değişmezliğini sağlamakta mıdır?
 - d. Sınıf İçi Değerlendirme Yaklaşımları Envanterinden elde edilen puanlar cinsiyete göre ölçme değişmezliği aşamalarından katı değişmezliği sağlamakta mıdır?

1.2. Araştırmanın Amacı

Araştırmanın amacı DeLuca vd. (2016b) tarafından geliştirilen Sınıf İçi Değerlendirme Yaklaşımları Envanterini (*Approaches to Classroom Assessment Inventory, ACAI*) Türkçe'ye uyarlamak ve envanterin cinsiyete göre ölçme değişmezliğini incelemektir. Bu amaç doğrultusunda envanterin öncelikle Türkçe'ye uyarlaması yapılmıştır. Uyarlama işleminden sonra envanterin ölçme değişmezliği sınanmıştır. Ölçme değişmezliğinin elde edilmesiyle ölçme aracının psikometrik özelliklerinin uygulandığı gruplarda aynı olması sağlanır ve böylece elde edilen ölçümlerin anlamlı olduğu kabul edilir. Farklı kültürlerden yapılan çevirilerde kültürler arası genellenebilirliğin yapılabilmesi ya da karşılaştırılması için de ölçme değişmezliğinin test edilmesi gerekmektedir (Önen, 2009). Benzer şekilde aynı kültür grubundan gelen katılımcılara uygulanan ölçme araçlarının psikometrik özelliklerinin cinsiyete, sosyo-ekonomik duruma vb. durumlara göre ölçme değişmezliği sınanır. Bu araştırmada benzer kültür grubundan gelen katılımcıların cinsiyete göre ölçme değişmezliğinin sınanması amaçlanmıştır.

1.3. Araştırmanın Önemi

Ölçme ve değerlendirme uygulamaları eğitimde hedeflenen çıktıların elde edilmesinde kilit rol üstlenir. Öğretmenler ölçme ve değerlendirme uygulamalarıyla öğrencilerin başarısını izler, onların eksik öğrenmelerini belirler ve ihtiyaç duydukları düzenlemeleri yapar. Ölçme ve değerlendirmenin eğitim sistemindeki bu önemli rolleri göz önüne alındığında, öğretmenlerin bu konuda yetkin olmasının ne kadar hayati olduğu ortaya çıkmaktadır. Değerlendirme okuryazarlığı kavramı bu noktadan doğmuştur ve önem kazanmıştır. Alanyazın incelendiğinde öğretmenlerin değerlendirme okuryazarlığıyla ilgili yapılmış birçok çalışmaya rastlanılmaktadır (Azrak ve Yalçınkaya 2019; Bütüner vd., 2010; Ergül, 2019; Gül, 2011). Fakat bu çalışmaların çoğunluğunda 1990 standartlarına göre geliştirilmiş ölçme araçlarının kullanıldığı görülmektedir. 1990 standartlarını Brookhart'ın (2011) modası geçmiş olarak nitelendirmesi ve Gotch ve French'in (2014) bu standartlara göre geliştirilmiş ölçme araçlarının zayıf psikometrik özellikler gösterdiğini iddia etmesi, 1990 standartlarını temel alarak geliştirilen ölçme araçlarından elde edilen ölçümlerin

geçerliđi üzerinde řüphe oluřturmaktadır. Sonuç olarak, öđretmenlerin deđerlendirme okuryazarlıklarının belirlenmesinde kullanılabilir ve geçerliđi kanıtlanmış yeni ölçme araçlarına ihtiyaç duyulduđu anlaşılmaktadır. Bu arařtırmada ise DeLuca vd. (2016b) tarafından geliřtirilen “Approaches to Classroom Assessment Inventory (ACAI)” Türkçe’ye “Sınıf İçi Deđerlendirme Envanteri” ismi ile uyarlanmış ve envanterin cinsiyete göre ölçme deđiřmezliđi incelenmiřtir. Böylece öđretmenlerin deđerlendirme okuryazarlıklarının belirlenmesinde kullanılabilir yeni bir ölçme aracı alanyazına kazandırılmıştır. Bu ölçme aracı vasıtasıyla öđretmenlerin ölçme ve deđerlendirmede eksik ya da zayıf kaldıkları alanlar belirlenebilecek ve öđretmenlerin bu konudaki eksiklerini gidermek amacıyla ihtiyaç duyulan eđitimlerin planlaması ilgililer tarafından yapılabilecektir.

1.4. Sayılılar

Arařtırmada katılımcıların ölçme aracına ait maddeleri içtenlikle cevapladıkları varsayılmıştır.

1.5. Sınırlılıklar

Arařtırmanın katılımcılarını 2022-2023 yılları arasında Milli Eđitim Bakanlığı’na bađlı okullarda çalıřan öđretmenler oluřturmaktadır.

1.6. Tanımlar

Deđerlendirme okuryazarlıđı: Eđitimcilerin eđitimle ilgili kararlarını etkileyen temel deđerlendirme kavramlarını ve süreçlerini bilmesidir (Popham, 2011).

Ölçek Uyarlama: Erkuř (2019) uyarlamayı,

bir psikolojik özelliđi ölçmek amacıyla geliřtirilmiş ve sađlam olduđu varsayılan bir ölçeđin, aynı dil ve kültürde deđiřen zamanlar ile başka demografik gruplar, aynı dil fakat farklı kültürler veya farklı dil ve farklı kültürler için irdelenip kullanıma uygun hale getirilmesi olarak (s. 4) tanımlamıştır.

Ölçme Deđermezliđi: Ölçme aracında yer alan maddelerin tüm gruplarda aynı řekilde anlaşılması ve yorumlanmasıdır (Byrne & Watkins, 2003). Ölçme

değişmezliđi bir özelliđin farklı kořullarda ölçülmesinden elde edilen puanların aynı anlama gelip gelmediđiyle ilgilidir (Meade & Lautenschlager, 2004).



BÖLÜM II

KAVRAMSAL ÇERÇEVE

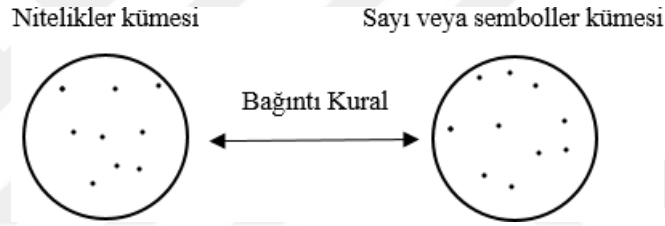
2.1. Eğitimde Ölçme ve Değerlendirme

İnsanlık var olduğu ilk günden beri yaşadığı ortamı gözlemlemiş, analiz etmiş ve karşılaştığı durumlara uyum sağlama çabası içinde olmuştur. Hayatta kalma güdüsü insanı sürekli değişen çevresine uyum sağlamaya itmiştir. Her bir birey bu değişim ve uyum sürecinden farklı şekillerde etkilenmiştir. İlk çağ filozoflarından Herakleitos'un akış teorisinde bu durum "*Aynı ırmaklara girenlerin üzerinden farklı sular akar*" (Herakleitos, 2005, s. 53, Fr. 12 akt. Yeşil, 2017) sözüyle vurgulanmıştır. Farklı şekillerde yorumlanan bu sözde ırmak bir sabittir, fakat ırmaktan akan sular kişiden kişiye değişim gösterir (Yeşil, 2017). Herakleitos'un bu yaklaşımı eğitim sistemi için düşünüldüğünde ırmak eğitim sisteminin kendisini, ırmaktan akan sular ise eğitimde girdiler ve süreç kısmını temsil eder. Peki ırmağın kişiler üzerinde yarattığı değişim - eğitimin çıktıları- nasıl belirlenecektir? Bu noktada ise eğitim sisteminin kontrol organı ölçme ve değerlendirme öne çıkmaktadır. Ölçme ve değerlendirme ırmakta ıslanan her bir öğrencinin istendik yönde değişim gösterip göstermediğini belirlemede rol alır. Dolayısıyla eğitimin sigortası konumunda olan ölçme ve değerlendirmenin eğitim paydaşlarınca ayrıntılı olarak bilinmesi ve uygulanması önemlidir.

2.1.1. Eğitimde Ölçme

Ölçme birçok araştırmacı tarafından farklı şekillerde tanımlanmıştır. Örneğin Erkuş (2021) ölçmeyi "*ilgilenilen niteliklerin (özelliklerin), amaca, araca ve olanaklara bağlı olarak nicelleştirilmesi (sayısallaştırılması) çabası, işlemi ve süreci*"

olarak açıklamıştır (s.7). Baykul (2015) ölçmeyi “*gerçek dünyanın elemanlarını sayı ve sembollerle ifade eden bir model*” olarak tanımlamıştır (s.86). Özgüven (2022) ise ölçmeyi “*herhangi bir objenin belirli bir niteliğini, belirli bir kurala göre, sayarak, sınıflandırarak, derecelendirerek ya da birimlerle sayısal olarak ifade etme süreci*” olarak tarif etmiştir (s.36). Bu tanımlamalardan yola çıkılarak ölçmenin bir gözlem süreci olduğu, gözlem sonucunda gözlemlenen özelliklerin sayı ya da semboller kullanılarak betimlendiği anlaşılmaktadır. Baykul (2015) bu süreci nitelikler kümesi ile sayı ve semboller kümesi arasında kurulan bir fonksiyonel bağıntı olarak tanımlamıştır. Bu fonksiyonel bağıntıda nitelikler kümesinde yer alan her bir özellik bağıntı(ölçme) kuralı kullanılarak sayı ve semboller kümesinde elemanlarla eşleştirilir. Bu bağıntı Şekil 1’de sunulmuştur.



Şekil 1. Ölçme (Baykul, 2015, s.86)

Nitelikler kümesi ölçülmesi amaçlanan özelliklerin yer aldığı kümedir. Bu kümeden bir ölçme kuralı vasıtasıyla ilgili özellikler sayı veya sembole eşleştirilerek ölçme işlemi gerçekleştirilir. Ölçme işleminin ilk adımında nitelikler kümesinin gözlem yoluyla oluşturulması yer alır. Nitelikler kümesinde yer alan her bir değişkenin kendine ait özellikleri farklı ölçme işlemlerinin yapılmasını gerektirir. Örneğin, bir insanın kişilik özelliklerinin ölçülmesi ile boy uzunluğunun ölçülmesi birbirinden farklı ölçme türleridir. Ölçme doğrudan, dolaylı ve türetilmiş olmak üzere üçe ayrılır. Ölçme türlerine ilişkin aşağıda kısa bir bilgiye yer verilmiştir.

1. *Doğrudan Ölçme*: Ölçülmesi hedeflenen özelliğin başka bir değişken kullanılmadan doğrudan ilgili ölçme aracı kullanılarak ölçme işleminin gerçekleştirilmesidir (Özer Özkan, 2019). Boy uzunluğunun metre ile ölçülmesi, ağırlığın terazi ile ölçülmesi doğrudan ölçmeye örnektir.

2. *Dolaylı Ölçme*: Ölçülmek istenen özelliğin bir başka özellik yardımıyla ölçülmesidir (Güler, 2015). Sıcaklığın termometreyle ölçülmesi, eğitim bilimlerinde ders başarısının sınavlarla ölçülmesi dolaylı ölçmeye örnektir. Örneğin ders başarısı doğrudan gözlemlenemez fakat öğrencinin verdiği doğru yanıtlar üzerinden ders başarısı ölçülebilir.
3. *Türetilmiş Ölçme*: Ölçülmek istenen özelliğin birden fazla özellik yardımıyla matematiksel işlemler kullanılarak ölçülmesidir (Özer Özkan, 2019). Örneğin, bir aracın hızının belirlenmesinde aracın belirli bir zamanda gittiği yol dikkate alınır. Aracın gittiği yolun zamana bölünmesiyle aracın hızı bulunur.

2.1.2. Eğitimde Değerlendirme

Ölçme gözlem yapılarak özelliklerin sayı veya sembollerle ifade edilmesidir. Değerlendirme ise ölçme sonuçlarının bir ölçütle karşılaştırılarak yorumlanması, yargıya varılması veya karar verilmesidir (Özer Özkan, 2019). Örneğin, bir öğrencinin matematik sınavından 60 alması bir ölçme işlemidir. Matematik dersinden 60 alan öğrencinin başarılı sayılması ise değerlendirmedir. Değerlendirme kullanılan ölçüte ve amaca göre ikiye ayrılır (Özer Özkan, 2019).

Ölçüte Göre Değerlendirme: Kullanılan ölçüte göre mutlak ve bağıl değerlendirme olarak ikiye ayrılır.

- a. *Mutlak Değerlendirme*: Mutlak ölçüte göre yapılan değerlendirmedir. Ders geçme notunun 60 olduğu bir sınavda, 70 alan kişinin dersi geçmesi mutlak değerlendirmeye örnektir.
- b. *Bağıl Değerlendirme*: Bağıl ölçüte göre yapılan değerlendirmedir. Ders geçme kalma kararının grubun not ortalamasına göre belirlenmesi bağıl değerlendirmeye örnektir.

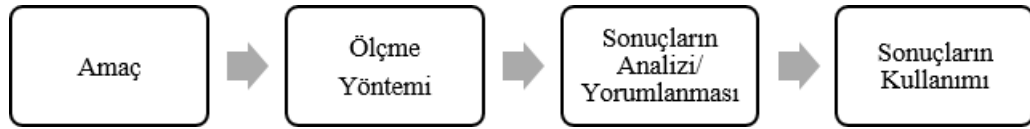
Amaca Göre Değerlendirme: Amaca göre değerlendirme eğitim öğretimin başında, süreç boyunca ve sonunda olmak üzere üç başlıkta sınıflandırılabilir.

- a. *Tanımayı-Yerleştirmeye Yönelik Değerlendirme*: Eğitim öğretim dönemi başında öğrencilerin derse hazırbulunuşluk seviyelerinin tespiti amacıyla yapılan değerlendirmedir (Doğan, 2020). Örneğin, yabancı dil seviye tespit

sınavları yapılarak öğrencilerin seviye sınıflarına ayrılması tanımayaylaştırmeye yönelik değerlendirmeye örnektir.

- b. *Biçimlendirici (Formatif) Değerlendirme*: Eğitim süreci boyunca belirli zaman aralıklarıyla öğrencilerin kısa sınav, ünite testi gibi ölçme araçları kullanılarak değerlendirilmesidir. Temel amaç öğrencilerin öğrenmeleri hakkında geri bildirim almak ve böylelikle eğitim sürecini öğrencilerin eksiklerini giderecek şekilde düzenlemektir. Biçimlendirici değerlendirmede öğrencilere not verme amacı güdülmez (Güler, 2015).
- c. *Değer Biçici (Summatif) Değerlendirme*: Eğitim süreci sonunda öğrenme hedeflerinin ne kadarının öğrenciler tarafından başarıldığına karar verme amacıyla yapılır. Ünite sonu testi, finaller, proje ödevi değer biçici değerlendirme örnekleridir (Özer Özkan, 2019). Amaç öğrenciye not vermek ve derse ait kazanımların ne derecede kavradığına karar vermektir (Güler, 2015).

McMillan (2001) ise değerlendirmenin *amaç, ölçme yöntemi, sonuçların analizi/yorumlanması ve sonuçların kullanımı* olmak üzere dört aşamadan oluştuğunu belirtmiştir. Değerlendirme süreci Şekil 2’de verilmiştir.



Şekil 2. Değerlendirme (McMillan, 2001)

Değerlendirme öncelikle amacın belirlenmesi ile başlayan bir süreçtir ve amaç değerlendiriciye ve kullanılan değerlendirme yöntemine göre farklılaşır. Örneğin, bir öğretmen dersin başında öğrencilerin ne bildiğini anlayabilmek, ders sırasında öğrencilerin öğrenmelerini takip edebilmek ve ders sonunda öğrencilerin konuyu ne kadar öğrenebildiğini anlayabilmek amacıyla değerlendirmeyi kullanabilirken, okul yönetimleri değerlendirmeyi öğrencilerin standart testlerde gösterdikleri performansı ölçmek amacıyla kullanabilir. Aileler çocuklarının ders durumlarını öğrenmek amacıyla değerlendirme sonuçlarına ihtiyaç duyabilirler. Eğitim politikalarına yön

veren yöneticiler ise okulların performanslarının değerlendirilmesi amacıyla öğrencilerin yüksek riskli sınavlardan aldıkları puanları kullanılabilir.

McMillan (2001) değerlendirmede amaca odaklanılırken sınıf içi değerlendirmeler ve geniş ölçekli değerlendirmeler arasındaki farkların iyi bir şekilde bilinmesi gerektiğini ve böylelikle amaca uygun değerlendirme yönteminin seçilebileceğini belirtmiştir. Sınıf içi değerlendirmelerde öğrenci eğitim sürecinde devamlı olarak değerlendirmeye tabi tutulur ve öğrenciye anlık geri bildirim verilir. Öğrencinin konuyu en iyi şekilde özümsemesine ve öğretmenin dersi öğrencinin en iyi anlayacağı şekilde düzenlemesine imkan tanınır. Geniş ölçekli standart testlerle yapılan değerlendirmelerde ise amaç öğretmen ve okul performansının öğrencinin sınav performansına göre değerlendirilmesidir. Öğrenciye anlık geri bildirim verilmez. Standart testler kullanılarak öğrencilerin birbirlerine göre değerlendirilmesi söz konusudur. Sınıf içi ve geniş ölçekli değerlendirmeler arasındaki farklar Tablo 1’de sunulmuştur.

Tablo 1. *Sınıf içi ve geniş ölçekli değerlendirme* (McMillan, 2001, s.7)

Sınıf İçi Değerlendirme	Geniş Ölçekli Değerlendirme
Öğrenciler bireysel olarak değerlendirilir.	Öğrenciler grup olarak değerlendirilir.
Ders öncesinde, esnasında ve sonrasında yapılabilir.	Ders öncesinde ve sonrasında yapılabilir.
Öğretmen yapımı değerlendirmedir.	Sınıf dışındaki uzmanlarca yapılır.
Farklı sınıflar için uyarlanabilir.	Bütün öğrenciler aynı değerlendirmeye tabi tutulur.
Öğrencilere bireysel, anlık geribildirim verilir.	Öğrencilere değerlendirme sonrasında genel bir geribildirim verilir.
Belli bir zaman içinde ya da zaman tutulmaksızın gerçekleştirilir.	Genellikle değerlendirme belirli bir zaman aralığında gerçekleştirilir.
Değerlendirme süreklidir.	Değerlendirme belirli aralıklarla yapılır.
Değerlendirme konunun belirli bir kısmı için yapılır.	Değerlendirme konunun tamamı için yapılır.
Değerlendirme sonuçları öğrencinin performansıyla karşılaştırılır.	Öğrencilerin puanları birbirleriyle mukayese edilir.
Teknik olarak zayıftır.	Teknik olarak güçlüdür.
Öğrenciye not verme ve öğrencinin gelişimi için kullanılır.	Öğretmen ve okulların performansının değerlendirilmesi amacıyla kullanılır.

McMillan (2001) değerlendirme yöntemlerini ise genel olarak nicel ve nitel değerlendirme olarak ikiye ayırmıştır. Nicel değerlendirme öğrenci performansının sayıyla ifade edilmesi iken, nitel değerlendirme de ise öğrenci performansının sayı

kullanmak yerine sözlü olarak betimlenmesi söz konusudur. Nitel ve nicel değerlendirmeleri ise Tablo 2’de gösterildiği gibi geleneksel ve alternatif değerlendirme yöntemleri başlıkları altında özetlemiş, bu yöntemleri kendi içinde yanıt-seçmeli ve açık uçlu yöntemler olarak ikiye ayırmıştır.

Tablo 2. *Değerlendirme türleri* (McMillan, 2001, s.9)

	Geleneksel Yöntemler	Alternatif Yöntemler
Yanıt-Seçmeli	Çoktan Seçmeli Test	Öz Değerlendirme
	Doğru Yanlış Testi	
	Eşleştirme	
	İki Seçenekli	
	Yapılandırılmış Gözlem	
	Yapılandırılmış Görüşme	
Açık Uçlu	Anketler	
	Cümle Tamamlama	Performans Değerlendirme
	Kısa Cevaplı	Gerçek Ortamda Değerlendirme
	Makale/Kompozisyon Yazma	Portfolyo Değerlendirme
	Anekdöt	Sergiler
	Rapor Yazma	Gösteriler
Yapılandırılmamış Görüşme	Öz Değerlendirme	

Amaca uygun ölçme yönteminin kullanımıyla birlikte öğrenci başarısı hakkında bilgi edinilir. Elde edinilen bu bilgiler bağıl veya mutlak ölçütler göz önünde bulundurularak analiz edilir ve yorumlanır. Son aşamada ise sonuçlar sınıf içi değerlendirme ya da geniş ölçekli değerlendirmenin amaçlarına uygun olacak şekilde kullanılır.

McMillan (2001) değerlendirmede temel amacın öğrenci başarısının iyi bir şekilde raporlanması ve böylece eğitim sürecinde istendik yönde değişiklikler yapılması olduğunu vurgulamıştır. Bu nedenle değerlendirme öğelerin neler olduğu, değerlendirmenin amacı, değerlendirme yöntemleri, değerlendirme sonuçların nasıl yorumlanacağı ve kullanacağı bilgisi eğitim paydaşlarınca iyi bir şekilde bilinmelidir. Bu durum ise alanyazında *değerlendirme okuryazarlığı* olarak isimlendirilmiştir.

2.2. Değerlendirme Okuryazarlığı

Okuryazarlık okuma ve yazma faaliyetleri aracılığıyla sembollerin kavranması ve ifade edilmesi sürecidir. Bu süreçte birey çevresindeki nesne ve olguları kavrayarak

anlamlandırır, özümser ve çevresiyle etkileşim kurar (Aşıcı, 2009). Eğitim sistemi içerisinde birçok değişkeni barındıran ve bu değişkenlerin sürekli etkileşim içerisinde olduğu geniş bir sosyal ağdır ve bu sosyal ağda sağlıklı iletişimin sürdürülmesi için eğitim paydaşlarının okuryazarlık seviyeleri yüksek olmalıdır. Eğitimle ilgili alanyazında teknoloji okuryazarlığı, bilgi okuryazarlığı, öğretmen veya öğrenci geribildirim okuryazarlığı gibi farklı okuryazarlık türlerinin tanımlandığı görülmektedir (Sur, 2022; Uzun ve Çelik, 2020). Bu okuryazarlık türlerinden birisi de *değerlendirme okuryazarlığı*dır.

Değerlendirme okuryazarlığı ilk olarak Stiggins (1991) tarafından tanımlanmıştır. Stiggins'e göre değerlendirme okuryazarı olan bir öğretmen; yüksek ve düşük kaliteli değerlendirmeleri ayırt edebilir, öğrenci başarısını değerlendirme sonucuna göre analiz edebilir ve değerlendirme sonucunun öğrenci üzerindeki olası etkilerini yordayabilir. Webb (2002) ise değerlendirme okuryazarlığı kavramını; öğrencilerin ne bildiği, bunun nasıl değerlendirileceği ve hem öğrencilerin başarısını hem de müfredatı geliştirmek için değerlendirme sonuçlarının nasıl analiz edilebileceği bilgisi olarak açıklamaktadır. Popham (2011) değerlendirme okuryazarlığını, eğitimcilerin eğitimle ilgili kararlarını etkileyen temel değerlendirme kavramlarını ve süreçlerini bilmesi olarak açıklamaktadır. Gürsoy (2017) ise değerlendirme okuryazarlığını,

öğretmenin öğrenci başarısını, uyguladığı yöntemlerin geçerliliğini ve öğretim programının uygunluğunu doğru bir şekilde değerlendirebilmesi için gerekli olan ölçme aracını etik kurallar ve ölçme prensipleri içinde seçme/geliştirme, uygulama, puanlama, yönetme, bilgilendirme ve sonuçları yayınlama gibi süreçleri doğru yönetme becerisi (s. 285) olarak tanımlamıştır.

Yapılan bu tanımlamalara göre değerlendirme okuryazarı olan bir öğretmen (Gürsoy, 2017; Popham, 2011; Stiggins, 1995; Web, 2002):

- Her bir değerlendirme yönteminin amacını, yöntemin sahip olduğu avantajları ve dezavantajları, değerlendirme sonuçlarının öğrenci başarısı hakkında ne söylediğini anlayabilir, analiz edebilir.
- Doğru ve hatalı değerlendirmeler arasındaki farkı bilir.
- Yapılan değerlendirmede nelerin ölçülüp ölçülmediğini anlayabilir ve değerlendirme sonuçlarını analiz edip yorumlayabilir.

- Değerlendirme sonuçlarını öğretimin geliştirilmesi ve öğretim programının etkinliğinin artırılması amacıyla kullanabilir.

Popham (2004) değerlendirme okuryazarlığı ve eğitimde hesap verilebilirlik olgularının ayrılmaz ikili olduklarını savunmuş, öğrencilerin testlerde gösterdikleri performansın doğru bir şekilde değerlendirilmesi ihtiyacını vurgulamıştır. Bu nedenle öğretmenlerin değerlendirme okuryazarı olmaları gerektiğini belirtmiş, değerlendirme okuryazarı olmama durumunu ise profesyonel intihar olarak tanımlamıştır. Öğretmenlerin bu konudaki yetersizliklerinin olası sebeplerini ise hem üniversitede hem de mezun olduktan sonra ölçme değerlendirme konusunda yeterince eğitim almamaları ve ölçme değerlendirmeyi karmaşık, nicel bir alan olarak algılamalarından kaynaklanabileceğini belirtmiştir. Alanyazınında gerçekleştirilen farklı araştırmalar (Acar Güvenir ve Özer Özkan, 2016; Azrak ve Yalçınkaya, 2019; Deluca & Klinger, 2010; Erdoğan ve Kurt, 2012; Ergül, 2019; Koh, 2011; Plake, Impara & Fager, 1993) Popham'ın bu iddiasını doğrular niteliktedir. Yapılan bu araştırmalarda elde edilen sonuçlara göre; ölçme değerlendirme ile ilgili eğitim alan kişilerin eğitim almayanlara göre değerlendirme okuryazarlık seviyelerinin daha yüksek olduğu, öğretmenlerin değerlendirmeyi nicel bir alan olarak gördükleri ve bu nedenle değerlendirme konusunda olumsuz tutuma sahip oldukları görülmüştür.

Stiggins (1995) ise öğretmenlerin değerlendirme okuryazarlığına sahip olmasını engelleyen etmenleri; değerlendirme konusunda duyulan korku, doğru değerlendirme için yeterli zamanın ayrılamaması ve değerlendirmeye ilişkin toplumsal algılar olarak sıralamıştır. Öğretmenlerin değerlendirmeye ilişkin korkularının kendi öğrencilik dönemlerinde maruz kaldıkları yanlış değerlendirme yöntemlerinden kaynaklanabileceğini ileri sürmüştür. Sürekli olarak başarıya göre sıralanma duygusunun yaşatıldığı bu dönemlerin öğretmenlerin değerlendirme konusunda olumsuz tutum geliştirmelerine sebep olduğunu iddia etmiştir. Aynı zamanda öğretmenlerin değerlendirme konusunda ilk defa eğitim aldıkları üniversite dönemindeki derslerin ise onları gerçek sınıf ortamındaki değerlendirmelere yeterince hazırlayamadığını belirtmiştir. Bu ölçme değerlendirme derslerinin teorik ağırlıklı olduğunu, gerçek sınıf ortamıyla bağlantı kurulamadığını ileri sürmüştür. Quilter ve Gallini (2000) ise geçmişte değerlendirme konusunda olumsuz tutuma sahip

öğretmenlerin farklı değerlendirme türlerine karşı korku duyduğunu ve bu yöntemleri derslerinde kullanmamayı tercih ettiklerini belirtilmektedir.

Değerlendirme okuryazarlığının önündeki engellerden bir diğeri ise doğru değerlendirme için yeterli zamanın ayrılamamasıdır (Stiggins, 1995). Öğretmenlerin üzerindeki sorumlulukların fazlalığı, müfredatta yaşanan değişimler, öğrenci başarısının kısıtlı zaman içerisinde değerlendirilip raporlandırılması isteği gibi faktörler öğretmenlerin çok zaman alan değerlendirme yöntemlerini tercih etmemesini ve bazı değerlendirme yöntemlerini ise daha çok kullanmasına neden olmaktadır. Bu durum doğru değerlendirmenin yapılamamasına ve öğretmenlerin değerlendirme okuryazarlığının gelişmemesine neden olmaktadır.

Toplumsal kalıplaşmış yargılar da değerlendirme okuryazarlığının gelişiminin önündeki engellerden birisidir (Stiggins, 1995). Öğrenci başarısının standartlaştırılmış testlerden alınan sonuçlara göre değerlendirilmeye çalışılması, yüksek riskli sınav sonuçlarına göre öğretmenlerin ve okulların performansının değerlendirilmesi, sınıf içi değerlendirmeye yeterince önem verilmemesi ve karne notlarının öğrencinin başarısını yüzde yüz yansıttığı yanılgısı bu kalıplaşmış yargılara örnektir. Bu tür toplumsal inanışlar öğretmenlerin farklı değerlendirme yöntemlerini kullanmalarına ket vurmakta ve neticede öğretmenlerin değerlendirme okuryazarlıkları gelişmemektedir.

2.2.1. Değerlendirme Okuryazarı Olmanın Önemi

Değerlendirme okuryazarlığı kavramının 1991 yılında Stiggins tarafından tanımlanmasından sonra, kavramın gelişimindeki en önemli dönüm noktalarından birisi Amerika Birleşik Devletleri'nde hayata geçirilen Hiçbir Çocuğun Geride Bırakılmaması Yasası'nın (No Child Left Behind Act, 2002) kabulüdür. Bu yasayla eyaletler ve okullar arasındaki öğrenci başarı farklarını gidermek amaçlanmıştır. Bu amaç doğrultusunda öğrenci başarısı standart testler kullanılarak yıllık olarak değerlendirilmiştir (Güçlü ve Bayrakçı, 2004). Bu reform hareketinde öğrenci başarısının doğru bir şekilde değerlendirilmesi için uygun yöntemlerin kullanılması ve velilerin öğrenci başarısı hakkında bilgilendirilmesi hedeflenmiştir (Coble & Azordegan, 2004). Bu yasanın okullara ve öğretmenlere yüklediği öğrenci başarısının yüksek standartlara çıkarılması sorumluluğu, değerlendirme okuryazarlığı kavramının

önemini artırmıştır. Öğretmenler öğrenci başarısını uygun yöntemler kullanarak doğru şekilde değerlendirmekten sorumlu tutulmuşlardır. Dolayısıyla değerlendirme okuryazarlığı kavramına olan ilgi artmıştır.

Webb (2002) araştırmasında değerlendirme okuryazarlığına artan ilginin sebeplerini, standart testlerle yapılan değerlendirmelerde öğrenci başarısının doğru bir şekilde ölçülmesinin istenmesi ve farklı değerlendirme türlerinin kabul edilmesi olarak sıralamıştır. Okulların ve öğretmenlerin başarısı öğrencilerin yüksek riskli sınavlarda gösterdikleri performansa göre değerlendirilmektedir (Koh, 2011). Yüksek riskli sınavların uygulanması ve bu sınavlardan alınan sonuçlarla öğrenci başarısının değerlendirilmeye çalışılması, öğretmenler üzerindeki baskıyı artırmış; eğitimde hesapverilebilirlik algısının artmasıyla birlikte öğretmenlerin farklı değerlendirme yöntem ve becerilerine sahip olması gerektiği ve böylece hem eğitim sürecinin geliştirilebileceği hem de öğrenci başarısının daha doğru bir şekilde ölçülebileceği inancı kabul görmüştür (Deluca & Klinger, 2010).

Değerlendirme yaklaşımındaki değişimler bunlarla sınırlı kalmamış, sonuç odaklı değerlendirmeden öğrencilerin sınıf içinde yapılan aktivitelerle değerlendirilmesi ve öğretmenlere değerlendirme konusunda destek verilmesi anlayışına geçiş süreci yaşanmıştır (Deluca vd., 2016a). 1990'lardan bugüne dek değerlendirme yaklaşımında yaşanan bu değişimler, değerlendirme okuryazarı olan öğretmenlerin sahip olması gereken temel standartları da şekillendirmiştir. Örneğin, Stiggins (1995) 21. yüzyılda değerlendirme okuryazarı olan bir öğretmenin sınıf içi değerlendirmenin beş temel standardını bilmesi gerektiğini öne sürmüştür. Bu standartlar;

1. Değerlendirmeye açık, net amaçlarla başlama
2. Başarı hedeflerine (*konu içeriğine hakimiyet, sorgulama yeterliliklerinin geliştirilmesi, performans becerilerinin artırılması, yüksek kaliteli ürünler elde etme*) odaklanma
3. Farklı başarı hedefleri için uygun değerlendirme yöntemlerinin seçimi
4. Kullanılan değerlendirme yöntem ve araçlarına göre öğrenci başarısının doğru bir şekilde örneklemlendirilmesi
5. Değerlendirmede yanlılık ve hatalardan kaçınmadır.

Bu standartların öğretmenlerce anlaşılması ve içselleştirilmesi öğrencilere iyi bir eğitim verilmesinin temel anahtarıdır. Bu standartlara göre değerlendirme okuryazarı olan bir öğretmen hem öğretim sürecini hem de öğrencilerin öğrenmesini geliştirebilir ve güçlendirebilir.

Popham 2018 yılında yayımladığı kitabında değerlendirme okuryazarı olan ve olmayan öğretmenlerin öğrencilerin eğitimlerine olan etkilerini incelemiştir. Değerlendirme okuryazarlığının hem öğretmenleri daha iyi eğitimci yapacağını hem de öğrencilerin daha iyi eğitim görmelerini sağlayacağını belirtmiştir. Popham'a göre değerlendirme okuryazarı olan bir öğretmen, eğitsel kararlar alırken doğru testleri doğru amaçlar çerçevesinde kullanır. Değerlendirme yönteminin avantaj ve dezavantajlarını göz önünde bulundurarak, eğitim süreci boyunca öğrencilerin ihtiyaçlarına yönelik faydalı testler kullanır. Öğrencilerin iyi bir eğitim alması, öğretmenlerin öğretim süreci boyunca doğru eğitsel kararlar vermelerine bağlıdır. Bu durum ise öğretmenlerin değerlendirme okuryazarı olmalarının doğal bir sonucudur.

2.2.2. Ölçme Değerlendirme Standartları

Öğretmenlerin öğrenci başarısını doğru bir şekilde değerlendirilebilmesi için belirli yeterliliklere sahip olması gerekir. Öğretmenlerin değerlendirme konusunda hangi yeterliliklere sahip olması gerektiği ve bu yeterliliklerin belirli bir çerçevede nasıl tanımlanabileceği üzerine yapılan çalışmalar 1990'lı yıllara dek geriye gitmektedir. Değerlendirme okuryazarlığı kavramının belirli standartlar altında tanımlanması için yapılan ilk çalışma ABD'de Öğretmenler Federasyonu (American Federation of Teachers, AFT), Ulusal Eğitimde Ölçme Konseyi (National Council on Measurement in Education, NCME) ve Ulusal Eğitim Kurumu'nun (National Education Association, NEA) iş birliğinde 1987-1990 yılları arasında gerçekleştirilmiştir. Bu çalışma sonucunda ölçme değerlendirme standartları yedi madde ile tanımlanmış (Tablo 3) ve öğretmenlerin değerlendirme konusunda sahip olması gereken yeterlilikler belirlenmiştir.

Türkiye'de ise öğretmen yeterlilikleri hakkındaki ilk çalışma 1999 yılında yapılmıştır (MEB, 2017). Bu çalışmanın sonucunda Öğretmen Yeterlilikleri Komisyonu tarafından Öğretmen Mesleği Genel Yeterlilikleri belgesi 2002 yılında yayımlanmıştır. Bu belgeye göre öğretmenlerin değerlendirme konusunda sahip

olması gereken yeterlilikler *Eğitme ve Öğretme Yeterlilikleri* başlığı içerisinde sunulmuştur (Seferoğlu, 2004). Türkiye’de öğretmen yeterlilikleri üzerindeki ikinci önemli çalışma ise Temel Eğitime Destek Programı (TEDP) kapsamında yapılmıştır. Bu çalışma sonucunda Öğretmen Mesleği Genel Yeterlilikleri taslağı oluşturulmuş ve yayımlanmıştır (MEB, 2006). Bu çalışmada öğretmenlerin değerlendirme konusundaki yeterlilikleri *Öğrenmeyi, Gelişimi İzleme ve Değerlendirme* başlığı içerisinde verilmiştir. Avrupa Yeterlilikler Çerçevesi’nin 2008 yılında kabulünden sonra ülkemizde bu çerçeveye uyum sağlamak amacıyla öğretmen yeterliliklerinin güncellemesi ihtiyacı görülmüştür. Türkiye Yeterlilikler Çerçevesi’nin 2015 yılında yürürlüğe girmesiyle birlikte Öğretmen Mesleği Genel Yeterlilikleri güncellenmiş ve yayımlanmıştır (MEB, 2017). Öğretmen mesleği genel yeterlilikleri *Mesleki Bilgi, Mesleki Beceri ve Tutum ve Değerler* başlıkları altında yapılandırılmış ve öğretmenlerin değerlendirme konusunda sahip olması gereken yeterlilikler beş madde ile tanımlanmıştır. İlgili maddeler Tablo 3’te verilmiştir.

Tablo 3. Ölçme değerlendirme standartları

Ölçme Değerlendirme Standartları (AFT, NCME & NEA, 1990)	Mesleki Beceriler (MEB, 2017, s. 22) Ölçme ve Değerlendirme Öğretmen Yeterlilikleri
1. Eğitimle ilgili karar alınırken uygun değerlendirme yöntemlerini seçmede yetenekli olma	1. Alanına ve öğrencilerin gelişim özelliklerine uygun ölçme ve değerlendirme araçları hazırlar ve kullanır
2. Eğitimle ilgili karar alınırken uygun değerlendirme yöntemlerini geliştirmede yetenekli olma	2. Ölçme ve değerlendirmede süreç ve sonuç odaklı yöntemler kullanır.
3. Öğretmen yapımı bir testin ya da hazır bir testin uygulanmasında, puanlanmasında ve sonuçlarının yorumlanmasında yetenekli olma	3. Ölçme ve değerlendirmeyi objektif ve adil olarak yapar.
4. Değerlendirme sonuçlarını; öğrenciler hakkında bireysel karar verme, öğretimi planlama, öğretim programını ve okulu geliştirme konularında kullanabilme	4. Ölçme ve değerlendirme sonuçlarına göre öğrenci ve diğer paydaşlara doğru ve yapıcı geribildirimler verir.
5. Öğrencilerin değerlendirmesinde geçerliği kanıtlanmış not verme yöntemlerini geliştirebilmede yetenekli olma	5. Ölçme ve değerlendirme sonuçlarına göre öğretme ve öğrenme süreçlerini yeniden düzenler.
6. Değerlendirme sonuçlarını öğrencilerle, velilerle, eğitimcilerle ve eğitimle alakalı diğer kişilerle paylaşabilme	
7. Etik olmayan, yasa dışı ve farklı nedenlerle uygun olmayan değerlendirme yöntemlerini fark etmede ve bu değerlendirme bilgilerinin kullanımını tanımadada yetenekli olma	

Ölçme değerlendirme standartlarının 1990 yılında yayımlanmasıyla birlikte, değerlendirme konusunda öğretmen yeterliliklerin belirlenmesi için yapılan çalışmalar hız kazanmıştır. 1990 yılından bugüne dek değerlendirme konusundaki temel algıların değişimi değerlendirme standartlarını yeniden şekillendirmiştir. Sonuç odaklı değerlendirme anlayışından biçimlendirici değerlendirme ve öğretmenlerin değerlendirme konusunda eğitilmesi, desteklenmesi anlayışına geçiş süreci yaşanmıştır. Son yirmi üç yıllık zaman diliminde bu değişim ve gelişmelere bağlı olarak bölgesel ya da ulusal çapta değerlendirme standartları oluşturulmuştur (Deluca vd., 2016a). Örneğin Brookhart (2011) çalışmasında 1990 yılında yayımlanan ölçme değerlendirme standartlarını biçimlendirici değerlendirmede yaşanan güncel değişimleri yansıtmaması ve öğretmenlerin standart tabanlı test kullanımında karşılaştıkları sorunlara yeterince değinmemesi nedeniyle modası geçmiş olarak nitelendirmiştir. Bu nedenle Brookhart (2011) eğitimde hesapverilebilirlik algısındaki değişimleri göz önünde bulundurarak on bir maddelik ölçme değerlendirme standartlarını geliştirmiştir. Bu standartlar sırasıyla verilmiştir.

1. Öğretmenler kendi alanlarına ilişkin detaylı bilgiye sahip olmalı ve öğrenci öğrenmelerinin bu alanda nasıl gerçekleştiğini anlayabilmelidir.
2. Öğretmenler öğretim ve değerlendirme sürecinin temellerini oluşturan, konu ve müfredatla uyumlu olan hedefler belirleyebilmelidir.
3. Öğretmenler belirledikleri hedeflere ulaşmada kullanacakları yöntemleri öğrencilere anlatmada geniş bir strateji ağına (söyleme, gösterme, keşfettirme vb.) sahip olmalıdır.
4. Öğretmenler kullanabilecekleri değerlendirme yöntemlerinin amaçlarını ve bu yöntemlerin nasıl kullanılması gerektiğini bilmeli, bu konuda yetenekli olmadır.
5. Öğretmenler, öğrencilerin düşünce becerilerini ve belirli değerlendirme görevlerini yapabilmek için sahip olmaları gerekenleri anlayabilmek için; test maddelerini, sınıf içinde sorulan soruları ve performans görevlerini analiz edebilmelidir.
6. Öğretmenler öğrencilere etkili ve kullanışlı geri bildirimler verebilmelidir.
7. Öğretmenler öğrencinin sınıf içi değerlendirmelerinden oluşan puanlama sistemlerini oluşturabilmelidir. Bu sistemlerle öğrenci, okul, sınıf ve bölge

hakkında karar verilme sürecine destek olmalı, böylelikle öğrencinin öğrenmesini ve gelişimini desteklemelidir.

8. Öğretmenler öğrencilere çeşitli değerlendirmeler uygulayabilmeli, değerlendirme sonuçlarını öğrenciler, sınıflar, okullar ve bölgeler için karar alma süreçleri için yorumlayabilmelidir.
9. Öğretmenler değerlendirme sonuçlarının toplu analiz edilebilmeli, sonuçları eğitsel kararlar alınırken yorumlayıp kullanabilmelidir.
10. Öğretmenler değerlendirme sonuçlarının öğrenciler tarafından doğru şekilde anlaşılıp değerlendirilmesi için öğrencilere rehberlik etmelidir.
11. Öğretmenler değerlendirme sürecinde etik ve yasal unsurlara dikkat etmelidir.

Benzer şekilde Eğitimsel Değerlendirme Standartları Ortak Komitesi tarafından eğitim paydaşlarının değerlendirme konusunda desteklenmesi amacıyla Sınıf İçi Değerlendirme Standartları (Klinger vd., 2015) oluşturulmuştur. Bu standartlara ait Temeller, Kullanım ve Kalite bölümleri ise Tablo 4’te verilmiştir.

Tablo 4. *Sınıf içi değerlendirme standartları* (Klinger vd., 2015)

Temeller	Kullanım	Kalite
Değerlendirme Amacı	Öğrenci Performansının Analizi	Kültürel ve Dilsel Çeşitlilik
Öğrenme Beklentileri	Etkili Geribildirim	Olağanüstülük ve Özel Eğitim Tarafsız ve Adil Değerlendirme
Değerlendirme Tasarımı	Öğretim Takibi	Güvenilirlik ve Geçerlilik
Değerlendirmede Öğrenci Katılımı	Notlar ve Özet Yorumlar	Yansıtma
Değerlendirme Hazırlığı	Raporlama	
Bilgilendirilmiş Öğrenciler ve Veliler/Vasiler		

Tablo 4 incelendiğinde, eğitsel değerlendirme standartlarının *Temeller*, *Kullanım* ve *Kalite* olmak üzere üç başlıkta yapılandırıldığı görülmektedir. *Temeller*, değerlendirmenin niçin ve nasıl yapılacağı, öğrenci ve velilerin değerlendirme hakkında bilgilendirilmesi, öğrenme beklentileri ve değerlendirme sürecine öğrencinin de dahil edilmesiyle ilgilidir. *Kullanım* ise değerlendirme yapılması, sonuçların analizi/raporlanması ve sonuçların iletimi konularını kapsamaktadır. *Eğitsel standartların kalitesi* ise değerlendirmenin yapıldığı ortamın özelliklerini ya

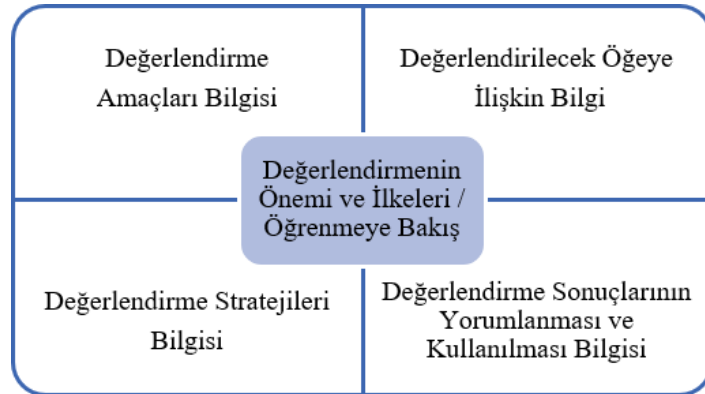
da değerlendirilen kişilere ait özel durumların göz önünde bulundurulması ve böylece değerlendirmede güvenilirlik ve geçerliğin sağlanmasıyla ilgilidir.

2.2.3. Değerlendirme Okuryazarlığı Modelleri

Öğrenci başarısının doğru bir şekilde değerlendirilmesi ihtiyacı ve eğitimde yüksek riskli sınavların uygulanmasıyla birlikte hesapverilebilirlik olgusunun önem kazanması, biçimlendirici değerlendirme anlayışının ön plana çıkması (Brookhart, 2011) ve bunlara ek olarak değerlendirme konusunda öğretmenlerin desteklenmesi fikri (Deluca vd., 2016a), değerlendirme okuryazarlığı kavramının yeniden ele alınmasına ve kavramla ilgili farklı modeller geliştirilmesine neden olmuştur (Abell & Siegel, 2011; Chan & Luk, 2022; Pastore & Andrade, 2019; Xu & Brown, 2016). Bu modeller geliştirilme tarihine göre kronolojik olarak aşağıda verilmiştir.

Birinci Model

Abell ve Siegel (2011) araştırmalarında fen bilimleri öğretmenlerinin değerlendirme okuryazarlıklarını modellemiştir. Modelin merkezinde öğretmenlerin öğrenmeye ilişkin görüşleri, değerlendirmenin önemi ve ilkeleri yer almaktadır. Modelin merkezini ise birbirleriyle ilişkili olan *Değerlendirme Amaçları Bilgisi*, *Değerlendirilecek Öğeye İlişkin Bilgi*, *Değerlendirme Stratejileri Bilgisi* ve *Değerlendirme Sonuçlarının Yorumlanması ve Kullanılması Bilgisi* çevrelemektedir. Abell ve Siegel (2011) tarafından oluşturulan model Şekil 3'te sunulmuştur.



Şekil 3. Fen bilimleri öğretmenleri için değerlendirme okuryazarlığı modeli (Abell & Siegel, 2011)

Şekil 3'teki modele göre öğretmenin değerlendirmeye ve öğrenci öğrenmelerine bakış açısı, öğretmenin değerlendirme konusunda verdiği kararları etkilemektedir. Öğretmenin neyi, niçin ve nasıl değerlendireceğini, hangi değerlendirme stratejisini kullanacağını ve değerlendirme sonuçlarını nasıl kullanacağını belirlemektedir.

İkinci Model

Xu ve Brown (2016) tarafından geliştirilmiştir. Bu modelde araştırmacılar öğretmen değerlendirme okuryazarlığının uygulamada nasıl modellenebileceği üzerine çalışma gerçekleştirmiştir. Çalışmada 1985 ile 2015 yılları arasında değerlendirme okuryazarlığı ile ilgili yayımlanmış araştırmalar incelenmiş ve değerlendirme okuryazarlığının gelişimini sağlayan ya da engelleyen etmenler belirlenmeye çalışılmıştır. Geliştirilen değerlendirme okuryazarlığı modeli aracılığıyla öğretmen eğitimi ile değerlendirme okuryazarlığı arasındaki bağın kuvvetlendirilmesi hedeflenmiştir. Piramitsel olarak aşamalı bir sınıflandırmayla kurgulanan modelde temel amaç, piramidin en tepe noktasında yer alan öğretmenlerin birer değerlendirici rolüne kavuşturulması ya da sahip oldukları bu rollerin yeniden yapılandırılmasıdır. Uygulamada değerlendirme okuryazarlığı modeli Şekil 4'te verilmiştir.



Şekil 4. Uygulamada değerlendirme okuryazarlığı modeli (Xu & Brown, 2016)

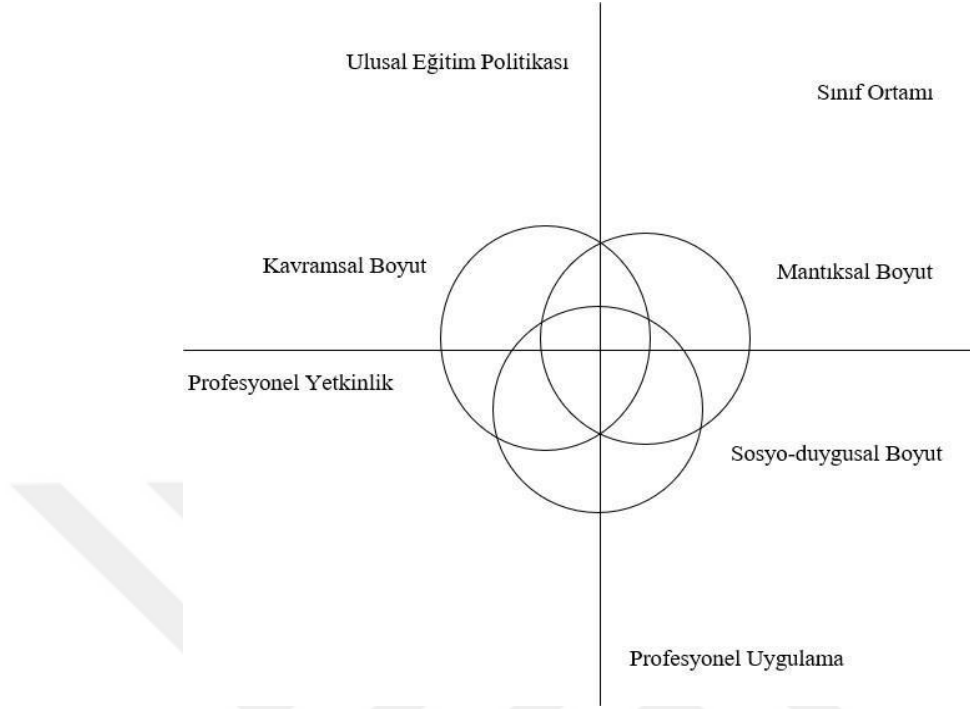
Bu modelde piramidin ilk basamağı değerlendirme bilgisi basamağıdır. Bu basamakta öğretmenin pedagojik alan bilgisi, değerlendirme amaçları, içeriği ve yöntemleri bilgisi, notlandırma bilgisi, geribildirim bilgisi, değerlendirmenin

yorumlanması ve paylaşımı bilgisi, değerlendirme etiği bilgisi ve öğrencinin değerlendirme süreçlerine katılımı bilgisi yer almaktadır. Piramidin ikinci basamağında ise öğretmenlerin değerlendirmeye ilişkin görüşleri bulunmaktadır. Bilişsel ve duyuşsal boyutta öğretmenin değerlendirme konusunda sahip olduğu görüşler bilgi basamağının yapılandırılmasını sağlamaktadır. Öğretmenin içerisinde bulunduğu sosyo-kültürel ortam ve çalıştığı kurumun dinamikleri değerlendirme okuryazarlığının gelişiminde önemli bir role sahiptir. Eğitim politikaları, hesapverilebilirlik olgusu, sınıf mevcudu gibi etmenler öğretmenlerin değerlendirmede kullandıkları yöntem, içerik ve stratejileri şekillendirmektedir. Bu durum geliştirilen değerlendirme modelin standart bilgi ve becerilerden ziyade dinamik, şartlara göre uyarlanabilen bir yapıda olmasını sağlamaktadır.

Piramidin bir üst basamağında ise değerlendirme yöntemlerini öğretmen öğrenmesine katkıları ve öğretmenlerin kendileri geliştirmesi yer almaktadır. Sınıf içi değerlendirme yöntemlerinin uygulanmasıyla öğretmenler değerlendirmede eksik oldukları alanları görebilmekte, değerlendirici kimliği kazanmadan önce eksiklerini giderebilmektedir. Piramidin önemli bir diğeri ögesi ise katılımıdır. Öğretmenlerin değerlendirme konusunda kendilerini geliştirebilecekleri atölyelere, çalışmalara ve sempozyumlara katılmaları değerlendirici kimliği kazanmada önemli bir aşamadır. Değerlendirici kimliği kazanmada modelin tüm basamaklarının kavranması ve öğretmenin değerlendirme konusunda farkındalığının artması şarttır. Fakat modelin basamakları standardize edilmiş değil, dinamik içinde bulunan şartlara göre değişen bir yapıdadır.

Üçüncü Model

Pastore ve Andrade (2019) ise araştırmalarında değerlendirme okuryazarlığını üç boyutlu bir modelle açıklamışlardır. Delphi tekniğini kullanarak farklı ülkelerden uzmanların katılımıyla değerlendirme okuryazarlığını kavramsal, mantıksal ve sosyo-duygusal olmak üzere üç boyutta modellemiştir. Geliştirilen model Şekil 5'te verilmiştir.



Şekil 5. Üç boyutlu değerlendirme okuryazarlığı modeli (Pastore & Andrade, 2019)

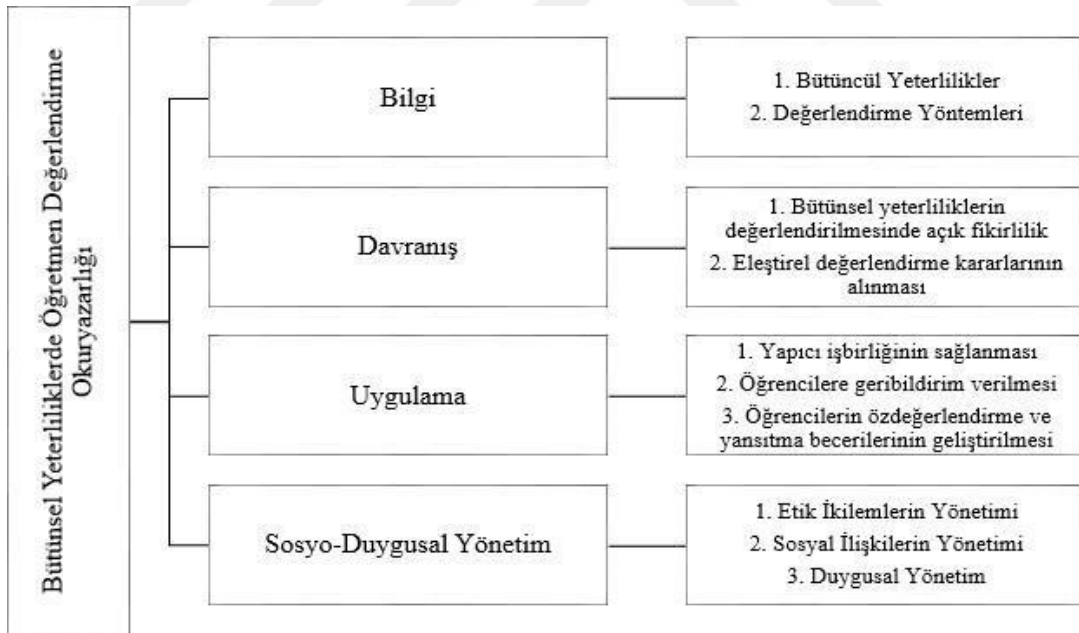
Şekil 5'teki modelde kavramsal boyutta değerlendirmenin ne olduğu, neden yapıldığı, hangi yöntemlerin kullanıldığı, veri analizi, sonuçların raporlanması ve paylaşımı başlıkları yer almaktadır. Mantıksal boyut ise değerlendirmenin uygulanma aşamasıdır. Değerlendirme hedeflerinin/ölçütlerinin belirlenmesi, uygun değerlendirme yöntemlerinin seçimi, sonuçların analizi, öğrencilere geribildirim verilmesi vb. uygulamaya dönüt başlıklar bu boyutta yer almaktadır. Sosyo-duygusal boyutta ise değerlendirmenin sosyal bir uygulama olduğu gerçeği vurgulanmaktadır. Öğretmenlerin eğitim paydaşlarıyla değerlendirme konusunda etkin bir şekilde çalışmaları, öğrencilere ait bilgilerin değerlendirme sürecinde korunması, etik durumlara dikkat edilmesi, değerlendirmenin öğretmen-öğrenci ilişkisindeki ve öğrencinin derse katılımındaki etkisi bu boyutta yer alan temel hususlardır. Bu modelle değerlendirme okuryazarlığını etkileyen boyutlar ve onların iç içe geçen dinamik yapısı vurgulanmış ve her bir boyut için araştırmalar yapılması önerilmiştir.

Dördüncü Model

Değerlendirme okuryazarlığının modellenmesi üzerine gerçekleştirilen bir diğer araştırma ise Chan ve Luk (2022) tarafından bütüncül yeterliliklerin

değerlendirilmesi konusunda yapılmıştır. Chan ve Luk (2022) bütüncül yeterlilikleri, bireylerin kişisel gelişimlerini ve toplumda yer edinmelerini sağlayan genel yeterlilikler (problem çözme becerisi, grup çalışması becerisi vb.) ve olumlu değer ya da davranışlar (kendine güven, psikolojik sağlamlık vb.) olarak tanımlamışlardır. Alanyazında değerlendirme okuryazarlığı ile ilgili geliştirilen modellerin (Abell & Siegel, 2011; Pastore & Andrade, 2019; Xu & Brown, 2016) akademik bilginin değerlendirilmesiyle ilgili olduğunu, bütüncül yeterliliklerin değerlendirilmesi ile ilgili bir model geliştirilmediğini belirtmişlerdir.

Chan ve Luk (2022) araştırmalarını beş farklı üniversiteden 24 akademisyenle yarı-yapılandırılmış görüşmeler ve odak grup görüşmesi yaparak gerçekleştirmiştir. İçerik analizi yaparak verileri, daha önceki geliştirilen değerlendirme okuryazarlığı modellerini inceleyerek elde ettikleri bilgi, davranış, uygulama ve sosyo-duygusal yönetim boyutlarında sınıflandırmışlardır. Şekil 6'da verilen bütüncül yeterliliklerde değerlendirme okuryazarlığı modeli dört boyut ve 10 alt boyuttan oluşmaktadır.



Şekil 6. Bütüncül yeterliliklerde öğretmen değerlendirme okuryazarlığı modeli (Chan & Luk, 2022)

Bilgi boyutunda öğretmenlerin bütüncül yeterlilikler ve bu yeterliliklerin değerlendirilmesinde farklı değerlendirme yöntemleri hakkında yeterli bilgiye sahip olması gerektiği vurgulanmaktadır. Davranış boyutu ise öğretmenlerin bütüncül

yeterliliklerinin geliştirilmesi ve değerlendirilmesinde sahip oldukları düşüncelerle ilgilidir. Öğretmenlerin açık fikirli ve eleştirel bir bakış açısıyla değerlendirme yapmaları gerektiği sonucuna varılmıştır. Uygulama boyutu ise öğretmen ve öğrenci bakış açısından değerlendirmelerin yapılandırılması, öğrencilere geri bildirim verilmesi ve öğrencilerin özdeğerlendirme ve özyansıtma becerilerinin geliştirilmesiyle ilgilidir. Sosyo-duygusal yönetim boyutunda ise öğretmenlerin bütüncül yeterliliklerin değerlendirilmesinde adilliğin sağlanması, olumlu öğretmen öğrenci iletişiminin kurulması ve öğrencilerin duygu durumlarının değerlendirme yaparken ve onlara geri bildirim verirken dikkate alınmasıyla ilgilidir.

Değerlendirme okuryazarlığı ile ilgili geliştirilen modeller incelendiğinde değerlendirme bilgisinin modellerin temelini oluşturduğu görülmektedir. Bilgi basamağından sonra ise değerlendirmede sosyo-kültürel ve sosyo-duygusal öğelerin etkilerinin incelendiği anlaşılmaktadır. Modellerin farklı disiplinler için değişebileceği ya da güncellenebileceği, dinamik bir yapıda oldukları ve modellerde tanımlanan boyutların kesin sınırlarla birbirlerinden ayrılmadıkları görülmektedir.

2.2.4. Değerlendirme Okuryazarlığı ile İlgili Geliştirilen Ölçme Araçları

Değerlendirme okuryazarlığının tanımlanması ve standartların belirlenmesi çalışmalarından sonraki en önemli aşama, öğretmenlerin bu konudaki yeterliliklerinin belirlenmesidir. 1990 yılındaki ölçme değerlendirme standartlarının yayımlanmasıyla birlikte öğretmenlerin değerlendirme okuryazarlığı seviyelerinin ölçülmesi amacıyla çeşitli ölçme araçları geliştirilmiştir. DeLuca vd. (2016a) araştırmalarında ölçme değerlendirme standartlarında 1990 yılından bugüne dek meydana değişimleri ve değerlendirme okuryazarlığı ile ilgili geliştirilen ölçme araçlarını analiz etmişlerdir. Geliştirilen ölçme araçlarının türlerini, içerdikleri maddelerin yapılarını, hangi standartlara göre geliştirildiklerini, hangi örneklerde uygulandıklarını ve uygulama sonucunda elde edilen psikometrik özellikleri incelemişlerdir. Benzer bir çalışma ise Türkiye’de Gürsoy (2017) tarafından gerçekleştirilmiştir. Gürsoy (2017) araştırmasında değerlendirme okuryazarlığı ile ilgili geliştirilen ölçme araçlarını kronolojik olarak incelemiştir.

Değerlendirme okuryazarlığı ile ilgili ilk ölçme aracı Plake, Impara ve Fager (1993) tarafından geliştirilen Öğretmen Değerlendirme Okuryazarlığı Ölçeği (TALQ)’dir. Bu ölçme aracı, 1990 yılındaki belirlenen her bir standart için beşer soru

olmak üzere toplamda 35 sorudan oluşan bir başarı testidir. Testin 555 öğretmene uygulanmasıyla birlikte testin güvenilirlik değeri ,54 ve ortalama puanı ise 23,2 olarak bulunmuştur. TALQ kendisinden sonra değerlendirme okuryazarlığı ile ilgili geliştirilen birçok ölçme aracının temellerini oluşturmuştur. Örneğin Campbell vd. (2002) TALQ'yu 220 öğretmen adayına uygulamış, ölçme aracının güvenilirliğini ,74 olarak bulmuşlardır. Bu uygulamada ölçme aracını Değerlendirme Okuryazarlığı Envanteri (ALI) olarak tekrardan isimlendirmişlerdir.

Mertler ise 2003 yılında TALQ'yu 197 öğretmen ve 220 öğretmen adayına uygulamıştır. Bu uygulamada ölçme aracının öğretmenler için hesaplanan güvenilirlik değeri ,57 olarak bulunurken öğretmen adayları için ,74 olarak bulunmuştur. Mertler ve Campbell 2005 yılında Plake vd. (1993) geliştirdiği TALQ'yu senaryo tabanlı sorulardan oluşacak şekilde yeniden yapılandırmıştır. Geliştirdikleri ölçme aracını 250 öğretmen adayına uygulamışlar ve ölçme aracının güvenilirlik değerini ,74 olarak hesaplamışlardır.

DeLuca vd. (2016a) 1990 ölçme değerlendirme standartlarının, değerlendirme okuryazarlığı ile geliştirilen ölçme araçlarının çoğunluğunun temellerini oluşturduğunu belirtmiştir. Farklı değerlendirme standartların çerçevesini çizdiği ölçme araçlarının ise kısıtlı sayıda olduğu görülmektedir (DeLuca vd., 2016a; Gürsoy 2017).

Sonuç olarak değerlendirme okuryazarlığı standartlarındaki gelişmeler ve ölçme araçlarının farklı örneklerde uygulanıp sonuçlarının doğrulanma çabası, değerlendirme okuryazarlığı ile ilgili birçok ölçme aracının geliştirilmesine neden olmuştur. Geliştirilen ölçme araçları Tablo 5'te verilmiştir. Plake vd. (1993) öncüsü olduğu bu çalışmalarda öğretmen ve öğretmen adaylarının değerlendirme konusundaki yeterlilikleri ölçme değerlendirme standartları göz önünde bulundurularak belirlenmeye çalışılmıştır.

Tablo 5. *Değerlendirme okuryazarlığı ile ilgili geliştirilen ölçme araçları*

Araştırmacılar	Ölçme Aracının Adı	Ölçme Aracı Türü / Madde Özellikleri	Kılavuz Çerçeve	Cevaplayıcılar	Psikometrik Özellikler
Plake vd. (1993)	TALQ	Başarı Testi 35 madde, her bir standart için beş madde.	Ölçme Değerlendirme Standartları (1990)	555 Öğretmen	r=,54 \bar{x} =23,2 S=3,3

Tablo 5 (Devamı)

Campbell, Murphy & Holt (2002)	ALI	Başarı Testi Toplam 35 madde, her bir standart için beş madde.	Ölçme Değerlendirme Standartları (1990)	220 Öğretmen Adayı	$r=,74$ $\bar{x}=21$ $S=4,6$
Mertler (2003)	CALI	Başarı Testi Toplam 35 madde, her bir standart için beş madde.	Ölçme Değerlendirme Standartları (1990)	197 Öğretmen 220 Öğretmen Adayı	$r=,57$ $\bar{x}=22$ $S=3,4$ $r=,74$ $\bar{x}=19$ $S=4,7$
Mertler & Campbell (2005)	Düzenlenmiş ALI	Başarı Testi Toplam 35 adet senaryo tabanlı soru (toplam beş senaryo)	Ölçme Değerlendirme Standartları (1990)	250 öğretmen adayı	$r=,74$ $\bar{x}=23,9$ $S=4,6$
Zhang & Burry-stock (1997)	API	5'li Likert Tipi Ölçek (1=Hiç Yetenekli Değilim, 5=Çok Yetenekliyim) Toplam 67 madde, yedi alt boyut	Ölçme Değerlendirme Standartları (1990)	311 öğretmen	Ölçeğin tamamı için $\alpha=,97$ ve alt boyutlar için α değeri ,79 - ,93 arasındadır.
Kershaw IV, 1993	AVCQ	5'li Likert Tipi Ölçek (1=Yetkin Değilim, 5=Oldukça Yetkinim) Toplam 26 madde	Ölçme Değerlendirme Standartları (1990)	393 öğretmen	$\alpha=,91$ $\bar{x}=97$ $S=12,9$
Gotch & French (2013)	Ölçme Bilgisi Ölçeği (MKS)	Başarı Testi dört seçenekli 20 soru	-	650 öğretmen	$\alpha=,47$ $\bar{x}=13,95$ $S=2,44$
	Ölçme Öz Yeterlilik Ölçeği (MSS)	7'li Likert Tipi Ölçek (1=Hiç İyi Değilim, 7=Çok İyiyim) Toplam 21 madde	Bandura (2006), Öz Yeterlilik Ölçeği Oluşturma Rehberi	650 öğretmen	$\alpha=,94$ $\bar{x}=112,03$ $S=19,88$
Tao (2014)	Sınıf İçi Değerlendirme Bilgisi Testi (CAK)	Başarı Testi dört seçenekli 27 soru	Ölçme Değerlendirme Standartları (1990)	108 öğretim Elemanı	$\alpha=,74$
Daniel & King (1998)	Ölçme Okuryazarlık Ölçeği	30 adet Doğru-Yanlış Tipi Madde	Değerlendirme Literatürü (Popham, 2015 vb.)	95 öğretmen	$\alpha=,6$ $\bar{x}=18,2$ $S=3,3$

Tablo 5 (Devamı)

Jarr (2012)	Değerlendirme Özgüven Ölçeği	7'li Likert Tipi Ölçek (1= <i>Emin</i> <i>Değilim</i> , 7= <i>Oldukça</i> <i>Eminim</i>) Toplam 15 madde	Bandura (2006), Öz Yeterlik Ölçeği Oluşturma Rehberi	201 öğretmen	$\alpha=,9$ $\bar{x}=64,9$ $S=14,2$
-------------	------------------------------------	--	---	--------------	---

DeLuca vd. (2016a) ve Gürsoy (2017)'den yararlanılarak hazırlanmıştır.

2.3. Konu İle İlgili Ulusal ve Uluslararası Araştırmalar ve Genel Değerlendirme

Değerlendirme okuryazarlığı ile ilgili çalışmalar uluslararası alanda 1991 yılında başlarken ulusal ölçekte kavramın ülkemiz alanyazında yer bulması çok daha geç bir tarihtir. Bütüner vd. (2010) tarafından Campell ve Mertler'in (2005) geliştirdiği Ölçme Değerlendirme Okuryazarlığı Envanteri'nin Türkçe'ye uyarlanması önemli bir adımdır. Değerlendirme okuryazarlığı ile ilgili ulusal ve uluslararası düzeyde yapılan çalışmaların çoğunluğu öğretmen veya öğretmen adaylarının katılımı ile gerçekleştirilmiştir. Alanyazındaki çalışmalara ulusal ve uluslararası düzeyde yapılmış çalışmalar olmak üzere iki temel başlık altında aşağıda detaylı olarak yer verilmiştir.

2.3.1. Ulusal Düzeyde Yapılmış Araştırmalar

Yapılandırmacı yaklaşımın benimsendiği Türk eğitim sisteminde öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki görüşleri ve bu yöntemlerin uygulamasında karşılaşılan sorunlar Gelbal ve Kelecioğlu'nun (2007) araştırmasında incelenmiştir. Ankara ilindeki ilköğretim okullarında görev yapan 242 öğretmene anket uygulanarak veriler toplanılmıştır. Araştırmada öğretmenlerin daha çok geleneksel ölçme ve değerlendirme yöntemlerinde kendilerini yeterli gördükleri ve bu nedenle daha çok bu yöntemleri kullandıkları görülmüştür. Diğer ölçme yöntemlerinin ise öğretmenler tarafından daha az tercih edildiği anlaşılmıştır. Bu durum öğretmenlerin yapılandırmacı anlayışa uygun ölçme ve değerlendirme teknikleri konusunda eğitime ihtiyaç duyduklarını göstermektedir. Ölçme ve değerlendirme uygulamalarında ise en sık karşılaşılan problemler sınıfların kalabalık oluşu ve zaman yetersizliğidir.

Bütüner vd. (2010) arařtırmalarında Campell ve Mertler (2005) tarafından geliřtirilen Ölçme Deęerlendirme Okuryazarlıęı Envanteri'ni Türkçe'ye uyarlamıřtır. Orijinal envanter beř senaryo ve her bir senaryo için hazırlanan yedi soru olmak üzere toplamda otuz beř çoktan seçmeli sorudan oluřmaktadır. Orijinal envanter 1990 yılında belirlenen ölçme deęerlendirme standartlarına göre hazırlanmıřtır. Uyarlama çalıřmasına ise bu standartların MEB tarafından belirlenen standartlarla karřılařtırması yapılmıř ve Türkiye'de belirlenen standartlarla uyumlu olmayanlar uyarlama çalıřmasına dahil edilmemiřtir. Uyarlama çalıřması sonucunda 30 maddelik envanter Türkçe'ye çevrilmiřtir. Çevirisi yapılan envanter iki farklı uygulamada toplamda 260 öęretmen adayına uygulanmıřtır. İkinci uygulamadan elde edilen bulgulara göre envanterin KR20 deęeri ,86, test maddelerinin ortalama güçlük deęeri ,644 ve ortalama ayırt edicilik deęeri ise ,49 olarak bulunmuřtur.

Gül (2011) yüksek lisans tezinde ilköęretim öęretmen adaylarının ölçme deęerlendirme okuryazarlık seviyelerini ve ölçme deęerlendirmeye karřı tutumlarını çalıřmıřtır. Arařtırmada öęretmen adaylarının ölçme deęerlendirme okuryazarlık seviyelerinin tespiti için Plake vd. tarafından geliřtirilen *Deęerlendirme Okuryazarlıęı Ölçeęi* (1993) ve yazar tarafından geliřtirilen *Ölçme Deęerlendirmeye İliřkin Tutum Ölçeęi* kullanılmıřtır. Arařtırma eęitim fakültesinde öęrenim gören 180 öęretmen adayıyla gerçekteřtirilmiřtir. Arařtırma sonuçlarına göre öęretmen adaylarının ölçme deęerlendirme okuryazarlık seviyeleri düşük bulunmuřtur. Deęerlendirme sonuçlarını paylařma yeterlilięi alanında ise öęretmen adaylarının en düşük puanı aldıkları görülmüřtür. Öęretmen adaylarının alanlarına göre ölçme deęerlendirme okuryazarlıkları karřılařtırılmıř, matematik alanında eęitim gören öęretmen adaylarının en yüksek ortalama puana sahip oldukları belirlenmiřtir. Öęretmen adaylarının ölçme deęerlendirmeye karřı tutumları ise ilgi, endiře-kaygı, önem ve yeterlilik faktörleri altında incelenmiřtir. Tutum faktörleri öęretmen adaylarının alanlarına göre karřılařtırmalı olarak analiz edilmiřtir.

Erdoędu ve Kurt (2012) arařtırmalarında öęretmenlerin ölçme ve deęerlendirme yeterlilik algılarını; cinsiyet, öęretmenlik alanı, ölçme dersi alma durumu, etik deęerlere baęlılık, mesleki geliřmeleri takip etme ve ölçme ve deęerlendirmedeki yenilikleri izleme deęiřkenlerine göre incelemeyi amaçlamıřlardır. Bu amaç doęrultusunda Nartgün (2008) tarafından geliřtirilen "Ölçme ve

Değerlendirme Genel Yeterlilik Algısı Ölçeği” ve araştırmacılar tarafından hazırlanan bilgi formu 35’i sınıf ve 154’ü branş öğretmeni olmak üzere toplamda 189 öğretmene uygulanmıştır. Araştırmada öğretmenlerin ölçme ve değerlendirmede genel olarak kendileri orta düzeyde yeterli gördükleri ve istatistiksel teknikler boyutunda ise kendilerini daha az yeterli gördükleri sonucuna varılmıştır. Cinsiyete dayalı anlamlı bir farklılık ise bulunmamıştır. İki dönem ölçme değerlendirme dersi alan öğretmenlerin tek dönem ders alanlara göre ölçme değerlendirme algıları daha yüksektir. Araştırmanın diğer önemli bulguları ise sırasıyla şunlardır: sınıf öğretmenlerinin ölçme teknikleri alanında branş öğretmenlere göre kendileri daha yeterli gördükleri, mesleki etike her zaman uygun davranan öğretmenlerin kısmen uygun davranış gösterenlere göre ölçme değerlendirme algılarının daha yüksek olduğu, mesleki ve ölçme değerlendirmedeki gelişmeleri takip eden öğretmenlerin ölçme değerlendirme algılarının daha yüksek olduğu ve örgütsel bağlılığı yüksek düzeyde olan öğretmenlerin ölçme değerlendirme algılarının daha yüksek olduğu.

Karaman ve Şahin (2014) araştırmalarında öğretmen adaylarının ölçme ve değerlendirme okuryazarlık düzeylerini ve ölçme ve değerlendirmeye ilişkin tutum ve düşüncelerini incelenmişlerdir. Araştırma dördüncü sınıfta öğrenim gören 289 öğretmen adayıyla gerçekleştirilen araştırmada Mertler ve Campell (2005) tarafından geliştirilen “Ölçme ve Değerlendirme Okuryazarlığı Envanteri”, Brown (2002) tarafından geliştirilen “Ölçme ve Değerlendirmeye İlişkin Düşünceler Ölçeği” ve Ogan-Bekiroglu (2009) tarafından geliştirilen “Ölçme ve Değerlendirmeye İlişkin Tutumlar Ölçeği” kullanılmıştır. Araştırma sonuçlarına göre; öğretmen adaylarının ölçme ve değerlendirme okuryazarlık seviyelerinin düşük olduğu, öğretmen adaylarının yapılandırmacı yaklaşıma uygun olan biçimlendirici değerlendirme anlayışını daha çok benimsedikleri ve ölçme değerlendirmeye ilişkin yapılandırmacı anlayışa uygun tutumlar sergiledikleri görülmüştür. Araştırmada öğretmen adaylarının ölçme değerlendirmeyi eğitim ve öğretimin geliştirilmesinde önemli bir araç olarak gördükleri sonucuna varılmıştır.

Ogan-Bekiroğlu ve Süzük (2014) fizik bölümü öğretmen adaylarının değerlendirme okuryazarlığı seviyelerini ve okuryazarlığın pratik uygulamalarda kullanımını karma desenli bir araştırma ile analiz etmiştir. Araştırma 28 öğretmen adayıyla gerçekleştirilmiştir. Nicel veriler ilk yazarın geliştirdiği ölçekle, nitel veriler

ise 11 açık uçlu soru ile elde edilmiştir. Araştırmada ayrıca katılımcılara ölçme değerlendirme dersi kapsamında proje ödevleri hazırlanmış ve bu projelerden elde edilen verilerle, okuryazarlığın pratik uygulamalarda kullanımı analiz edilmiştir. Analiz sonuçlarına göre öğretmen adaylarının ölçme değerlendirme seviyeleri oldukça yüksek bulunmuştur. Katılımcıların ölçme değerlendirme konusunda yapılandırmacı bir anlayışa sahip oldukları görülmüştür. Dolayısıyla ölçme ve değerlendirmeyi öğrenmenin bir parçası olarak gördükleri ve öğrenci kapasitesinin belirlenmesinde farklı ölçme değerlendirme yöntemlerinin kullanılması gerektiğini vurguladıkları belirlenmiştir. Araştırmada nitel verilerden elde edilen sonuçların, nicel verilerle uyumlu olduğu görülmüştür.

Yalvaç (2016) yüksek lisans tezinde fen bilgisi öğretmen adaylarının ölçme değerlendirme okuryazarlık düzeyleri ile bilişsel stratejileri arasındaki ilişkileri incelemiştir. Araştırma 149 öğretmen adayıyla gerçekleştirilmiştir. Araştırmada Bütüner vd. (2010) tarafından Türkçe'ye uyarlanan “Ölçme Değerlendirme Okuryazarlığı Envanteri” ile Witkin (1971) tarafından geliştirilen “Grup Saklı Figürler Testi” kullanılmıştır. Araştırma sonucunda; öğretmen adaylarının ölçme değerlendirme okuryazarlıklarının düşük olduğu, okuryazarlık ile bilişsel strateji arasında ise anlamlı bir ilişki olduğu görülmüştür. Alan bağımsız öğretmen adaylarının alan bağımlı adaylara göre okuryazarlık düzeyleri daha yüksek bulunmuştur. Öğretmen adaylarının en zayıf olduğu ölçme değerlendirme okuryazarlığı boyutunun ise “Ölçme-Değerlendirme Sonuçları ile Öğrenciler, Ebeveynler ve Diğer Eğitimcilerle İletişim Kurabilme” olduğu tespit edilmiştir.

Ülkemizde değerlendirme okuryazarlığı ile ilgili yapılan en kapsamlı araştırmalarından birisi Gürsoy (2017) tarafından gerçekleştirilmiştir. Yazar araştırmasında değerlendirme okuryazarlığının kavramsal bir analizini yapmıştır. Kavramın farklı araştırmacılar tarafından nasıl tanımlandığını, değerlendirme okuryazarı olmanın önemini, öğretmenlerin değerlendirme okuryazarlığını etkileyen nedenleri ve öğretmenlerin kavrama ilişkin tutumlarını incelemiştir. Yazar değerlendirme okuryazarlığının belirlenmesinde temel alınan standartların tarihsel gelişimini analiz etmiş ve ülkemizde belirlenen standartlarla karşılaştırmıştır. Ölçme değerlendirme dersinin ülkemizde yıllar içinde nasıl yapılandırıldığını ve bu durumun öğretmenlerin değerlendirme okuryazarlıklarını nasıl etkilediğini eleştirmiştir. Yazar

araştırmasında değerlendirme okuryazarlığının ölçülmesi amacıyla geliştirilen ölçme araçlarının tarihsel gelişimini de yer vermiştir. Araştırmanın değerlendirme okuryazarlığı kavramına ilişkin yapılacak çalışmalara rehber olması amaçlanmıştır.

Çambay ve Kazanç (2018) araştırmalarında fen bilgisi, sınıf ve okul öncesi programlarında dördüncü sınıfta öğrenim gören toplam 135 öğretmen adayının değerlendirme okuryazarlığı ve değerlendirme okuryazarlığına ilişkin öz yeterlilikleri arasındaki ilişkileri karma desenli bir araştırmayla incelemiştir. Araştırmada veri toplama aracı olarak Bütüner vd. (2010) tarafından Türkçe'ye uyarlanan "Ölçme Değerlendirme Okuryazarlığı Envanteri" ve Kılınç'ın (2011) geliştirdiği "Eğitimde Ölçme Değerlendirmeye Yönelik Öz-Yeterlik Algı Ölçeği" kullanılmıştır. Ayrıca katılımcılarla yarı yapılandırılmış görüşmeler yapılmıştır. Araştırma sonucunda öğretmen adaylarının değerlendirme okuryazarlığı seviyeleri ile değerlendirme okuryazarlığına ilişkin öz yeterlilikleri arasında anlamlı ilişkiler bulunmuştur. Öğretmen adaylarının branşları ile değerlendirme okuryazarlığı seviyeleri ve değerlendirme okuryazarlığına ilişkin öz yeterlilikleri arasında ise anlamlı ilişkiler bulunmamıştır. Araştırmada yarı yapılandırılmış görüşmelerden elde edilen bulguların, nicel analizlerden elde edilen bulguları desteklediği belirtilmiştir.

İzci vd. (2018) araştırmalarında öğretmen adaylarının ölçme değerlendirme okuryazarlıkları ile ölçme değerlendirmeyi kavrama düzeyleri, ölçme değerlendirme yeterlilikleri ve ölçme değerlendirme ders başarıları arasındaki ilişkiyi YEM ile analiz etmişlerdir. Araştırmanın temelini ise Xu ve Brown (2016) tarafından oluşturulan "Uygulamada Değerlendirme Okuryazarlığı" modeli oluşturmaktadır. Araştırmanın katılımcıları ise ölçme ve değerlendirme dersini alan ilköğretim matematik, Türkçe ve bilgisayar ve öğretim teknolojileri bölümlerinde öğrenim gören 203 öğretmen adaydır. Araştırma sonucunda okuryazarlık ile ders başarıları ve kavrama düzeyi arasında anlamlı ilişkiler bulunmuştur. Yeterlilik ve ölçme değerlendirme okuryazarlığı arasında ise anlamlı bir ilişki bulunmamıştır.

Ergül (2019) ise yüksek lisans tezinde öğretmenlerin ölçme değerlendirme okuryazarlık düzeylerini cinsiyet, öğrenim düzeyi, ölçme-değerlendirme ile ilgili hizmet içi eğitim alma durumu, mesleki kıdem ve branş değişkenlerine göre incelemiştir. Araştırmaya 2018-2019 yılında Van ilinde görev yapan 189 ortaokul öğretmeni katılmıştır. Araştırmada veriler araştırmacı tarafından hazırlanan kişisel

bilgi formu, soru inceleme beceri testi ile Mertler ve Campell (2005) tarafından geliştirilen, Bütüner vd. (2010) tarafından uyarlanan Ölçme Değerlendirme Okuryazarlığı Envanteri kullanılarak toplanmıştır. Araştırma sonucunda öğretmenlerin ölçme değerlendirme okuryazarlık seviyeleri düşük çıkmıştır. Öğretmenlerin ölçme değerlendirme okuryazarlık düzeyleri cinsiyet ve öğrenim düzeyine göre anlamlı olarak farklılaşmamakta; ölçme-değerlendirme ile ilgili hizmet içi eğitim alma durumu, mesleki kıdem ve branşa göre ise anlamlı olarak farklılaşmaktadır. Araştırmada matematik branş öğretmenlerine soru inceleme beceri testi verilmiş ve bu test ile öğretmenlerin ölçme değerlendirme okuryazarlıkları arasında pozitif yönde orta düzeyde bir ilişki ($r=,40$; $p<,05$) olduğu görülmüştür.

Azrak ve Yalçinkaya (2019) araştırmalarında sosyal bilgiler öğretmen adaylarının ölçme ve değerlendirme okuryazarlık düzeylerini Mertler ve Campell (2005) tarafından geliştirilen ve uyarlaması Bütüner vd. (2010) tarafından yapılan Ölçme Değerlendirme Okuryazarlığı Envanteri'ni kullanarak incelemişlerdir. Toplamda 282 öğretmen adayına uygulanan envanterde öğretmen adaylarının ölçme ve değerlendirme okuryazarlık düzeyleri; cinsiyet, sınıf düzeyi, ölçme-değerlendirme dersi alma durumu, mezun olunan lise türü ve akademik başarı ortalaması değişkenlerine göre karşılaştırılmıştır. Araştırma sonucunda sosyal bilgiler öğretmen adaylarının ölçme ve değerlendirme okuryazarlık düzeylerinin düşük olduğu görülmüştür. Ölçme değerlendirme okuryazarlık düzeyinin cinsiyet ve mezun olunan lise türüne göre anlamlı olarak farklılaşmadığı, sınıf düzeyi, ölçme-değerlendirme dersi alma durumu ve akademik başarı ortalaması değişkenlerine göre ise anlamlı olarak farklılaştığı sonucuna varılmıştır. Bu sonuçlara göre; üçüncü ve dördüncü sınıfta öğrenim gören öğretmen adaylarının ölçme-değerlendirme okuryazarlık düzeylerinin birinci ve ikinci sınıfta öğrenim görenlere göre daha yüksek olduğu, ölçme-değerlendirme dersi alanların almayanlara göre ölçme-değerlendirme okuryazarlık düzeylerinin daha yüksek olduğu, akademik başarı ortalaması 3,50 ve üzeri olan öğretmen adaylarının ölçme-değerlendirme okuryazarlık düzeylerinin diğer akademik not ortalamalarından anlamlı bir şekilde yüksek olduğu görülmüştür.

Koç (2019; Koç & Bulut 2020) araştırmasında farklı üniversitelerde öğrenim gören Fen Bilgisi öğretmen adaylarının ölçme değerlendirme okuryazarlık düzeylerinin; üniversite, genel ağırlıklı not ortalaması (GANO) ve ölçme

değerlendirme dersi başarı durumuna göre anlamlı bir şekilde farklılaşıp farklılaşmadığını analiz etmiştir. Araştırma yedi farklı üniversitede, dördüncü sınıfta öğrenim gören toplam 290 öğretmen adayı ile gerçekleştirilmiştir. Araştırmada Bütüner vd. (2010) tarafından Türkçe'ye uyarlanan “Ölçme Değerlendirme Okuryazarlığı Envanteri” kullanılmış ve ayrıca “Ölçme Değerlendirme Okuryazarlığına İlişkin Yarı Yapılandırılmış Görüşme Formu” kullanılarak öğretmen adaylarının görüşleri alınmıştır. Araştırma sonucunda öğretmen adaylarının ölçme değerlendirme okuryazarlığı seviyelerinin düşük olduğu; ölçme değerlendirme okuryazarlıklarının üniversiteye ve ders başarısına göre anlamlı bir şekilde farklılaştığı, GANO'ya göre anlamlı bir farklılığın olmadığı görülmüştür. Araştırmada öğretmen adaylarının lisans döneminde tek bir dönemde verilen ölçme değerlendirme dersini yeterli bulmadıklarını ifade etmişlerdir. Araştırmada elde edilen nitel ve nicel veriler öğretmen adaylarının ölçme değerlendirme okuryazarlıklarının düşük olduğunu göstermektedir.

Korkut (2019; Demirdögen & Korkut, 2021) araştırmasında fen edebiyat ve eğitim fakültesi mezunu fen bilimleri öğretmenlerin ölçme değerlendirme okuryazarlık düzeylerini nitel bir araştırmayla incelemiştir. Çoklu durum çalışmasının yürütüldüğü araştırmaya iki eğitim fakültesi ve iki fen edebiyat fakültesi mezunu fen bilimleri öğretmeni katılmıştır. Katılımcı öğretmenler iki devlet ortaokulunda fen dersleri vermektedir. Araştırmada veriler ön-görüşme, gözlem, son-görüşme ve dokümanlar aracılığıyla toplanmıştır. İçerik analizinin yapıldığı araştırmada farklı fakülte mezunu öğretmenlerin; ölçme değerlendirme kullanım amaçlarının farklılaştığı, öğretim sürecinde öğretim rollerinin değiştiği, ölçme değerlendirme araçlarıyla öğrencilerin hazırbulunuşluk düzeylerini ölçtüğü, ölçme değerlendirmeyi bilgi ve uygulama düzeyinde tanılayıcı, biçimlendirici ve özetleyici amaçlar için kullandıkları görülmüştür. Eğitim fakültesi mezunu öğretmenlerin öğrencilere geri bildirim sağlamada öğrencileri destekleyici rol üstlenirken, fen edebiyat fakültesi mezunu olan öğretmenlerin etkili geri bildirim sağlayamadıkları görülmüştür. Araştırmada fen bilimleri öğretmenlerinin ölçme değerlendirme konusunda eğitimine ihtiyaç duydukları sonucuna varılmış ve fen bilimleri öğretmenlerinin ölçme değerlendirme okuryazarlıklarına ilişkin daha fazla araştırma yapılması önerilmiştir.

Tünkler (2019) araştırmasında Sosyal Bilgiler bölümü öğretmen adaylarının değerlendirme okuryazarlık düzeyleri ile değerlendirme öz-yeterlik algılarını incelemiştir. Araştırma 90 öğretmen adayının katılımıyla gerçekleştirilmiştir. Araştırmada Bütüner vd. (2010) tarafından Türkçe'ye uyarlanan "Öğretmen Değerlendirme Okuryazarlığı Envanteri" ile Nartgün (2008) tarafından geliştirilen "Ölçme Değerlendirme Genel Yeterlilik Algısı Ölçeği" kullanılmıştır. Araştırma sonuçlarına göre öğretmen adaylarının değerlendirme okuryazarlığı seviyelerinin düşük olduğu; cinsiyete göre değerlendirme okuryazarlık düzeyinde anlamlı farklılık görülürken, akademik başarıya göre ise anlamlı bir farklılaşmanın olmadığı sonucuna varılmıştır. Değerlendirme öz-yeterlik algılarının ise cinsiyet ve akademik başarıya göre anlamlı bir şekilde farklılaşmadığı görülmüştür. Okuryazarlık düzeyi ile öz-yeterlik algısı arasında anlamlı bir ilişki bulunmamıştır.

Natır (2020) yüksek lisans tezinde ilköğretim matematik öğretmenlerinin ölçme ve değerlendirme okuryazarlıklarına ilişkin sınıf içinde kullandıkları uygulamaları nitel bir araştırmayla incelemiştir. Araştırmada sınıf içi ölçme ve değerlendirme uygulamalarında sunulan içeriğin niteliği ve öğrencilere verilen dönütler incelenmiştir. Araştırma 0-10 yıl arası mesleki deneyime sahip altı ilköğretim matematik öğretmeni ile gerçekleştirilmiştir. Araştırma sonucunda öğretmenlerin ölçme ve değerlendirmeyi işlemsel bilgi düzeyindeki uygulamalarla gerçekleştirdiği, taksonominin üst basamaklarına ilişkin uygulamaları ise derslerinde daha az kullandıkları görülmüştür. Araştırmada öğretmenlerin öğrencilere değerlendirici düzeyde dönütler verdiği, öğrenci gelişimini destekleyecek betimleyici dönütlerin ise daha az sunulduğu görülmüştür.

Acar Güvendir ve Özer Özkan (2021) araştırmalarında öğretmen adaylarının ölçme ve değerlendirme genel yeterlik algıları ile değerlendirme okuryazarlıkları arasındaki ilişkide akademik başarının aracılık rolünü incelemişlerdir. Araştırmada Nartgün (2008) tarafından geliştirilen "Öğretmen Adayları İçin Ölçme ve Değerlendirme Genel Yeterlilik Algısı Ölçeği", Bütüner vd. (2010) tarafından Türkçe'ye uyarlanan "Ölçme Değerlendirme Okuryazarlığı Envanteri" ve öğretmen adaylarının ölçme ve değerlendirme dersi dönem sonu notları kullanılmıştır. Araştırma 126 öğretmen adayıyla gerçekleştirilmiştir. Araştırmada akademik başarının ölçme

değerlendirme yeterlilik algısı ile değerlendirme okuryazarlığı arasındaki ilişkide tam aracı rolü üstlendiği görülmüştür.

Akpolat ve Genç (2022) araştırmalarında KPSS kursuna devam eden öğretmen adaylarının epistemolojik inançları ile ölçme ve değerlendirme okuryazarlığı arasındaki ilişkileri; cinsiyet, yaş, branş, eğitim durumu ve KPSS'ye giriş sayısı değişkenlerine göre analiz etmiştir. Araştırmada Bütüner vd. (2010) tarafından Türkçe'ye uyarlanan "Ölçme Değerlendirme Okuryazarlığı Envanteri" ile Deryakulu ve Büyüköztürk (2005) tarafından Türkçe'ye uyarlanan "Epistemolojik İnanç Ölçeği" kullanılmıştır. Araştırma 216 öğretmen adayının katılımı ile gerçekleştirilmiştir. Araştırma sonucunda, epistemolojik inanç ölçeğinin sadece "tek bir doğru olduğuna ait inanç" alt boyutu ile cinsiyet değişkeni arasında anlamlı ilişki bulunmuştur. Ayrıca, öğretmen adaylarının sadece cinsiyet değişkenine göre ölçme değerlendirme okuryazarlıkları anlamlı bir şekilde farklılaşmaktadır.

Atasoy (2022) yüksek lisans tezinde kimya öğretmenlerinin ve öğretmen adaylarının ölçme değerlendirme okuryazarlıklarını ve ölçme değerlendirmeye ilişkin tutumlarını incelemiştir. Araştırmada Bütüner vd. (2010) tarafından Türkçe'ye uyarlanan "Ölçme Değerlendirme Okuryazarlığı Envanteri" ve Gül (2011) tarafından geliştirilen "Ölçme Değerlendirmeye İlişkin Tutum Ölçeği" kullanılmıştır. Araştırma 119 kimya öğretmeni ve 56 kimya öğretmen adayı ile yürütülmüştür. Araştırma sonucunda öğretmen ve öğretmen adaylarının ölçme değerlendirme okuryazarlığı seviyelerinin düşük olduğu görülmüştür. Ölçme değerlendirme okuryazarlığı ile tutum arasında öğretmenler için anlamlı bir ilişki bulunmazken, öğretmen adayları için ise pozitif yönlü düşük düzeyli, anlamlı bir ilişki bulunmuştur. Araştırmada ölçme değerlendirme okuryazarlığı ile cinsiyet, hizmet yılı, çalışılan okul türü, çalışan okulun niteliği (resmi-özel) ve hizmet içi eğitim alıp almama değişkenleri arasındaki ilişkiler incelenmiştir. Analiz sonuçlarına göre cinsiyet değişkeni için kadın öğretmenler lehine anlamlı bir fark bulunmuştur. Hizmet yılı 15-20 yıl arasında olan öğretmenlerin ve fen lisesinde çalışan öğretmenlerin ölçme değerlendirme okuryazarlığı seviyeleri anlamlı bir şekilde daha yüksek bulunmuştur. Çalışılan okulun türü ve niteliği ile hizmet içi eğitim alıp almamaya göre ise öğretmenlerin ölçme değerlendirme okuryazarlığı seviyesi anlamlı bir şekilde farklılaşmamaktadır. Katılımcıların ölçme değerlendirmeye ilişkin tutumları ile cinsiyet, hizmet yılı, okul türü, hizmet içi eğitim

alıp almama ve çalışılan okulun niteliği (resmi-özel) değişkenlerine göre anlamlı farklılıklar bulunmamıştır.

Değerlendirme okuryazarlığı ile ilgili ulusal düzeyde yapılan araştırmalar incelendiğinde, araştırmaların çoğunluğunun nicel yöntemler ile yürütüldüğü ve Campell ve Mertler (2005) tarafından geliştirilen Bütüner vd. (2010) tarafından Türkçe'ye uyarlanan "Ölçme Değerlendirme Okuryazarlığı Envanteri" kullanılarak gerçekleştirildiği görülmektedir. Senaryo tabanlı 30 sorudan oluşan bu envanterin aslında bir başarı testi olduğu gözden kaçırılmamalıdır. Ayrıca envanterin uyarlanması aşamasında 1990 ölçme değerlendirme standartlarının Milli Eğitim Bakanlığı'nca (2017) belirlenen standartlarla uyumsuzluğu önemli bir eksikliklerdir. Araştırmalar genel olarak öğretmen veya öğretmen adaylarının katılımıyla gerçekleştirilmiştir.

2.3.2. Uluslararası Düzeyde Yapılmış Araştırmalar

Plake vd. (1993) araştırmalarında öğretmenlerin değerlendirme okuryazarlıklarını ölçmek amacıyla 35 maddeden oluşan bir ölçek (*Öğretmen Değerlendirme Okuryazarlığı Ölçeği- TALQ*) geliştirmiştir. Ayrıca araştırmada öğretmenlere ait mesleki tecrübe, cinsiyet vb. kişisel bilgiler toplanılmış, öğretmenlere eğitimsel karar verme sürecinde değerlendirmeye ilişkin algıları ve standart test puanlarının yorumlamadaki tutumlarına ilişkin sorular yöneltilmiştir. Ölçek geliştirilirken AFT, NCME ve NEA'nın 1990 yılında belirlediği öğretmen değerlendirme standartları kullanılmıştır. Geliştirilen ölçeğe toplamda 555 öğretmen katılım göstermiştir. Ölçeğin tümüne ait KR-20 değeri ,54 olarak bulunmuştur. Ölçeğin on maddesinin öğretmenlerin %90 ve daha fazlasının doğru cevapladığı görülmüştür. Bu maddeler öğretmenlerin değerlendirme yöntemi seçimi, standart testlerin uygulanması, velilere öğrenci notlarının açıklanması ve etik olmayan standart test uygulamaları ile ilgilidir. Ölçeğin beş maddesinin ise öğretmenlerin %30 ve daha azı tarafından doğru cevaplandığı tespit edilmiştir. Bu maddelerin değerlendirmenin puanlanması, test sorularının güvenilirliğinin artırılması ve etik olmayan uygulamaların farkına varılması ile ilgilidir. Araştırmada öğretmenlerin standart testlerin uygulanmasına ilişkin algıları ile ölçekten aldıkları toplam puan arasında önemli bir farklılık bulunmamıştır. Diğer taraftan standart testlerden alınan puanların analizinde kendisini daha iyi hisseden öğretmenlerin ölçekten daha yüksek puan aldığı

görülmüştür. Araştırmada yapılan analizlere göre ölçme ve değerlendirme konusunda eğitim almış öğretmenlerin eğitim almayanlara göre ölçekten aldığı puanların ortalaması anlamlı bir şekilde daha yüksektir.

Quilter ve Gallini (2000) araştırmalarında öğretmenlerin değerlendirme okuryazarlıkları ile farklı değerlendirme türlerine ilişkin tutumları arasındaki ilişkiyi incelemişlerdir. 117 öğretmen ile gerçekleştirilen araştırmada, öğretmenlerin değerlendirme okuryazarlık düzeylerini belirlemek için Plake vd. (1993) tarafından geliştirilen Değerlendirme Ölçeği'nin araştırma için düzenlenen versiyonu kullanılmıştır. Benzer şekilde öğretmenlerin değerlendirmeye ilişkin tutumlarını belirlemek için Green ve Stager (1986) tarafından geliştirilen ölçek kullanılmıştır. Öğretmenlerin standartlaştırılmış değerlendirme ve sınıf içi değerlendirme ilişkin tecrübeleri hakkındaki bilgiler, mesleki tecrübeleri ve değerlendirme ile ilgili katıldığı kurs ve çalıştaylar ile ilgili bilgiler de ayrıca araştırmacılar tarafından toplanmıştır. Araştırma sonuçları: öğretmenlerin değerlendirme ile ilgili geçmiş tecrübelerinin öğretmenlerin değerlendirmeye ilişkin tutumlarını doğrudan etkilediğini; geçmişte ölçme ve değerlendirme ile ilgili olumsuz tutumlara sahip öğretmenlerin, farklı değerlendirme türlerine karşı olumsuz tutum sergilemeye meyilli olduklarını; geçmişte standartlaştırılmış test kullanan öğretmenlerin alternatif değerlendirme metotlarını kullanımının düşük olduğunu göstermiştir. Araştırmada değişkenler arasındaki korelasyonlar, kanonik korelasyon analiziyle incelenmiştir.

Webb (2002) araştırmasında Amerika'da Milwaukee eyaletinde bulunan devlet okullarındaki değerlendirme sistemini analiz etmiştir. Değerlendirme sisteminde öğretmenlerin durumunu, sisteme olan yönetsel baskıları son yıllık süreç içerisinde detaylı olarak incelemiştir. Okullarda uygulanan değerlendirme sistemlerini ilkökul, ortaokul ve lise düzeyinde eyalet yönetiminin yaptığı testler ve okullarca yapılan performans değerlendirmeleri başlıkları altında sunmuştur. Okullardan mezuniyet şartlarını incelemiş ve dengeli değerlendirme sistemini kriterleriyle birlikte vermiştir. Dengeli değerlendirme sisteminde öğrenciler hem eyalet bazında yapılan sınavlardan hem de okullarda yapılan değerlendirmelerden sorumlu tutulmuşlardır. Yazar bütün bu değerlendirme süreçlerinde öğretmenlerin değerlendirme okuryazarı olmaları gerektiğinin önemini vurgulamıştır.

Mertler 2003 yılında gerçekleştirdiği araştırmasında öğretmen adaylarının ve öğretmenlerin değerlendirme okuryazarlık düzeylerini Sınıf İçi Değerlendirme Envanterini (*Classroom Assessment Inventory-CALI*) kullanarak karşılaştırmıştır. Sınıf İçi Değerlendirme Envanteri (CALI) 1990 yılında belirlenen değerlendirme okuryazarlığı standartlarına göre geliştirilmiştir (AFT, NCME & NEA, 1990). Araştırma 67 öğretmen adayı ve 197 öğretmen üzerinde gerçekleştirilmiştir. Araştırma sonuçları iki grubun değerlendirme okuryazarlık düzeyleri arasında önemli farklılıklar bulunduğunu ve öğretmenlerin değerlendirme okuryazarlık seviyelerinin öğretmen adaylarından daha yüksek olduğunu göstermiştir. Öğretmenler en yüksek puanı değerlendirme sonuçlarının yorumlanması, puanlanması ve yönetimi becerisinde alırken, öğretmen adayları ise uygun değerlendirme yöntemi seçme becerisinde en yüksek puanı almıştır. Her iki grup en düşük puanı ise geçerli notlandırma sistemi geliştirme becerisinde almıştır.

Mertler 2005 yılında yayımladığı araştırmasında öğretmen adaylarının değerlendirme okuryazarlıklarını ölçmek amacıyla Değerlendirme Okuryazarlığı Envanterini (*Assessment Literacy Inventory-ALI*) geliştirmiştir. Envanter 1990 yılında belirlenen değerlendirme okuryazarlığı standartlarına göre geliştirilmiştir (AFT, NCME & NEA, 1990). Envanter toplam beş senaryo ve her bir senaryoda yedi çoktan seçmeli soru olacak şekilde yapılandırılmıştır. Geliştirilen ölçme aracının iki farklı kurumda deneme uygulaması yapılmış ve envanterin KR20 güvenirlik katsayısı ,74 olarak bulunmuştur. Ölçme aracına ait madde ve test analizleri SPSS (v. 11) ve TAP (v. 5. 2. 7.) programları kullanılarak yapılmıştır.

Volante ve Fazio (2007) araştırmalarında öğretmen adaylarının değerlendirme okuryazarlıklarını etkileyen değişkenleri hazırladıkları anketle analiz etmişlerdir. Araştırmanın katılımcılarını, okul öncesi ve ilköğretim düzeyinde çalışan 69 öğretmen adayı oluşturmaktadır. Dört yıllık bir programda eğitim gören katılımcılar araştırmacıların hazırladıkları ankette yer alan açık uçlu ve değerlendirme ölçeklerinden oluşan toplam dokuz soruya cevap vermiştir. Araştırma sonucunda öğretmen adaylarının değerlendirme konusunda öz yeterliklerinin düşük olduğu, danışman öğretmenlerin aday öğretmenleri değerlendirme konusunda yeterince desteklemedikleri, öğretmen adaylarının daha çok geleneksel değerlendirme türlerini kullanmaya meyilli oldukları görülmüştür. Öğretmen adaylarının değerlendirme

konusundaki eksikliklerinin devlet yönetimince bu alana daha çok yatırım yapılarak giderilmesi gerektiği önerilmiştir.

DeLuca ve Klinger 2010 yılında gerçekleştirdikleri araştırmada Kanada Ontario'da bir üniversitede eğitim gören öğretmen adaylarının ölçme ve değerlendirme eğitimlerinin analizinin yapılması amaçlamıştır. Bu amaç doğrultusunda 288 öğretmen adayına araştırmacılar tarafından hazırlanan 57 maddelik ölçek uygulanmıştır. Öğretmen adaylarının ölçme ve değerlendirme konusundaki özyeterlilikleri ve algılarının belirlenmesi amaçlanmıştır. Ayrıca öğretmen adaylarına ölçme ve değerlendirme dersinde ne tür eğitimler almak istedikleri de sorulmuştur. Katılımcılar ölçme ve değerlendirme dersi alıp almama ve okul öncesi-ilkokul öğretmenlikleri/ortaokul-lise öğretmenlikleri bölümlerinde eğitim görenler olarak dörtlü çapraz gruplara bölünmüştür. Uygulanan ölçek için faktör analizi yapılmış, değerlendirme uygulamaları, değerlendirme teorisi bilgisi ve değerlendirme felsefesi alanları altında faktörler belirlenmiştir. ANOVA yapılarak oluşturulan gruplar arasında değerlendirme uygulamaları, teorisi ve felsefesi alanlarında belirlenen faktörler için anlamlı farklılıklar olup olmadığı belirlenmiştir. Öğretmen adaylarının ölçme ve değerlendirme dersinde daha kapsayıcı ve pratiğe yönelik bir müfredat istedikleri görülmüştür.

Alkharusi 2011 yılında gerçekleştirdiği araştırmasında Plake vd. (1993) tarafından geliştirilen Öğretmen Değerlendirme Okuryazarlığı Ölçeği'ni (*Teacher Assessment Literacy Questionnaire-TALQ*) kullanmıştır. Araştırma Sultan Qaboos Üniversitesi Eğitim Fakültesi'nde farklı alanlarda eğitim gören ve ölçme ve değerlendirme dersi alan 259 öğretmen adayıyla gerçekleştirilmiştir. Araştırmada Öğretmen Değerlendirme Okuryazarlığı Ölçeği'nin psikometrik özelliklerinin incelenmesi ve böylelikle bu ölçekten elde edilen sonuçların öğretmenlerin değerlendirme okuryazarlıklarının tespitinde ne derece kullanışlı olduğunun ortaya çıkarılması amaçlanmıştır. Bu amaç doğrultusunda ölçeğin KR20 değeri ,84; ölçüt geçerliliği değeri için hesaplanan Pearson korelasyon katsayısı ,32 (*ölçek toplam puanları ile öğretmen adaylarının ölçme ve değerlendirme dersi notları arasındaki korelasyon değeri*); Doğrulamalı faktör analizi sonuçları RMSEA = ,08 , TLI(NNFI) = ,89 , CFI = ,89 ve madde-test istatistikleri hesaplanmıştır. Araştırma sonuçları Plake vd. (1993) tarafından geliştirilen ölçeğin farklı bir örnekleme uygulandığında,

öğretmenlerin değerlendirme okuryazarlıklarının tespitinde güvenilir ve geçerli kanıtlar sunduğunu göstermiştir.

Koh (2011) öğretmenlerin değerlendirme okuryazarlıklarının geliştirilmesinin öğrenciler üzerindeki etkilerini boylamsal ve yarı-deneysel bir çalışmayla analiz etmiştir. İki yıllık eğitim öğretim sürecinde sekiz farklı okulda çalışan öğretmenlere somut değerlendirme metotları konusunda eğitim verilmiştir. Araştırmanın deney grubu ve kontrol grubu dörder okulda çalışan öğretmenlerden oluşturulmuştur. Deney grubundaki okullarda çalışan öğretmenlere düzenli ve sürekli olarak iki eğitim yılı boyunca somut değerlendirme metotları ve rubrik kullanımı konusunda eğitim verilmiştir. Her eğitim yılı sonunda bu öğretmenlerle sınıflarında kullandıkları somut değerlendirme metotlarının analizinin yapıldığı değerlendirme toplantıları düzenlenmiştir. Kontrol grubundaki öğretmenlere ise somut değerlendirme metotları konusunda kısa süreli, bir günlük eğitimler ve iki günlük danışma seansları düzenlenmiştir. Araştırma sonucunda uzun süreli ve sürekli eğitim verilen öğretmenlerin değerlendirme okuryazarlıklarının, kısa süreli eğitim gören öğretmenlere göre önemli bir miktarda artış gösterdiği görülmüştür. Bu öğretmenlerin aynı zamanda somut değerlendirme konusunu daha iyi kavradıkları rapor edilmiştir.

Davidheiser (2013) doktora tezinde lise düzeyinde çalışan öğretmenlerin değerlendirme okuryazarlık düzeylerini 1990 yılında belirlenen ölçme değerlendirme standartları çerçevesinde branş ve mesleki kıdem değişkenlerine göre belirlemiş ve karşılaştırmıştır. Yazar tezinde karma desenli bir araştırma yürütmüştür. Tezin nicel boyutunda 180 lise öğretmenine Mertler (2003) tarafından geliştirilen Değerlendirme Okuryazarlığı Envanteri uygulanırken, nitel boyutta ise dört öğretmenle görüşme gerçekleştirilmiştir. Araştırmada öğretmenlerin Değerlendirme Okuryazarlığı Envanteri ortalama puanın 24,51 (maksimum puan 35) olduğu görülmüştür. Araştırmanın nitel bulguları ise *Değerlendirme Varsayımları*, *Profesyonel Gelişim* ve *Değerlendirme Hedefleri* temaları altında gruplandırılmıştır.

Gotch ve French (2014) araştırmalarında öğretmenlerin değerlendirme okuryazarlıkları ile ilgili 1991 ile 2012 yılları arasında yayımlanan çalışmalarını incelemişlerdir. Toplamda 50 farklı araştırmada öğretmenlerin değerlendirme okuryazarlığı ile ilgili 36 farklı ölçme aracının geliştirildiği görülmüştür. Araştırmada bu ölçme araçlarının psikometrik özellikleri; kapsam geçerliği, iç tutarlılık, ölçümlerin

kararlılığı, yapı geçerliği ve öğrenci başarısını yordama gücü olmak üzere altı kriter temel alınarak analiz edilmiştir. Araştırma sonucunda bu ölçme araçlarının geliştirilme sürecinde yeterli güvenilirlik ve geçerlik kanıtlarının sunulmadığı ve dolayısıyla bu ölçme araçlarının zayıf psikometrik özelliklere sahip olduğu öne sürülmüştür. Araştırmacılar öğretmenlerin değerlendirme okuryazarlıklarının belirlenmesinde kullanılabilecek geçerli ve güvenilir ölçme araçlarının geliştirilmesini ve bu konuda daha fazla çalışma yapılmasını önermiştir.

Gareis ve Grant (2015) araştırmalarında değerlendirme okuryazarlığını tanımlamışlar ve kavramın kavramsal çerçevesini oluşturmuşlardır. Değerlendirme okuryazarlığı yeterliliklerini ise ölçmenin türü, ölçmenin kalitesi ve ölçme sonuçları ve sonuçların kullanımı olmak üzere üç başlık altında yapılandırmışlardır. Öğretmen adaylarının değerlendirme konusunda eksik olduklarını belirtmişler ve bu eksikliğin giderilmesi amacıyla devlet üniversitesinde bir kredili ve toplamda 12,5 saat süren, sınıf içi değerlendirmeyi temel alan bir ders oluşturmuşlardır. Ders süresince öğretmen adaylarının değerlendirme ile ilgili kavramları nasıl aldıklarını ve algılarındaki değişimi analiz etmişlerdir. Araştırma sonucunda öğretmen adaylarının değerlendirme okuryazarlık seviyelerinin artırılması gerektiğini belirtmişlerdir.

Xu ve Brown (2016) araştırmalarında değerlendirme okuryazarlığı kavramıyla ilgili 1985 ile 2015 yılları arasında yayımlanan araştırmaları incelemişlerdir. Değerlendirme okuryazarlığının nasıl tanımladığını, kavramla ilgili ne tür araştırmalar yapıldığını ve hangi sonuçlara ulaşıldığını detaylı olarak sunmuşlardır. Değerlendirme okuryazarlığını ise yeni bir kavramsal çerçevede tanımlamışlardır. Uygulamada Öğretmen Değerlendirme Okuryazarlığı Çerçevesi (*Teacher Assessment Literacy in Practice-TALIP*) adını verdikleri sistemde değerlendirme okuryazarlığını altı basamaklı bir piramitsel bir döngüde yapılandırmışlardır. Piramitin en alttan en üst basamağına kadar sırasıyla; bilgi basamağı, öğretmenlerin değerlendirmeye ilişkin algıları, kurumsal ve sosyo-kültürel bağlamlar, temel konsept, öğretmen öğrenmesi, öğretmenlerin kendilerini değerlendirmeci olarak tanımlaması olarak isimlendirilmiştir. Bu yeni kavramsal çerçevenin her bir basamağı için araştırmalar yapılmasını önermişlerdir.

DeLuca vd. (2016a) arařtımlarında İngilizce dilinin konuřulduđu beř ũlkede (Amerika Birleřik Devletleri, Avusturalya, Birleřik Krallık, Kanada, Yeni Zelanda) ve Avrupa anakıtasında eđitim sistemlerinde kullanılan deđerlendirme okuryazarlıđı standartlarını incelemiřlerdir. Arařtırmada ayrıca 1990 yılı ve sonrasında deđerlendirme okuryazarlıđının belirlenmesi amacıyla geliřtirilen ōlçme araçları alanyazınından taranmıřtır. Arařtırmada yũrũtũlen tematik analiz sonucu on beř farklı deđerlendirme standardı ve deđerlendirme okuryazarlıđının ōlçũlmesi amacıyla geliřtirilen sekiz farklı ōlçme aracı belirlenmiřtir. Arařtırmada deđerlendirme okuryazarlıđının çok boyutlu yapısı ve deđerlendirme standardı iin geliřtirilen ōlçme araçlarının bu boyutları ne kadar temsil ettiđi karřılařtırmalı tablolar halinde sunulmuřtur. Tematik analiz sonucu deđerlendirme okuryazarlıđının boyutları: *Deđerlendirme Amaçları, Deđerlendirme Sũreci, Deđerlendirme Sonularının Paylařımı, Deđerlendirmenin Adilliđi, Deđerlendirmede Etik İlkeler, Őlçme Teorisi, Őđrenmenin Deđerlendirilmesi, Deđerlendirme Eđitimi ve Őđretmenlerin Desteklenmesi* olarak isimlendirilmiřtir. Arařtırmada yazarlar deđerlendirme standartlarında on yıllık dŐngũler halinde (1990'dan gũnũmũze) gerekleřen deđiřimleri incelenmiř, standart tabanlı ve sonu odaklı deđerlendirme anlayıřından, biimlendirici deđerlendirme ve Őđretmenlerin deđerlendirme konusunda eđitimi ve desteklenmesi anlayıřına geiř sũrecini ayrıntılı olarak sunmuřlardır.

Wylie ve Lyon (2017) arařtımlarında Őđretmenlerin deđerlendirme okuryazarlıklarının geliřimi iin beř ařamadan oluřan, sıralı bir yapı Őnermiřlerdir. Yapının ilk basamađını biimlendirici deđerlendirme oluřturmaktadır. Biimlendirici deđerlendirmenin en sık yapılan ve en dũřũk riskli deđerlendirme tũrũ olduđu ve bu nedenle biimlendirici deđerlendirmenin temel alınarak Őđretmenlerin deđerlendirme okuryazarlıđının geliřimi iin ařamalı bir yapı oluřturulabileceđi belirtilmiřtir. Arařtırmada dũřũk riskli deđerlendirmeden yũksek riskli deđerlendirmeye gŐre sıralanan ařamalar ise řunlardır: biimlendirici deđerlendirme, Őđretim programına uygun performans deđerlendirmeleri, sınıf ii sonu odaklı deđerlendirme, ara deđerlendirmeler ve ulusal boyutta yũksek riskli deđerlendirmeler.

Xu ve Brown (2017) arařtımlarında Plake vd. (1993) tarafından geliřtirilen *Őđretmen Deđerlendirme Okuryazarlıđı Őleđi*'nin (TALQ) uyarlanmıř formunu, in'de ũniversitede alıřan İngilizce Őđretmenlerinin deđerlendirme okuryazarlıđını

ölçmek amacıyla kullanmıştır. Uyarlanmış ölçek 24 maddeden oluşmaktadır ve 891 öğretmene uygulanmıştır. Ölçeğin psikometrik özellikleri, madde tepki kuramına göre 3PL modelinde incelenmiştir. DFA yapılarak ölçeğin faktör yapısına bakılmıştır. Analiz sonuçlarına göre TALQ'nun orijinal maddelerinin psikometrik özelliklerinin düşük olduğu ve Çin'de kullanımında detaylı bir şekilde gözden geçirilmesi gerektiği belirtilmiştir. Ayrıca TALQ'un çoklu faktör yapısının araştırmada doğrulanmadığı, dolayısıyla ölçek maddelerinin gözden geçirilmesi ve yeniden yazılması gerektiği belirtilmiştir.

Xu ve He (2019) araştırmalarında öğretmen adaylarının, öğretmenlik uygulaması dersi öncesi ve sonrasında değerlendirme okuryazarlıklarındaki değişimi incelemiştir. Araştırma üç öğretmen adayıyla gerçekleştirilmiştir. Veri toplama aracı olarak öğretmen adaylarıyla odak grup görüşmesi ve yarı yapılandırılmış görüşmeler yapılmış, ayrıca sınıf içi gözlem formlarından ve ders planlarından yararlanılmıştır. Araştırma sonucunda öğretmen adaylarının öğretmenlik uygulaması dersi sonrasında değerlendirme okuryazarlık seviyelerinin arttığı görülmüştür. Bu durumun nedenleri ise; kişisel faktörler (değerlendirmede aktif görev alma), deneyimsel faktörler (öğrencilerle etkileşim, okul ortamında değerlendirme pratiği yapma, geri bildirim vb.), bağlamsal faktörler (danışmanlık, gerçek sınıf ortamı, okul değerlendirme kültürü, ulusal değerlendirme politikası) olarak sıralanmıştır.

Adamson (2020) araştırmasında, Mertler ve Campbell (2005) tarafından geliştirilen Değerlendirme Okuryazarlığı Envanteri'ni kullanmıştır. Araştırmacı tarafından envanterden bir senaryo silinmiş ve kalan senaryolara biçimlendirici değerlendirme ile ilgili maddeler eklenmiştir. Oluşturulan envanter mesleklerinde 0-4 yıllık deneyime sahip, Alberta okul bölgesinde çalışan 10 öğretmene uygulanmış ve bu öğretmenlerin değerlendirme okuryazarlığında eksik kaldığı alanlar belirlenmiştir. Araştırma sonucuna göre, değerlendirme yöntemlerin seçimi/geliştirilmesi, uygulama, puanlama, sonuçların analizi/kullanımı boyutları öğretmenlerin eksik kaldığı alanlar olarak ifade edilmiştir.

DeLuca vd. (2021) ABD, Kanada ve Çin'deki öğretmenlerin değerlendirmeye ilişkin yaklaşımlarını örtük sınıf analizi ile incelemiştir. Araştırmacılar bu üç ülke arasındaki göçmen öğrenci sayısının son yıllarda arttığına dikkat çekerek, bu öğrencilerin farklı değerlendirme yaklaşımlarına maruz kaldıklarını belirtmişlerdir. Bu

nedenle bu ülkelerdeki öğretmenlerin değerlendirme yaklaşımlarının, *Sınıf İçi Değerlendirme Yaklaşımları Envanteri*'ni (Approaches to Classroom Inventory, ACAI) kullanarak belirlemişlerdir. Araştırma 710 öğretmenin katılımıyla gerçekleştirilmiştir. Araştırma sonucunda öğretmenlerin değerlendirme yaklaşımları beş sınıfta gruplandırılmış ve ülkelere göre karşılaştırılmalı olarak analiz edilmiştir. Bu sınıflandırmalar sırasıyla; *öğretmen-merkezli değerlendiriciler*, *çekingen değerlendiriciler*, *kısmen öğrenci-merkezli değerlendiriciler*, *çoğunlukla öğrenci-merkezli değerlendiriciler* ve *istekli değerlendiriciler*'dir.

Atjonen vd. (2022) 168 öğretmen adayının değerlendirme Okuryazarlığı düzeylerini nitel bir araştırma ile incelemiştir. Öğretmen adaylarının değerlendirme okuryazarlık düzeyleri bir değerlendirme modülü aracılığıyla değerlendirilmiş, öğretmen adaylarının değerlendirme hakkında görüşleri yazılı olarak toplanmıştır. Elde edilen veriler Xu ve Brown'un (2016) geliştirdiği *Uygulamada Değerlendirme Okuryazarlığı Modeli*'nin *Bilgi*, *Öğretmenlerin Değerlendirmeye İlişkin Görüşleri* ve *Öğretmenin Öğrenmesi* basamaklarına göre analiz edilmiştir. Araştırma sonuçlarına göre; öğretmen adaylarının değerlendirme hakkında yeterli bilgi düzeyine sahip olduğu, akran öğrenmesinin değerlendirme konusunda işlevsiz kaldığı ve kısa süreli değerlendirme eğitimlerinin bile değerlendirme okuryazarlığına olumlu etkilediği belirlenmiştir.

Uluslararasıda yapılan çalışmalarda öncelikle değerlendirme okuryazarlığının kavramsal analizinin yapıldığı, ölçme değerlendirme standartlarının incelendiği ve bu standartlara göre ölçme araçları geliştirildiği görülmektedir. Yapılan araştırmaların katılımcıları öğretmen veya öğretmen adaylarıdır. Değerlendirme okuryazarlığı düzeyleri, değerlendirme okuryazarlığını etkileyen değişkenler ve bu değişkenler arasındaki ilişkiler, okulların sahip olduğu değerlendirme sistemlerinin analizi temel araştırma konularını oluşturmaktadır.

2.4. Psikolojik Ölçme Araçlarının Uyarlanması

Teknolojinin gelişmesiyle birlikte mesafeler kısalmış, iletişim olanakları artmıştır. Günümüzde farklı dil ve kültür gruplarından insanlar daha rahat bir şekilde iletişim kurabilmektedir. Bir akademik çalışma grubunun farklı ülkelerden gelen katılımcılardan oluşması sıradanlaşmıştır. Küreselleşmenin bu boyuta ulaştığı bir

dönemde herhangi bir kültür veya dil grubu için geliştirilen bir ölçme aracının farklı kültür veya dil gruplarında uygulanması ve elde edilen sonuçların karşılaştırılması istenilen bir durumdur. Örneğin, 2021 PIRLS sınavı yetmiş yakın ülke ve kıyaslama kuruluşunun katılımıyla gerçekleştirilmiştir (Mullis & Martin, 2019). Bu sınavda yer alan ölçme araçlarının sadece farklı dillere çevrilerek uygulanması ve elde edilen sonuçların karşılaştırılması hatalı olacaktır. Ölçme sonuçlarına dilsel, psikolojik ve kültürel öğelerden kaynaklı hatalar karışacaktır. Ölçme sürecine karışacak bu hatalardan kurtulmanın yolu ise ölçme aracını uyarlamaktan geçmektedir.

Erkuş (2019) uyarlamayı,

bir psikolojik özelliği ölçmek amacıyla geliştirilmiş ve sağlam olduğu varsayılan bir ölçeğin, aynı dil ve kültürde değişen zamanlar ile başka demografik gruplar, aynı dil fakat farklı kültürler veya farklı dil ve farklı kültürler için irdelenip kullanıma uygun hale getirilmesi (s. 4) olarak tanımlamıştır.

Bu tanıma göre psikolojik özelliklerin dile, kültüre ve uygulandığı zamana göre değişim gösterdiği, dolayısıyla bir kültür veya dil grubu için geliştirilen bir ölçeğin farklı bir kültür ve dil grubunda uygulandığında uyarlamaya ihtiyaç duyulduğu görülmektedir. Başarı, zekâ gibi psikolojik yapıların farklı kültürlerde farklı davranış göstergelerinden oluştuğu ve uyarlama sürecinde bu göstergelerin dikkate alınması gerektiği vurgulanmaktadır. Uyarlamada amaç anlamsal, dilsel, psikometrik ve psikolojik boyutlarda kaynak ve hedef grup arasındaki farkı en aza indirmektir (Hambleton & Jong, 2003).

Hambleton ve Patsula (1999) ise ölçme araçlarının uyarlanmasının beş temel sebebinin olduğunu belirtmiştir. Bu sebepler sırasıyla verilmiştir.

1. İkinci bir dile ölçek uyarlamanın yeni baştan bir ölçme aracını geliştirmekten daha ucuz ve hızlı olması.
2. Uyarlanmış testin amacı kültürler ya da uluslararası değerlendirme yapmak olduğunda, uyarlanmış testin ikinci dilde eşdeğer bir test üretmenin en etkili yolu olması.
3. İkinci bir dilde ölçek geliştirme konusunda yeterli uzmanlığın olmaması.
4. Uyarlanması hedeflenen ölçeğin tanınan, sonuçlarına güvenilen ve bu nedenle birçok dile uyarlanan bir ölçek olması.
5. Ölçmede eşitliğin, bir ölçeğin farklı dillere uyarlanmış formlarının kullanılmasıyla sağlanması.

Ölçek geliştirme; amacın belirlenmesi, ilgili kavramsal yapının tanımlanması, ölçek geliştirme tekniğine karar verilmesi, pilot uygulama, geçerlik ve güvenirlik kanıtlarının elde edilmesi gibi adımlardan oluşan zorlu ve kapsamlı bir süreçtir (Erkuş, 2021; Şahin, 1994). Bu nedenle araştırmacılar daha kolay ve hızlı bir yöntem olan ölçek uyarlamayı tercih edebilmektedir. Akademisyenler üzerindeki yayın baskısı da bu durumu etkilemektedir (Erkuş, 2007). Öner (2008) yaptığı tarama araştırmasında Türkiye’de 2005 yılında kadar kullanılan 350 psikolojik ölçme aracından sadece dörtte birinin özgün olarak geliştirildiğini, geri kalanın ise çeviri veya uyarlama olduğunu tespit etmiştir (akt. Erkuş, 2019).

Spielberger’in Durumluk-Sürekli Kaygı Envanteri ve Wechler Çocuklar İçin Zekâ Ölçeği gibi alanyazında tanınmış ölçme araçlarının çok sayıda dile uyarlamasının yapıldığı görülmektedir (Hambleton & Jong, 2003).

Uyarlama çalışmalarının yapılmasının bir başka nedeni ise kültürlerarası karşılaştırma çalışmalarının gerçekleştirilmek istenmesidir (Van de Vijver & Hambleton, 1996). Örneğin; PISA, PIRLS, TIMMS gibi geniş ölçekli sınavlarda kullanılan ölçme araçları farklı kültür ve dil grupları için uyarlanmaktadır. Bu sınavlardan elde edilen sonuçlarla farklı kültür ve dil gruplarından gelen bireylerin sınav performansları karşılaştırılabilmektedir. Aynı şekilde Microsoft gibi firmalarca hazırlanan akreditasyon sınavlarının birçok ülkede kullanımı, uyarlama çalışmalarının yapılmasının başlıca nedenlerindedir (Fitzgerald, 2004; Hambleton & Jong, 2003). Son olarak ölçmede adillığın sağlanması amacıyla aynı sınavın farklı dil formlarının hazırlanması uyarlama çalışmaları ile mümkün olabilmektedir (International Test Commission, 2017).

Uyarlama her ne kadar popüler ve tercih edilen bir uygulama olarak görülse de uyarlamamanın da kendine özgü problemleri mevcuttur. Ölçek uyarlamadaki en temel sorun ölçeğin geliştirildiği dilden farklı bir dile çevirisinin yapılmasıdır (Öztürk vd., 2015). Bu sorunla başa çıkmanın yolu ise çevirinin her iki dile ve kültüre hâkim, ölçek uyarlamayla ilgili bilgiye sahip çevirmenlerce yapılmasından geçer (Deniz, 2007). Bu özelliklere sahip çevirmenleri bulma zorluğu, uyarlama sürecinde yaşanan diğer bir sıkıntıdır. Bu durum uyarlama sürecine konu ile ilgili uzmanların dâhil edilmesiyle giderilebilir. Fakat bu uzmanların birarada ve işbirliğine dayalı bir şekilde çalışmaları sağlanmalıdır. Uyarlamada karşılaşılan bir diğer problem ise uyarlanan ölçeğin

psikometrik özelliklerinin orijinal ölçekten daha düşük çıkmasıdır (Deniz, 2007). Yüksek geçerlik ve güvenilirlik değerlerine sahip bir ölçme aracının uyarlanmış halinin de yüksek güvenilirlik ve geçerlik değerlerine sahip olması beklenir. Hambleton ve Patsula'ya (1999) göre ise ölçek uyarlamada birer efsaneye dönüşmüş beş hatalı kabul bulunmaktadır. Bu hatalı kabuller sırasıyla:

1. İkinci bir dilde test geliştirmektense, her zaman mevcut testin uyarlamasının yapılması tercih edilmelidir.
2. Her iki dili de bilen herhangi biri ölçeğin kabul edilebilir bir çevirisini yapabilir.
3. Çevirisi iyi yapılmış bir testten elde edilen puanlar kültürlerarası karşılaştırmalarda ikinci dilde veya kültürde geçerli ölçümler verir.
4. Yapılar evrenseldir ve bu nedenle tüm testler diğer dillere ve kültürlere çevrilebilir.

Ölçek uyarlamada yaşanan bu sıkıntılar ve hatalı kabuller, uyarlama işlemlerinin belirli bir sistem içerisinde adım adım yapılmasıyla önlenabilir. Böylece uyarlanan ölçeklerin farklı araştırmacılar tarafından gözden geçirilmesi ve kullanımı kolaylaşacaktır. Nitekim Türkiye'de eğitim ve psikoloji dergilerinde yayımlanmış çeşitli ölçek geliştirme ve uyarlama çalışmalarında sistematik bir bütünlük olmadığı ve eksikler görüldüğü rapor edilmektedir (Acar Güvendir ve Özer Özkan, 2015; Çüm ve Koç, 2013; Öztürk vd., 2015).

2.4.1. Ölçek Uyarlama Adımları

Ölçek uyarlama çalışmalarında takip edilecek adımlar özünde aynı olmakla birlikte araştırmacılarca farklı şekillerde sınıflandırılmaktadır (Çapık vd., 2018; Erkuş, 2019; Hambleton & Patsula, 1999; International Test Commission, 2017). Örneğin Erkuş (2019) ölçek uyarlamayı üç adımda sınıflandırmıştır. Bu adımlar:

1. Dilsel Çeviri
2. Öndeneme, Deneme Uygulaması ve Örneklem
3. Yapı/Madde Analizi'dir.

Çapık vd. (2018) ise kültürlerarası ölçek uyarlama aşamalarına ilişkin güncellenmiş bir rehber oluşturmuştur. Bu rehberde ölçek uyarlama yedi adımdan oluşmaktadır. Bu adımlar sırasıyla:

1. Çeviri
2. Semantik(anlamsal) açıklamalar
3. Uzman paneli
4. Geri çeviri
5. Pilot uygulama ve bilişsel inceleme
6. Son sürümün elde edilmesi
7. Dokümantasyon'dur.

Hambleton ve Patsula (1999) ise Uluslararası Test Komisyonun yayımladığı Test Çeviri ve Uyarlama Rehberi'nden (1. baskı) (International Test Commission, 2005) ve farklı deneysel çalışmalardan yararlanarak ölçek uyarlama adımlarını oluşturmuştur. Ölçek uyarlama adımlarının sürekli bir gelişim içerisinde olduğunu vurgulamışlar ve adımların farklı çalışmalarda kullanımıyla birlikte güncellenebileceği belirtilmişlerdir. Hambleton ve Patsula (1999) ölçek uyarlamayı 13 adımda açıklamıştır. Bu adımlar sırasıyla verilmiş ve açıklamalar yapılmıştır.

1. Adım: İlgili dil ve kültür gruplarında yapı eşdeğerliğinin sağlandığından emin olunması

İlgili yapının kaynak ve hedef grupta var olup olmadığı sorgulanmalı, eğer yapı mevcutsa yapısal eşitlik sağlanmalıdır. Uyarlama çalışmasının temelleri yapının her iki grupta var olması ve eşit bir şekilde temsil edilmesiyle atılır. Eğer bu durum sağlanamıyorsa uyarlama çalışması yapılmamalıdır. Harkness (1998) yapısal eşitliğin sağlanmasında araştırmacının cevaplaması gereken dört soru olduğunu öne sürmüştür (akt. Hambleton ve Patsula, 1999).

- Araştırmacının ölçmeyi hedeflediği yapı her iki grupta da var mıdır?
- Hedef ve kaynak grubu bu yapı üzerinde karşılaştırmak anlamlı mıdır?
- Bu yapı için yapılacak herhangi bir kültürlerarası karşılaştırma anlamlı olacak mıdır?
- Ölçülen yapı karşılaştırılacak tüm kültür gruplarında aynı anlama gelmekte midir?

Uyarlama çalışması yapmayı hedefleyen araştırmacılar bu dört soruyu cevaplandırmadan çalışmalarına başlamamalıdır. Hambleton ve Patsula (1999) kaynak ve hedef gruplar arasındaki yapısal eşitliğin her iki gruba da hâkim olan araştırmacılar ve uzmanlarca sağlanabileceğini belirtmiştir.

2. Adım: Yeni bir test geliştirmeye ya da var olan bir testin uyarlanmasına karar verme

Ölçülmek istenen yapıyla ilgili test geliştirmenin mi yoksa var olan bir testi uyarlamanın mı daha anlamlı olduğu sorunsalı yapılacak olan ölçmenin amacıyla ilgilidir. Eğer kültürlerarası karşılaştırma yapmak amaçlanıyorsa uyarlama çalışması tercih edilmelidir. Öte yandan uyarlama, test geliştirmeden her ne kadar daha pratik bir yol olarak görünse bile hedef ve kaynak grup arasındaki dil farklılıkları uyarlamayı güçleştirebilir. Dolayısıyla test geliştirmenin ve uyarlamanın avantajları ve dezavantajları göz önünde bulundurulmalı, amaca uygun olan seçilmelidir (Şahin, 1994).

3. Adım: Nitelikli çevirmenlerin seçimi

Uyarlama çalışmalarında en çok hata yapılan adımlardan birisidir. Nitelikli çevirmen her iki dile ve kültüre hâkim, ölçülecek yapı ve test geliştirme konusunda bilgili kişi demektir. Test geliştirme konusunda bilgili çevirmen bulmanın güçlüğü düşünüldüğünde, çevirmenlere bu konuda eğitim verilebilir ya da çeviri sürecine bu konuda uzman bir kişi eklenebilir. Uyarlaması yapılacak ölçme aracının hedef dile çevirisi panel veya komiteler düzenlenerek de gerçekleştirilebilir. Böylece tek bir çevirmenin yapacağı çeviri yerine birden fazla uzmanın yapmış olduğu çeviriler kullanılabilir.

4. Adım: Testin çevrisinin yapılması ve uyarlanması

Hambleton ve Patsula (1999) çeviride *İleriye Doğru Çeviri* ya da *Geriye Doğru Çeviri* yöntemlerinin kullanılabilceğini belirtmiştir. İleriye doğru çeviride ölçme aracı orijinal dilden hedef dile bir veya birden fazla çevirmen tarafından çevrilir. Yapılan çeviri farklı bir veya birden fazla çevirmen tarafından kontrol edilir ve gerekliyse düzenlemeler yapılır. Geriye doğru çeviride ise ölçme aracı öncelikle orijinal dilden hedef dile bir veya birden fazla çevirmen tarafından çevrilir. Daha sonra ölçek hedef dilden orijinal dile farklı çevirmenler tarafından tekrar çevrilir. Ölçme

aracının orijinal hali ile geri çevirisi yapılmış hali karşılaştırılır ve denkliklerine karar verilir (Öztürk vd., 2015).

Hambleton ve Patsula (1999) ileriye doğru yapılan çevirilerde hedef dile odaklanılmasından dolayı avantaj sağlandığını, fakat çeviri de her iki yöntemin de kullanılmasını gerektiğini önermiştir.

5. Adım: Uyarlanan testin gözden geçirilmesi ve gerekli değişikliklerin yapılması

İleri veya geri çeviri yöntemleri kullanılarak çevirisi yapılmış testler, farklı çevirmenlerce gözden geçirilmeli ve gerekliyse düzenlemeler yapılmalıdır.

6. Adım: Pilot uygulamanın yapılması

Ölçek geliştirildiği hedef grubu temsil eden daha küçük bir grupta uygulanmalıdır. Pilot uygulamayla ölçeğin uygulanma koşullarına (test yönergesi, zaman vb.) ve ölçek maddelerine ilişkin geri dönütler alınır. Toplanan bu geri dönütler ışığında ölçek ile ilgili düzenlemeler gerçekleştirilir.

7. Adım: Testin uygulanması

Uyarlanması yapılmış test, amaca uygun daha büyük bir örnekleme uygulanmalıdır.

8. Adım: Kaynak ve hedef gruplarının elde edilen sonuçların karşılaştırılması için istatistiksel bir desen seçimi

Kültürlerarası bir karşılaştırma, test normlarının veya performans standartlarının karşılaştırılması söz konusu olduğunda, uygun bir istatistiksel desen seçilerek testin farklı formlarından elde edilen puanlar karşılaştırılmalıdır. En çok kullanılan üç desen sırasıyla:

- İki Dilli Grup Deseni
- Tek Dilli Karşılaştırmalı Grup Deseni
- Tek Dilli Grup Deseni'dir.

9. Adım: Kültürler arası karşılaştırma amacıyla dilsel eşdeğerliğin sağlanması

Test formunun orijinal ve uyarlanmış halleri aynı örneklem üzerinde uygulanarak testlerden alınan puanlar karşılaştırılır. Herhangi bir test maddesinin dilden kaynaklı bir yanlılığa sebep olup olmadığı istatistiksel analizlerle belirlenir.

Yanlılık gösterdiği tespit edilen maddelerin yeniden çevirisi yapılmalı ve analizler tekrarlanmalıdır (Deniz, 2007).

10. Adım: Geçerlik kanıtlarının elde edilmesi

Uyarlaması yapılan ölçeğin güvenirlik ve geçerlik kanıtları elde edilmelidir. Faktör analizi, deneysel ya da ilişkisel analizler yapılmalı, geçerlik kanıtları sunulmalıdır.

11. Adım: Uyarlama sürecinin belgelendirilmesi ve kullanıcılar için rehber hazırlanması

İlk on adımda elde edilen sonuçlarla ilgili belgeler hazırlanmalı ve uyarlanmış test için bir kullanıcı rehberi oluşturulmalıdır. Hazırlanacak rehber, testin uygulama yönergesine ve sonuçların nasıl yorumlanacağına ilişkin bilgiler içermelidir. Bu adım oldukça önemlidir, fakat yine de göz ardı edilmektedir.

12. Adım: Kullanıcıların eğitimi

Uygun koşullar mevcutsa test kullanıcılarına eğitim verilmelidir. Her ne kadar uyarlama süreçlerinin belgelendirilmesi ve rehber kılavuz hazırlanması söz konusu olsa da, test kullanıcılarına eğitim verilmesi oldukça önemlidir.

13. Adım: Uyarlanmış testin sürekli izlenmesi

Kültürlerarası karşılaştırmalar kesitsel bir araştırma olmasına rağmen, zekâ ve kişilik testleri gibi bazı testler uyarlandıkları dilde uzun süreli kullanılabilir. Dolayısıyla araştırmacılar uyarladıkları testlerde zamanla oluşabilecek hatalara karşı testleri sürekli izlemeli ve gerekli gördüklerinde yeniden geçerlik ve güvenirlik çalışmaları gerçekleştirmelidir.

Hambleton ve Patsula'nın (1999) ölçek uyarlama adımlarından sonra ise Uluslararası Test Komisyonunun (International Test Commission-ITC) kısaca tarihi anlatılacak ve ITC tarafından yayımlanan ölçek uyarlama rehberi ile ilgili bilgiler verilecektir.

2.4.2. Uluslararası Test Komisyonu (ITC)

Uluslararası Test Komisyonu (ITC) resmi olarak 1976 yılında Uluslararası Psikoloji Bilimleri Birliği Kongresi'nde (IUPsyS) kurulmuştur (Oakland vd., 2001). ITC'nin Temmuz 2000 itibarıyla 27'si tam (ulusal profesyonel psikoloji kurumları) ve

58'i bağılı kuruluş (test ile ilgili çalışan diđer komisyonlar, üniversiteler, şirketler) olmak üzere 85 üyesi bulunmaktadır (Deniz, 2007; Oakland vd., 2001).

ITC'nin temel amacı; üyeleri ile üyesi olmayan fakat test ilgili uygulamaları geliştirmek isteyen topluluklar, kuruluşlar ve bireyler arasında test geliştirme ve kullanımına ilişkin bilgi alışverişine yardımcı olmaktır. Bu amaç doğrultusunda ITC tarafından Testlerin Çevirisi ve Uyarlanması, Test Güvenliđi, Teknoloji Tabanlı Deđerlendirme vb. konularda rehberler yayımlamakta, çeşitli uluslararası dergi ve kitap çalışmaları yürütölmektedir (Oakland vd., 2001).

ITC, 1992 yılında Testlerin Çeviri ve Uyarlanması konusunda bir rehber hazırlama projesi başlatmıştır. Bu projede aralarında ölkemizden Prof. Dr. Işık Savaşır'ın da yer aldığı 13 üye yer almıştır. Proje sonunda 22 maddeden oluşan Test Çeviri ve Uyarlama Rehberi (1. baskı) oluşturulmuştur (Deniz, 2007; Oakland vd., 2001). Rehberin ilk baskısı 2005 yılında yayımlanmıştır (International Test Commission, 2005).

ITC tarafından hazırlanan rehber, 2005 ve 2015 yılları arasında test geliştirme ve uygulama süreçlerinde yaşanan gelişmeler doğrultusunda güncellenmiştir. Rehberin ikinci baskısı 2017 yılında yayımlanmıştır (International Test Commission, 2017). Rehberin ikinci baskısına ait adımlar sırasıyla açıklanmıştır.

ITC tarafından hazırlanan rehberin ikinci baskısında test çeviri ve uyarlama süreci 18 adımda açıklanmıştır. Bu adımlar altı kategoride sınıflandırılmıştır. Bu kategoriler sırasıyla *Önkoşullar*, *Test Geliştirme*, *Dođrulama*, *Uygulama*, *Puanlama Ölçekleri ve Yorumlanması*, *Belgelendirme*'dir.

Önkoşullar:

1. Uyarlama sürecine başlamadan önce ölçme aracının fikri mülkiyet hakları sahibinden gerekli izin alınmalıdır.
2. Test tarafından ölçölen yapının tanımı ve içeriđi ile hedef gruptaki madde içerikleri arasındaki örtüşme miktarının, ölçme aracından elde edilecek puanlarının amaçlı kullanımı için yeterli olup olmadığı deđerlendirilmelidir.
3. Çalışılan gruplarda testin amaçlanan kullanımlarıyla ilgisi olmayan kültürel ve dilsel farklılıkların etkisi en aza indirilmelidir.

Test Geliştirme:

4. İlgili uzmanlar seçilerek çeviri ve uyarlama süreçlerinde çalışılan gruplardaki dilsel, psikolojik ve kültürel farklılıkların göz önünde bulundurulduğundan emin olunmalıdır.
5. Çalışılan gruplarda test uyarlamada uygunluğu en üst düzeye çıkarmak için uygun çeviri yöntemleri kullanılmalıdır.
6. Çalışılan gruplarda test yönergesinin ve madde içeriklerinin aynı anlama geldiğine ilişkin kanıt sunulmalıdır.
7. Çalışılan gruplarda madde türlerinin, derecelendirme ölçeklerinin, puanlama kategorilerinin, test kurallarının, test uygulama biçimlerinin ve diğer süreçlerin uygunluğuna ilişkin kanıt sunulmalıdır.
8. Uyarlanmış testte gerekli düzenlemelerin yapılabilmesi için pilot uygulama yapılmalıdır. Toplanan verilerle madde analizi, güvenilirlik ve geçerlik analizleri gerçekleştirilmelidir.

Doğrulama:

9. Ölçme aracının kullanım amacına uygun ve yeterli büyüklüğe sahip bir örneklem seçilmelidir.
10. Çalışılan gruplarda yapı eşdeğerliği, yöntem eşdeğerliği ve madde eşdeğerliği hakkında ilgili istatistiksel kanıtlar sunulmalıdır.
11. Çalışılan gruplarda testin uyarlanmış formunun normlarını, güvenilirliğini ve geçerliliğini destekleyen kanıtlar sunulmalıdır.
12. Bir testin farklı dil formlarının puan ölçekleri ilişkilendirirken uygun bir test eşitleme tasarımı ve veri analizi süreçleri kullanılmalıdır.

Uygulama:

13. Puanlardan elde edilen çıkarımların geçerliliğini etkileyebilecek, uygulama süreçlerinden ve yanıtlama türlerinden kaynaklanan kültür ve dille ilgili sorunları en aza indirmek için uygulama materyalleri ve yönergeleri hazırlanmalıdır.
14. Çalışılan gruplarda dikkatlice takip edilmesi gereken test koşulları belirtilmelidir.

Puanlama Ölçekleri ve Yorumlanması:

15. Herhangi bir grup puan farkı elde edilen bütün bilgilere göre yorumlanmalıdır.
16. Çalışılan gruplar arasındaki puanlar sadece puanların rapor edildiği ölçekte değişmezlik düzeyi belirlendiğinde karşılaştırılmalıdır.

Belgelendirme:

17. Bir test başka bir evrende kullanılmak üzere uyarlandığında, eşdeğerliği desteklemek için elde edilen kanıtların bir açıklaması da dahil olmak üzere herhangi bir değişikliğin teknik belgeleri sağlanmalıdır.
18. Uyarlanmış bir testin farklı bir evrende kullanımında iyi uygulamayı desteklemek amacıyla test kullanıcıları için belgeler sunulmalıdır.

Bu araştırmada Hambleton ve Patsula (1999) ve ITC (2017) tarafından hazırlanan ölçek uyarlama adımları takip edilerek Sınıf İçi Değerlendirme Yaklaşımları Envanteri Türkçe'ye uyarlanmıştır.

2.5. Yapısal Eşitlik Modellemesi (YEM)

Yapısal eşitlik modellemesi içerisinde doğrulayıcı faktör analizi ve yol analizi gibi çok değişkenli istatistikleri barındıran, gözlenen ve gözlenmeyen değişkenlerin arasındaki nedensel ilişkileri inceleyeme yarayan güçlü bir istatistiksel tekniktir (Şen, 2020). Alanyazınında kovaryans yapı analizi, kovaryans yapı modeli ve kovaryans yapıların analizi olarak da isimlendirilmektedir (Kline, 2016).

Yapısal eşitlik modellemesi analizinde gözlenen ve gözlenmeyen değişkenler arasındaki ilişkileri en iyi tanımlayan modelin tahmin edilmesi ve doğrulanması amaçlanır. YEM'in geleneksel istatistiksel yöntemlere göre çeşitli üstünlükleri bulunmaktadır (Sümer, 2000). Birden fazla ölçüm yapılarak örtük yapılar arasındaki ilişkilerin test edilebilmesi, kurulan modelin kalitesi hakkında bilgi vermesi ve ölçme hatalarını dikkate alması örnek olarak verilebilir (Sümer 2000; Şen, 2020). YEM analizinde kullanılan değişkenlerin ve değişkenler arasındaki ilişkilerin gösteriminde kullanılan semboller Tablo 6'da verilmiştir.

Tablo 6. YEM'de kullanılan şekil ve semboller (Çokluk vd., 2021, s. 260)

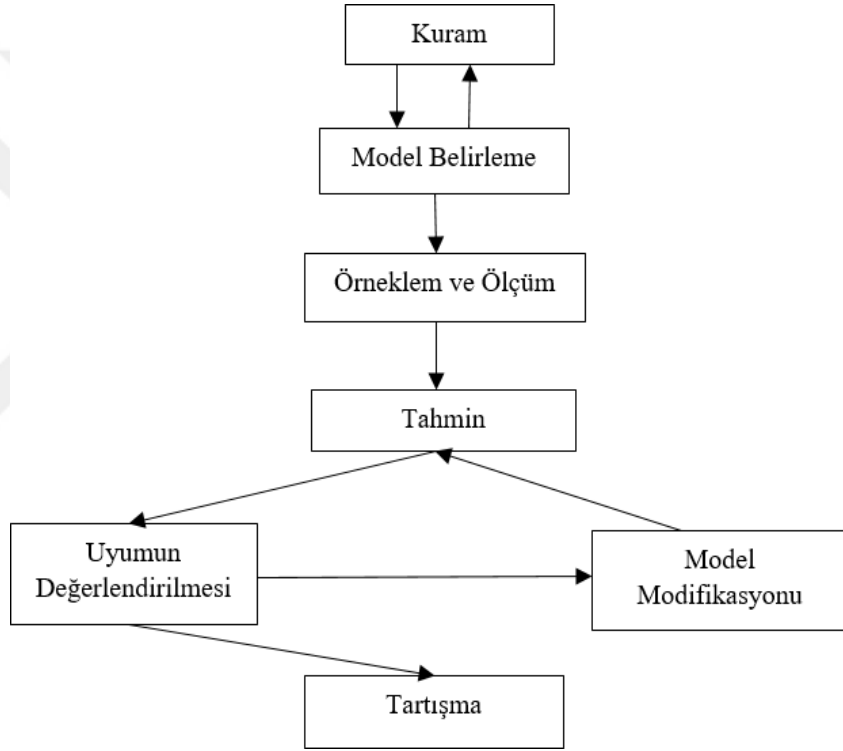
Gözlenen Değişken (Observed)	
Gizil Değişken (Latent)	
Doğrudan Etki (Direct Effect)	
Karşılıklı Etki (Reciprocal Effect)	
Korelasyon (Correlation) ya da Kovaryans (Covariance)	
Açıklanamayan Varyans (Disturbance)	<i>D</i>
Ölçme Hatası (Measurement Error)	<i>E</i>

Tam bir yapısal eşitlik modellemesi iki bölümden oluşur: *Ölçme modeli* ve *Yapısal model* (Şen, 2020). Ölçme modelinde gözlenen ve gizil değişkenler doğrulayıcı faktör analiziyle birbiriyle ilişkilendirilir, yapısal modelde ise örtük değişkenler arasındaki bağlantılar eş zamanlı olarak kurulur (Çokluk vd., 2021). Şen (2020) ölçme modelini YEM'in faktör analitik kısmı ve yapısal modeli ise YEM'in regresyon kısmı olarak açıklamıştır. YEM analizinin beş adımda gerçekleştirilmesi önerilmektedir (Bollen ve Long, 1993). Bu adımlar:

1. Modelin betimlenmesi
2. Modelin tanımlanması
3. Hesaplama
4. Uyumun test edilmesi
5. Yeniden betimleme'dir.

Araştırma modelin betimlenmesi adımı araştırma sorularına uygun olacak şekilde gözlenen ve gizli değişkenler arasındaki ilişkileri gösteren başlangıç modelini oluşturulur. Oluşturulan başlangıç modelinin tanımlanabilirliği ise YEM'in ikinci adımıdır. Bu adımda parametrelerin yordama güçleri doğrultusunda model tanımlanır. Modelin tanımlanmasının ardında ise veriler toplanır. Bir sonraki adımda ise çeşitli yazılımlar (AMOS, LISREL, Mplus vb.) kullanılarak parametre kestirimleri gerçekleştirilir. Tanımlanan verinin türüne göre maksimum olabilirlik (ML),

ağırlıklandırılmış en küçük kareler (WLS), ağırlıklandırılmamış en küçük kareler (ULS) ve genelleştirilmiş en küçük kareler (GLS) gibi parametre kestirim yöntemleri kullanılabilir (Şen, 2020). Parametre kestiriminin ardından ise model uyumu değerlendirilir. Model uyumunun değerlendirilmesinde ki-kare iyilik uyumu, GFI, AGFI, RMSEA, RMR, CFI, NFI, TLI(NNFI), PGFI uyum indeksleri kullanılır. Son adımda ise modifikasyon işlemleri yapılarak model yeniden betimlenir (Çokluk vd., 2021; Şen, 2020). Çokluk vd. (2021, s.256) tarafından oluşturulan YEM adımları ise Şekil 7’de verilmiştir.



Şekil 7. YEM için geleneksel yaklaşım (Çokluk vd., 2021, s.256)

2.5.1. Uyum İndeksleri

Model uyumu, gözlenen değişkenlere ait kovaryans matrisi ile tanımlanan modelde elde edilen kovaryans matrisinin ne derecede uyduğuyla ilgilidir (Sümer, 2000). Tam bir uyuma ise mükemmel uyum anlamına gelir. YEM analizinde model uyumunun değerlendirmesinde çeşitli uyum indeksleri kullanılmaktadır. Kline (2016) uyum indekslerini dört başlık altında sınıflandırmıştır, fakat RMSEA gibi bazı indekslerin aynı anda farklı sınıflarda da yer alabileceğini belirtmiştir. Şen (2020) ise

bu sınıflandırmaları *Mutlaka Uyum*, *Artırımlı Uyum*, *Tutumluluk Düzeltmesi* ve *Yordayıcı Uyum* olarak dilimize kazandırmıştır. Bu sınıflandırmaların temel özellikleri sırasıyla verilmiştir (Kline, 2016; Şen, 2020).

Mutlak Uyum (Absolute fit): Gözlenen verilerin model tarafından ne derecede iyi üretildiğine bakılır. Kötülük uyum ölçütleridir. İndeks bire yaklaşıkça uyum kötüleşir (Chen, 2007).

Artırımlı Uyum: Temel model (baseline) ile tahmini model arasındaki uyum ki-kare testiyle karşılaştırılır. İyilik uyum ölçütleridir. İndeks bire yaklaşıkça uyum mükemmelleşir (Chen, 2007).

Tutumluluk Düzeltmesi (Parsimony Correction): Tutumluluk ifadesiyle modelin tahmin ettiği parametre sayısının azlığı, modelin sadeliği kastedilir. Karmaşık modellemeler için ceza uygulanır. Mutlak ve artımlı uyum ölçütlerinin düzeltilmiş halleridir.

Yordayıcı Uyum (Predictive Fit Indexes): Modelin, orijinal örneklemin yer aldığı evrenden eşdeğer büyüklükte rastlantısal olarak seçilmiş örneklere uygunluğu test edilir. Dolayısıyla model, evren tabanlı olarak nitelendirilir. Uyum indekslerine ilişkin sınıflandırmalar Tablo 7’de verilmiştir.

Tablo 7. *Uyum indekslerinin sınıflandırılması* (Şen, 2020)

Mutlak Uyum (Absolute Fit)	Artırımlı Uyum (Incremental)	Tutumluluk Düzeltmesi (Parsimony Correction)	Yordayıcı Uyum (Predictive)
Ki Kare (χ^2) testi	NFI (Normed fit index)	PGFI (GFI’nin düzeltilmiş versiyonu)	AIC (Akaike’s Information Criteria)
χ^2 /sd oranı	CFI (comparative fit index)	PNFI (NFI’nin düzeltilmiş versiyonu)	BIC (Bayes Information Criteria)
AIC (Akaike’s Information Criteria)	TLI (Tucker-Lewis index)	PNFI2 (IFI’nin düzeltilmiş versiyonu)	SABIC (Sample size Adjusted BIC)
BIC (Bayes Information Criteria)	RNI (Relative non-centrality index)	PCFI (CFI’nin düzeltilmiş versiyonu)	CAIC (Consistent AIC)
ECVI (Expected Cross-validation Index)	IFI (Incremental Fit Index)	PR (Parsonomy ratio - tutumluluk oranı)	
RMSEA (Root mean square error approximation)			
RMR (Root mean square residual)			
SRMR (Standardized root mean square residual)			

Tablo 7 (Devamı)

GFI (Goodness of fit index)
AGFI (Adjusted Goodness of fit index)

Araştırmalarda hangi uyum ölçütlerinin kullanılması gerektiği konusu ise görecelidir. Temel prensip araştırmanın amacına en uygun indeksleri raporlamak olmalıdır. LISREL, AMOS gibi paket programlar birçok uyum indeksini çıktı olarak sunmaktadır. Bu çıktı sonuçlarına göre mükemmel uyum gösteren indekslerin raporlanıp, diğerlerinin göz ardı edilmesi hatalı olacaktır (Şen, 2020). Her bir uyum indeksinin kendisine özgü özellikleri olduğu unutulmamalıdır (Iacobucci, 2010). Bu doğrultuda Hoyle ve Panter (1995) ki-kare değerinin yanında mutlak ve artırılmış indekslerinin de raporlanması önermiştir. Kline (2016) ise ki-kare, RMSEA, CFI ve SRMR indekslerinin mutlaka raporlanması gerektiğini ifade etmiştir. Ki-kare raporlanırken sd ve p değerlerinin, RMSEA raporlanırken de güven aralıklarının belirtilmesi önerilmiştir (Şen, 2020). Araştırmacıların bu önerileri doğrultusunda öncelikle Ki-kare, RMSEA, SRMR, CFI ve TLI(NNFI) uyum indekslerine ait bilgiler sırasıyla verilmiştir.

Ki-kare (χ^2) Testi

Beklenen değerler ile gözlenen değerlerin farkına dayanır ve uyum anlamlıysa H_0 reddedilir ($p > .05$) (Erkuş, 2019). H_0 “model popülasyon kovaryans matrisine uymaktadır” anlamı taşır (Şen, 2020, s. 29). Bentler (1995) ise ki-kare testini iki kovaryans arasındaki uyum değeri ile örneklem büyüklüğünün bir eksiğinin çarpımı olarak açıklamıştır (akt. Çokluk vd., 2021). Bu açıklamadan ki-kare testinin örneklem ve korelasyon büyüklüğüne karşı duyarlı olduğu anlaşılmaktadır. Büyük örneklemelerde χ^2 değerinin arttığı ve H_0 hipotezinin anlamlı çıktığı belirtilmektedir (Erkuş, 2019; Şen, 2020). Bu nedenle χ^2 testi kötülük-uyumu testi (badness-of-fit) olarak da isimlendirilmiştir (Çokluk vd., 2021). Ayrıca χ^2 testinin kullanılacağı verinin çok değişkenli normallik varsayımını karşılaması da gerekir (West vd., 2012). χ^2 testinin bu dezavantajlı durumları göz önünde bulundurularak χ^2/sd oranının model uyumu için kullanılması önerilmiş, üç ve altındaki oranlar için iyi uyum, üç ile beş arasındaki oranlar için ise modelin veriye yeterli uyum gösterdiği ifade edilmiştir (Çokluk vd., 2021). Şen (2020) ise χ^2/sd oranlarının hangilerinin iyi uyumu

hangilerinin ise kötü uyumu gösterdiğine dair alanyazınında görüş birliği olmadığını ifade etmiştir.

Yaklaşık Hataların Ortalama Karekökü (Root Mean Square Error of Approximation (RMSEA))

Örnekleme ile model kovaryansları arasındaki farkın, dolayısıyla hatanın, derecesine göre uyum değerinin hesaplanmasıdır. Bu indeks 0 ile 1 arasında değer alır. Hatanın 0'a yaklaşması ise mükemmel uyuma işaret eder (Çokluk vd., 2021). RMSEA örnekleme sayısının yeterli olmadığı durumlarda ise doğru sonuç vermemektedir (Chen, 2007).

Standardize Edilmiş Artık Ortalamaların Karekökü (Standardized Root Mean Square Residuals, SRMR)

Evren ve model kovaryans matrisleri arasındaki artık kovaryans ortalamasıdır (Çokluk vd., 2021). Kötülük uyum ölçütüdür, değerin sıfıra yaklaşması uyumu mükemmelleştirir.

Karşılaştırmalı Uyum İndeksi (Comparative Fit Index, CFI)

Model uyumu, bağımsızlık (independence) ya da yokluk (null) olarak adlandırılan temel bir modele ait kovaryans matrisi ile YEM de elde edilen kovaryans matrisinin karşılaştırılması ile elde edilir (Çokluk vd., 2021). İyi uyum ölçütüdür, indeksin bire yaklaşması uyumu mükemmelleştirir.

Tucker-Lewis İndeksi (Tucker-Lewis Index, TLI)

Model uyum değerini modelin karmaşıklığı dikkate alınarak hesaplar. Bu hesaplamada modele ilişkin serbestlik değeri kullanılır (Sümer, 2000). İndeksin bire yaklaşması mükemmel uyumu işaret eder. AMOS programı TLI uyum indeksini NNFI şeklinde raporlamaktadır.

Uyum ölçütlerine ait özet bilgiler Tablo 8'de verilmiştir.

Tablo 8. *Uyum ölçütleri* (Çokluk vd., 2021; Hu & Bentler, 1999; Şen, 2020)

Uyum İndeksi	Mükemmel Uyum Ölçütü	Kabul Edilebilir Uyum Ölçütü
Ki-kare (χ^2)($p > .05$ olmalı)	-	-
CFI	$0,95 \leq CFI \leq 1$	$0,90 \leq CFI < 0,95$
RMSEA	$0 \leq RMSEA \leq 0,05$	$0,05 \leq RMSEA \leq 0,08$
SRMR	$0 \leq SRMR \leq 0,05$	$0,05 \leq SRMR \leq 0,08$
TLI (NNFI)	$0,95 \leq TLI \leq 1$	$0,90 \leq TLI \leq 0,95$

2.5.2. Varsayımlar

Çok değişkenli istatistik analizlerinde kullanılacak veri setinin bazı temel varsayımları karşılaması gerekir. Kayıp değerler, tekli/çoklu uç değerler, tekli/çoklu normallik, çoklu bağıntı ve eşvaryanslılık incelenmesi gereken varsayımlardır (Çokluk vd., 2021; Şen, 2020; Tabachnick & Fidell, 2007). Dolayısıyla YEM analizi öncesinde bu varsayımlar incelenmeli ve veri seti üzerinde gerekli değişiklikler yapılmalıdır.

Kayıp Değerler

Kullanılan örneklerde kayıp verilerin seçkisiz olarak sıralanması ve belirli bir örüntü içerisinde olmaması istenilen durumdur. Bu nedenle analize başlamadan önce kayıp verilere ilişkin örüntüler değerlendirilmelidir. Çünkü örüntülü kayıp veriler az sayıda bile olsa sonuçların genellenebilirliğini etkilemektedir (Tabachnick & Fidell, 2007). Kayıp verilerle mücadelede farklı yöntemler kullanılmaktadır. Örneğin SPSS gibi paket programlar aracılığıyla kayıp veri içeren değişkenleri veya cevaplayıcıları silme, kestirim yapma ve yaklaşık değer atama yapılabilmektedir. Bu yöntemlerden hangisinin kullanılması gerektiği ise örneklemin büyüklüğüne ve analiz edilen değişkenlerin özelliklerine göre farklılaşmaktadır. Silme işleminin kullanılması ciddi veri kayıplarına sebep olabileceği gibi, kestirim ve değer atama işlemleri de sadece nicel değişkenlerle sınırlıdır (Çokluk vd., 2021). Ayrıca örneklem büyüklüklerine göre ne kadar kayıp verinin göz ardı edilebileceğine dair ise kesin bir ölçüt bulunmamaktadır (Tabachnick & Fidell, 2007).

Uç Değerler

Bir örneklemden elde edilen puanlardan aşırı derecede farklılaşan puanlara uç değer denir (Kline, 2016). Bir veri setinde uçdeğer görülmesinin birçok sebebi bulunmaktadır (Tabachnick & Fidell, 2007). Bu sebepler sırasıyla;

1. Verilerin hatalı girilmesi,
2. Kayıp verilerin kodlanması sırasında yapılan hatalar ve kayıp veri atamaların program tarafından uç değer olarak algılanması,
3. Cevaplayıcının veri toplanması amaçlanan örneklemin bir üyesi olmaması,
4. Cevaplayıcının örneklemin kalan kısmından daha aşırı değerlere sahip olması ve normal dağılıma uymamasıdır.

Tek bir deęişkenle ilgili aşırı deęerler tek deęişkenli (yönlü) uç deęer olarak isimlendirilir. İki ya da daha fazla deęişkenin aşırı deęerlerine ise çok deęişkenli (yönlü) uçdeęerler denir (Çokluk vd., 2021; Kline, 2016). Uçdeęerler istatistiksel analizlerin doğru sonuçlara ulaşmasına engel olur. Birçok istatistiksel teknik (ANOVA, T Testi vb.) ortalama farklılıklarına dayanmaktadır. Uç deęerlerin varlığı ise ortalama deęerleri bozacak ve belki de yapılacak analizlerin anlamlı çıkmasına sebep olacaktır (Mertler & Vannatta, 2005 akt. Çokluk vd., 2021). Bu nedenle uç deęerler istatistiksel analizlere başlanmadan önce belirlenmeli ve incelenmelidir.

Tek deęişkenli uç deęerler verilerin standart Z puanlarına dönüştürülmesi yolu ile incelenebilir. +3 ile -3 standart puan dışında kalan veriler uçdeęer olarak kabul edilmektedir (Kline, 2016). Benzer şekilde histogram ve kutu grafiklerinin incelenmesi yoluyla da tek deęişkenli uç deęerler belirlenebilir. Normal dağılımın dışında kalan deęerler uçdeęer olarak kabul edilir (Tabachnick & Fidell, 2007). Çok deęişkenli uçdeęerlerin incelenmesinde ise Mahalonobis uzaklığı yöntemi kullanılmaktadır. Bu yöntemde cevaplayıcının dięer cevaplayıcıların merkezinden uzaklığının $p < ,001$ düzeyinde anlamlılığına bakılır. Anlamlı çıkan deęerler ise çok deęişkenli uç deęer olarak kabul edilir (Çokluk vd., 2021).

Uçdeęerlerin belirlenmesinden sonra ise uçdeęerin olası nedenleri üzerinden deęerlendirme yapılır. Veri girişi hatası kolaylıkla telafi edilebilirken, dięer olası nedenlerin düzeltilmesi ise zordur (Çokluk vd., 2021). Eęer uçdeęer cevaplayıcının örneklemin üyesi olmamasından kaynaklanıyorsa, Kline (2016) cevaplayıcının örneklem dışı bırakılmasını önermiştir. Ayrıca Kline (2016) uçdeęerlerle ilişkili deęişkene matematiksel dönüşümler yapılabileceğini veya uçdeęerin ± 3 standart sapma içerisinde yer alan en büyük ya da küçük deęere dönüştürülebileceğini belirtmiştir.

Normallik

Tek deęişkenli normallik grafikler ya da betimsel istatistikler incelenerek belirlenebilir. Histogram grafiklerinin normal eğrisine uygunluęuna ya da Q-Q saçılma grafiğinde gözlenen ve beklenen deęerlerin köşegenler odağında çizilen bir doğruya yakın dağılımda olmasına bakılabilir. Betimsel istatistiklere bakılarak çarpıklık ve basıklık deęerlerinin -1 ile +1 aralığında olması ya da Kolmogorov-Smirnov testi sonucuna göre normal dağılıma da karar verilebilir (Çokluk vd., 2021).

Çok değişkenli normallik ise gözlenen verilerin tüm değişkenlerin kombinasyonu için normal dağılım göstermesidir (Çokluk vd., 2021). Çok değişkenlik için (Kline, 2016):

1. Tek değişkenli normallik sağlanmalıdır.
2. İki değişkenli normallik, yani değişken çiftlerinin tüm kombinasyonları normal dağılıma sahip olmalıdır.
3. Tüm ikili değişken kombinasyonlarına ait saçılım grafikleri eşvaryanslı artıklarla doğrusal olmalıdır. Değişken çiftlerine ait saçılma grafiği elips şeklinde olmalıdır (Çokluk vd., 2021).

Normalliğin sağlanamadığı durumlarda ise veri dönüştürme yollarına başvurularak normallik varsayımı karşılanmalıdır (Çokluk vd., 2021). Çok değişkenli normalliğin test edilmesine yönelik alanyazında birçok yöntemin geliştirildiği görülmektedir. Enerji testi, Henze-Zirkler testi, Mardia'nın basıklık/çarpıklık testi, Royston H testi bunlardan bazılarıdır (Uysal ve Kılıç, 2022). Çok değişkenli istatistik analizleri yapılmadan önce bu testler kullanılarak normallik varsayımı test edilmelidir.

Çoklu Bağıntı

Bağımsız değişkenler arasındaki güçlü ilişkiler ($r > .90$) çoklu bağıntı olarak isimlendirilmektedir (Çokluk vd., 2021). Bu durumda VIF (varyans artış faktörü), CI (durum indeksi), tolerans değeri gibi katsayılar incelenerek çoklu bağıntıya sebep olan değişkenlerden bir ya da birden fazlası modelden çıkarılmalıdır (Tabachnick & Fidell, 2007). VIF değerinin 10 ve üzerinde olması; tolerans değerinin ise ,10 dan küçük olması bağımsız değişkenler arasında çoklu bağıntı problemine işaret eder (Gujarati, 1995; Webster, 1992 akt. Çokluk vd., 2021). Durum indeksinin ise 10 ile 30 arasında olması orta düzeyde, 30'dan büyük olması ise bağımsız değişkenler arasında yüksek düzeyde çoklu bağıntı problemini gösterir (Çokluk vd., 2021).

Eşvaryanslılık (Homojenlik)

Tek değişkenli istatistik analizlerinde varyansların homojenliği olarak da adlandırılır. Homojenlikle anlatılmak istenen sürekli değişkenlere ait puanlardaki değişimin her bir değişkende benzer şekilde gözlemlenmesidir (Çokluk vd., 2021). Tek değişkenli istatistik analizlerinde Levene testi (F puanı için $p > .05$ olmalı) ile çok değişkenli istatistiklerde Box M testi ($p < .05$ olmalı) ile homojenlik analizi gerçekleştirilir (Tabachnick & Fidell, 2007). Çok değişkenli istatistik analizlerinde ise normallik varsayımının karşılanmaması durumunda homojenlik sağlanamaz.

Değişkenler arasındaki normallik, değişkenlerin eşvaryanslı olmasını gerektirir (Tabachnick & Fidell, 2007). Stevens (1992) Box M testinin sonucunun çok değişkenli normalliğe fazlasıyla duyarlı olduğunu belirtmiş, normalliğin sağlanmadığı durumlarda Box M testi sonuçlarının reddedilmesinin hatalı olabileceğini ifade etmiştir (akt. Çokluk vd., 2021). Tabachnick ve Fidell (2007) ise tek ve çok değişkenli istatistiklerde homojenliğin kritik bir varsayım olmadığını ifade etmiştir.

2.6. Ölçme Değişmezliği

Ölçek geliştirme ve uyarlama çalışmaları ile elde edilen ölçme araçları bireyler ya da gruplar arasındaki farklılıkları belirlemek için kullanılır. Bireyler ya da gruplar arasındaki olası farklılıkların ölçme aracının ölçtüğü özellikten kaynaklandığı düşünülür. Dolayısıyla kullanılan ölçme aracının geçerli sonuçlar verdiği kabul edilir. Fakat elde edilen farklılıklar; ölçülen özellik dışında bireye ait farklı özelliklerin etkisinden veya ölçme aracının kendisinden de kaynaklanıyor olabilir (Uzun ve Öğretmen, 2010; Yandı vd., 2017). Bu nedenle kullanılan ölçme araçlarına ilişkin geçerlik çalışmaları yapılır ve çeşitli kanıtlar sunulur. Ölçme değişmezliği de yapılan bu çalışmalardan birisidir.

Byrne ve Watkins'e (2003) göre ölçme değişmezliği, ölçme aracında yer alan maddelerin tüm gruplarda aynı şekilde anlaşılması ve yorumlanmasıdır. Ölçme değişmezliği bir özelliğin farklı koşullarda ölçülmesinden elde edilen puanların aynı anlama gelip gelmediğiyle ilgilidir (Meade & Lautenschlager, 2004). Testin uygulanma zamanı, test uygulama yöntemleri ya da testin uygulandığı gruplar bu koşullara örnek olarak verilebilir. Ölçümler sonucunda elde edilen bireyler/gruplar arasındaki farklılıklar bu koşulların etkisi olmadan elde ediliyorsa bu durumda ölçme değişmezliğinden söz edilebilir (Kline, 2016).

Ölçme değişmezliği bir ölçeğin farklı uygulamalarında aynı parametre değerlerine sahip olmasıdır (Şen, 2020). Örneğin tekrarlı ölçümlerin yapıldığı bir durumda birinci ölçümdeki bir gösterge değişkeni ile örtük değişken arasındaki ilişki, ikinci ölçümde de mevcut ve aynı ise ölçme değişmezliği sağlanmıştır. Bu durum gösterge değişkeninin her iki ölçümde de örtük değişkenden aynı regresyon ağırlığına ya da sabitine sahip olması şeklinde de açıklanabilir (Clark & Donnellan, 2021). Psikometrik eşdeğerliğin ya da diğer bir ifadeyle ölçme değişmezliğinin

sağlanabilmesi için ölçek maddelerine ait faktör yükleri, faktörler arası korelasyon ve hata varyansları farklı gruplarda aynı olmalıdır (Başusta, 2010). Bu durumda gruplar için yapı geçerliği kanıtları elde edilmiş olacaktır. Fakat bu kanıtlar her bir grup için beklenen faktör yapısıyla uyumlu olmalıdır (Başusta ve Gelbal, 2015).

Ölçme değişmezliği gruplar arası karşılaştırma yapabilmek için bir önkoşul olarak kabul edilmektedir (Horn & Mcardle, 1992). Eğer bu koşul sağlanamıyorsa gruplar arasında niceliksel karşılaştırmalar yapmak yerine niteliksel yorumların yapılması daha uygun bulunmaktadır (Başusta ve Gelbal, 2015).

Ölçme değişmezliği çalışmalarında kullanılan yöntemler gözlenen ve örtük değişkenlerin özelliklerine göre farklılaşmaktadır. Yapısal Eşitlik Modellemesi (YEM), Madde Tepki Kuramı(MTK) ve Örtük Sınıf Analizi'ni (ÖSA) temel alan yöntemler kullanılmaktadır. Değişkenlerin sürekli yapıda olduğu durumlarda YEM ve MTK temelli yöntemler tercih edilirken, değişkenlerin kategorik olduğu durumlar da ise ÖSA içerisinde yer alan yöntemler kullanılmaktadır (Yandı vd., 2017). Vanderbeg ve Lance'e (2000) göre ise ölçme değişmezliğinin çalışmalarının %80'ni YEM modelleri kullanılarak gerçekleştirilmiştir. Bu çalışmalarda ise genellikle Çok Gruplu Doğrulayıcı Faktör Analizi'nin (ÇGDFA) kullanıldığı görülmektedir.

Alanyazını incelendiğinde ölçme değişmezliği ile ilgili birçok çalışmanın yapıldığı, farklı gruplarda (cinsiyet, sınıf düzeyi, ülke, kültür, bölge, dil vb.) ölçme değişmezliğinin incelendiği görülmektedir. Bu çalışmada ise Sınıf İçi Değerlendirme Yaklaşımları Envanteri'nin cinsiyet gruplarında ölçme değişmezliği incelenmiştir.

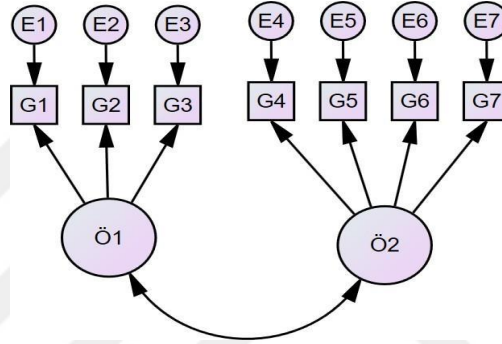
2.6.1. Doğrulayıcı Faktör Analizi (DFA)

Doğrulayıcı faktör analizi, gizil değişkenler ile gözlenen değişkenler arasındaki ilişkileri inceleyen ve ileri düzey araştırmalarda kullanılan gelişmiş bir tekniktir (Tabachnick & Fidell, 2007). DFA, gözlenen değişkenler (test maddeleri, test puanları vb.) ile gizil değişkenler arasındaki ilişkileri analiz eden yapısal eşitlik modellemesi çeşitlerinden birisidir (Brown & Moore, 2012).

DFA'da araştırmacı sahip olduğu teorik modelin uygunluğunu gözlenen değişkenler ile örtük değişkenler arasındaki ilişkileri inceleyerek test eder (Şen, 2020). Dolayısıyla DFA'da amaç bir yapının ortaya çıkarılması yerine hali hazırda mevcut olan yapının eldeki veriler kullanılarak doğrulanmaya çalışılmasıdır (Erkuş, 2019).

Gözlenen ve örtük değişkenler arasındaki ilişkiler ölçme hataları dikkate alınarak tek bir analizle incelenir.

DFA’da gözlenen ve gizil değişkenler arasında ilişkiler varyans ve kovaryans matrisleri ile test edilir. Değişkenler arasındaki ilişkinin derecesi faktör yükü (λ) ile ifade edilir ve ilişki yol (path) adı verilen tek yönlü ok ile gösterilir (Çokluk vd., 2021). Örtük değişkenler arasındaki ilişki ise çift yönlü ok ile gösterilir. DFA YEM’in ölçme modeli kısmını oluşturur (Şen, 2020). Şekil 8’de iki örtük değişken (Ö) ve yedi gizil değişken (G) arasında kurulan DFA modeli verilmiştir. Modelde hatalar ise E harfi ile ifade edilmiştir.



Şekil 8. DFA modeli (Çokluk vd., 2021, s. 279)

DFA’da gözlenen ve gizil değişkenler arasındaki ilişki doğrusaldır ve hata varyansı dikkate alınır. DFA denklemi ise (Erkuş, 2019, s. 41):

$$y = \lambda_y \eta + \varepsilon \text{ 'dir.}$$

y: gözlenen değişkenler

λ : faktör yükleri

η : gizil faktörler

ε : hata varyansı’dır.

2.6.2. Çok Gruplu Doğrulayıcı Faktör Analizi (ÇGDFA)

İstatistiksel işlemlerde kullanılan verilerin tek bir örneklem grubundan toplandığı düşünülerek analizler yapılması hatalı olabilir. Veriler kültür, cinsiyet, dil vb. değişkenlere göre farklılaşan alt gruplardan elde edilmiş olabilir. Bu durumda ölçümlerin veri toplanılan tüm gruplar için değişmezliği incelenmelidir (Şen, 2020). Böylece her bir grup için ayrı ayrı yapı geçerliği kanıtları elde edilmiş olunacaktır.

ÇGDFA ile iki ya da daha fazla grupta DFA analizleri eş zamanlı olarak gerçekleştirilir. Böylece ölçme aracının psikolojik yapısının farklı gruplarda aynı olup olmadığı değerlendirilir. Gruplar arasındaki ölçme aracından kaynaklanan olası yanlılıklar bu yöntemle belirlenebilmektedir (Önen, 2009).

ÇGDFA'da temel mantık kurulan modelde parametrelerin aşamalı (stepwise) olarak sınırlandırılmasıdır. Parametrelerin sınırlandırılması her bir modelde artış gösterir. Model uyumun değerlendirilmesi en az sınırlandırılmış modelle başlar ve daha çok sınırlandırılmış model ile daha az sınırlandırılmış model arasındaki uyum test edilerek devam eder (Önen, 2009). ÇGDFA kullanılarak ölçme değişmezliği dört aşamada incelenmektedir.

2.6.3. Ölçme Değişmezliği Aşamaları

Meredith'e (1993) göre ölçme değişmezliği dört aşamada incelenir. Her aşamada kurulan modele ait parametreler test edilir. Bu aşamalar sırasıyla verilmiştir.

1. Yapısal Değişmezlik (Configural Invariance)
2. Metrik Değişmezlik (Metric/Weak Factorial Invariance)
3. Ölçek Değişmezliği (Scalar/Strong Invariance)
4. Katı Değişmezlik (Strict Invariance)

Ölçme değişmezliği aşamalı, hiyerarşik bir yapıda incelenir. Her bir aşamada kurulan hipotezler model uyum katsayıları kullanılarak test edilir (Başusta ve Gelbal, 2015). Değişmezlik aşamaları hiyerarşik bir sırayla kurulur ve her bir aşamada kurulan model bir önceki aşamadaki modelle karşılaştırılır. Kurulan her bir yeni modelde parametrelere ek kısıtlamalar getirilerek, bir önceki model ile karşılaştırılır ve uyum katsayıları incelenir. Eğer yeni modelin veriye uyum katsayıları uygun düzeyde ise değişmezliğin sağlandığı kabul edilir ve bir sonraki aşamaya geçilir. Dolayısıyla metrik değişmezliğin sağlanmadığı bir modellemede skaler değişmezlik incelenmez (Uzun ve Öğretmen, 2010).

1. Yapısal Değişmezlik (Configural invariance)

Ölçme değişmezliğinin en temel aşamasıdır. Bu aşamada ölçme aracına ait maddelerin gruplar arasında aynı yapıyı ölçüp ölçmediği test edilir (Şen, 2020). Eğer aynı yapı ölçülüyorsa gruplar arası karşılaştırmalar yapılmamalı, elmalar ile

portakallar karşılaştırılmamalıdır (Clark & Donnellan, 2021; Erkuş, 2019). Temel model olması nedeniyle yapısal değişmezlik genel model uyumuna göre değerlendirilir (Şen, 2020).

2. Metrik Değişmezlik (Metric/Weak Factorial Invariance)

Ölçme değişmezliğinin bu aşamasında gruplar arasındaki faktör yüklerinin eşit olduğu hipotez test edilir (Önen, 2009). Ölçme aracının birimlerinin gruplar arasındaki eşitliğinin sınanması için kurulan modelde faktör yüklerinin eşit olduğu, fakat diğer parametrelerin serbestçe hesaplandığı kabul edilir (Erkuş, 2019). Metrik değişmezliğin sağlanması grupların maddeleri aynı şekilde anlayıp, yorumladıkları anlamına gelirken yapının aynı şekilde ölçüp ölçülmediği ise belirsizdir (Şen, 2020). Bu nedenle ölçek değişmezliği aşamasına geçilir.

3. Ölçek Değişmezliği (Scalar/Strong Invariance)

Bu aşamada faktör yüklerinin eşitliğinin yanı sıra gösterge sabitlerinin de eşit olduğu varsayımı test edilir (Şen, 2020). Diğer bir ifadeyle ölçümlerin orjinleri, regresyon sabitleri, gruplar arasında eşittir (Önen, 2009). Skalere değişmezliğin sağlanması gruplar arası karşılaştırma yapmak için şarttır. Fakat bu aşamada ölçme değişmezliğinin sağlanamaması ise olası bir yanlılık göstergesidir (Erkuş, 2019; Şen, 2020).

4. Katı Değişmezlik (Strict Invariance)

Katı değişmezlik aşamasında faktör yükleri ve regresyon sabitlerine ek olarak hata varyanslarının da gruplar arasında eşit olduğu kabul edilir. Katı değişmezliğin sağlanması zordur, modifikasyonlara başvurulabilir veya kısmi ölçme değişmezliği (partial measurement invariance) çalışmaları gerçekleştirilebilir (Erkuş, 2019). Ölçme değişmezliği aşamalarına ait özet bilgiler Tablo 9'da verilmiştir.

Tablo 9. Ölçme değişmezliği aşamaları (Erkuş, 2019; Önen, 2009; Şen, 2020)

Değişmezlik Türü	Hipotez	Formül
Yapısal Değişmezlik (Configural invariance)	Serbest ve sabit faktör yükleri örüntüsü açısından gruplar arası farklılık yoktur.	-
Metrik Değişmezlik (Metric/Weak Factorial invariance)	Gruplar arasında faktör yükleri (λ) eşittir.	$\lambda_k^g = \lambda_k^{g'}$
Ölçek Değişmezliği (Scalar/Strong invariance)	Gruplar arasında faktör yükleri (λ) ve gösterge sabitleri (τ) eşittir.	$\tau_k^g = \tau_k^{g'}$

Tablo 9 (Devamı)

Katı Değişmezlik (Strict invariance)	Gruplar arasında faktör yükleri (λ), gösterge sabitleri (τ) ve hata varyansları (ϵ) eşittir.	$\epsilon^g = \epsilon^{g'}$
--------------------------------------	--	------------------------------

Ölçme değişmezliği çalışmalarında farklı modeller arasındaki uyumun değerlendirilmesinde ki-kare fark testi kullanılır. Birbiri içine geçmiş ya da kümelenmiş (nested) olarak isimlendirilen bu modellerin karşılaştırılmasında ki-kare testinin anlamlı ($p > .05$) çıkması istenir ve böylece kısıtlanmış modelin değişmezliğin sağlandığı kabul edilir (Şen, 2020). Ki-kare testinin ise çok değişkenli normal dağılımdan sapma, örneklem büyüklüğü arttıkça testin manidar çıkma olasılığının artması, parametre sayısı artışından aşırı etkilenme gibi olumsuz özellikleri mevcuttur (Clark & Donnellan, 2021; Erkuş, 2019; Şen, 2020). Bu nedenle uyum istatistiklerin incelenmesinde farklı yöntemler önerilmiştir.

Ki-kare fark testi yerine modeller arası karşılaştırmada RMSEA (*Yaklaşık Hataların Ortalama Karekökü*), CFI (Karşılaştırmalı Uyum İndeksi) ve SRMR (*Standardize Edilmiş Artık Ortalamaların Karekökü*) en çok tercih edilen üç yöntemdir (Chen, 2007; Clark & Donnellan, 2021). Örneklem sayısının 300 ve üzerinde ve alt gruplarda eşit olduğunda iki model arasındaki $\Delta CFI \leq .01$, $\Delta RMSEA \leq .015$ ve $\Delta SRMR \leq .03$ karar ölçütleri ölçme değişmezliğinin sağlandığını gösterir (Chen, 2007). Chen (2007) RMSEA ve SRMR uyum indekslerinin küçük örneklerde ölçme değişmezliğine karşı duyarlı olduğunu, bu nedenle ölçme değişmezliğinin değerlendirilmesinde genel olarak CFI uyum indeksinin kullanıldığını belirtmiştir.

BÖLÜM III

YÖNTEM

Bu bölümde araştırma modeli, çalışma grubu, veri toplama aracı, verilerin toplanması ve analizi ele alınacaktır.

3.1. Araştırma Modeli

Bu araştırmada Sınıf İçi Değerlendirme Yaklaşımları Envanteri'nin Türkçe'ye uyarlanması ve envanterin cinsiyete göre ölçme değişmezliğinin incelenmesi amaçlanmıştır. Araştırmada ölçek uyarlama adımlarından oluşan bir süreç mevcuttur. Bu nedenle araştırma betimsel bir nitelik taşımaktadır. Betimsel araştırmalarda, araştırmada yer alan bireylere, gruplara ya da fiziksel ortama ilişkin (okul vb.) özellikler ayrıntılı ve eksiksiz bir biçimde tanımlanır (Büyüköztürk vd.,2022).

3.2. Çalışma Grubu

Ölçek geliştirme ve uyarlama çalışmalarında çalışma grubunun kimlerden oluşması ve büyüklüğünün ne kadar olması gerektiği konusu, ölçülen özelliğin yapısına ve ölçek/madde analizinde kullanılacak yöntemlere bağlıdır (Erkuş, 2021). Ölçülecek özelliğin tek veya çok boyutlu olması, bu özelliğin ölçülebilmesi için gerekli kişi sayısını doğrudan etkilemektedir. Tek boyutluluktan çok boyutluluğa gidildikçe, çalışma grubuna her bir boyutu temsil eden daha fazla kişi eklenir. Ölçek geliştirme ve uyarlama çalışmalarında çok değişkenli analizler de yapılacağından çalışma grubunun çok sayıda katılımcıdan oluşması istenilen bir durumdur (Erkuş, 2021). Bu araştırmanın çalışma grubu ise envanter maddelerinin özellikleri dikkate

alınarak ve yapılacak çok değişkenli analizler göz önünde tutularak 36 farklı öğretmenlik branşından 405 öğretmenin katılımıyla oluşturulmuştur. Araştırmanın çalışma grubunu 2022-2023 yılları arasında Milli Eğitim Bakanlığı'na bağlı okullarda öğretmenlik yapan 154'ü erkek (% 38,1) 251'i kadın (% 61,9) olmak üzere toplam 405 kişi oluşturmaktadır. Çalışma grubunda yer alan öğretmenlerin 126'sı 0-5 yıl; 114'ü 6-10 yıl; 67'si 11-15 yıl; 54'ü 16-20 yıl; 28'i 21-25 yıl; 12'si 26-30 yıl ve 4'ü ise 31-35 yıl arasında mesleki kıdeme sahiptir. Çalışma grubuna ait betimsel istatistikler Ek 2'de verilmiştir.

3.3. Veri Toplama Aracı

Araştırmada veri toplama aracı olarak Deluca vd. (2016b) tarafından geliştirilen (*Approaches to Classroom Assessment Inventory, ACAI*) ve Türkçe'ye Sınıf İçi Değerlendirme Yaklaşımları Envanteri ismiyle uyarlanan ölçme aracı kullanılmıştır. Deluca vd. (2016a) envanteri Sınıf İçi Değerlendirme Standartları'nı (Klinger vd., 2015) temel alarak ve altı farklı bölgede (Amerika Birleşik Devletleri, Avrupa, Avusturalya, Birleşik Krallık, Kanada ve Yeni Zelanda) tanımlanan ölçme değerlendirme standartlarını inceleyerek geliştirmiştir. Bu standartları *Değerlendirme Amaçları*, *Değerlendirme Süreçleri*, *Değerlendirme Adilliği* ve *Değerlendirme Teorisi* temaları altında gruplandırmışlar ve envanteri bu temalara uygun olarak yapılandırmışlardır. Bu temalara ve temaların alt boyutlara ilişkin bilgiler Tablo 10'da verilmiştir.

Tablo 10. *Değerlendirme temaları ve alt boyutlar* (Deluca vd., 2021, s.10)

Değerlendirme Temaları	Alt Boyutlar	Açıklamalar
Değerlendirme Amaçları	Öğrenmenin Değerlendirilmesi	Öğrenci öğrenmelerinin özetlenmesi ve notlandırılması amacıyla değerlendirilme yapılmasıdır.
	Öğrenme İçin Değerlendirme	Biçimlendirici değerlendirme yöntemlerinin kullanılarak eğitsel faaliyetlerin sonraki aşamalarına rehberlik edilmesidir. Öğretmen tarafından yönlendirilen öğrenci merkezli bir yaklaşımdır.
	Öğrenirken Değerlendirme	Öğrencilerin üstbilişsel yeteneklerini ve öğrenme becerilerini (öz değerlendirme, hedef belirleme, öğrenme planları oluşturma vb.) geliştirmek için onlara geribildirim verilmesini amaçlayan, temelde öğrenci-merkezli fakat öğretmenin de sürece dâhil olduğu yaklaşımdır.

Tablo 10. (Devamı)

Değerlendirme Süreçleri	Tasarım	Öğrenme hedefleriyle ilişkili olarak öğrencinin öğrenmesini ölçen güvenilir değerlendirmelerin ve ölçme araçlarının geliştirilmesidir.
	Kullanım/Puanlandırma	Değerlendirme senaryolarının yanıtlanmasında puanlandırma anahtarlarının düzenlenmesi ve kullanımına odaklanan alt boyuttur.
	İletişim	Öğrenciler ve velilerle iletişim yoluyla değerlendirme sonuçlarının ve geri bildirim yorumlanmasına odaklanan alt boyuttur.
Değerlendirme Adilliği	Standart Yaklaşım	Tüm öğrenciler için eşit değerlendirme süreçlerinin yürütülmesidir.
	Eşitlikçi Yaklaşım	Özel eğitime ihtiyaç duyan öğrenciler için (BEP'li vb.) değerlendirme süreçlerinin yapılandırılmasıdır.
	Farklılaştırılmış Yaklaşım	Her öğrencinin kendisine özgü öğrenme ihtiyaçlarının ve hedeflerinin göz önünde bulundurularak öğrenme fırsatlarının ve değerlendirmelerin bireyselleştirilmesidir.
Değerlendirme Teorisi	Tutarlı	Farklı öğretmenler ya da farklı zaman aralıklarında yapılan değerlendirmeler arasında tutarlılığın sağlanmaya çalışılmasıdır.
	İçeriksel	Ölçme ve değerlendirmenin amacına ve ölçülmesi hedeflenen kazanımlara uygun yapıldığından emin olunması, sonuçlarının analizinde geçerliliğin sağlanmasıdır.
	Dengeli	Ölçme ve değerlendirmenin amacına, ölçülmesi hedeflenen kazanımlara uygunluğuna ve ne derecede gerçekleştirilebildiğine odaklanılması, ölçme ve değerlendirmede tutarlılığın sağlanmaya çalışılmasıdır.

Temalar ve alt boyutlara ait bilgilerden sonra ise envanterin bölümlerine ilişkin bilgiler sırasıyla verilmiştir.

Bölüm 1. Sınıf İçi Değerlendirme Yaklaşımları

Envanterin bu bölümü birbirinden bağımsız dört senaryodan oluşmaktadır ve her bir senaryo 12 madde içermektedir. Senaryolar öğretmenlerin ikilemede kaldığı sonuç odaklı değerlendirme, not verme, farklılaştırılmış değerlendirme ve bütünleştirilmiş değerlendirme konularıyla ilgilidir.

Senaryodaki maddeler *Değerlendirme Amaçları*, *Değerlendirme Süreçleri*, *Değerlendirme Adilliği* ve *Değerlendirme Teorisi* temalarına göre kategorik olarak yapılandırılmıştır. Cevaplayıcıların bu maddelere verdikleri cevaplara göre değerlendirme yaklaşımları elde edilmektedir. Değerlendirme yaklaşımlarına ilişkin bilgiler Tablo 11'de verilmiştir.

Tablo 11. *Değerlendirme yaklaşımlarının temasal dağılımı* (Deluca vd., 2016b, s. 252)

Değerlendirme Teması	Değerlendirme Yaklaşımları		
	Yaklaşım A (Sonuç Odaklı)	Yaklaşım B (Biçimlendirici)	Yaklaşım C (Bireyselleştirilmiş)
Değerlendirme Amaçları	Öğrenmenin Değerlendirilmesi	Öğrenme İçin Değerlendirme	Öğrenirken Değerlendirme
Değerlendirme Süreçleri	Tasarım	Kullanım/Puanlandırma	İletişim
Değerlendirme Adilliği	Standart Yaklaşım	Eşitlikçi Yaklaşım	Farklılaştırılmış Yaklaşım
Değerlendirme Teorisi	Tutarlı	İçeriksel	Dengeli

Tablo 11'deki yaklaşımlar incelendiğinde sonuç odaklı yaklaşıma sahip öğretmenlerin öğretim sürecinde güvenilir değerlendirme tasarımlarını ve standartlaştırılmış değerlendirme uygulamalarının kullanarak tutarlı değerlendirmeler yapmaya odaklandıkları anlaşılmaktadır. Biçimlendirici yaklaşıma sahip öğretmenlerin ise öğretim sürecinde değerlendirmeyi öğrencilerin özel durumlarını gözeterik yaptıkları, değerlendirme sonuçlarını kullanarak öğrencilere geribildirim verdikleri ve böylece öğretim sürecini şekillendirdikleri görülmektedir. Bireyselleştirilmiş yaklaşıma sahip öğretmenlerin ise öğrencilerin üstbilişsel ve kendi kendine öğrenme becerilerini geliştirebilmek için onlara geri bildirim verdikleri, değerlendirmeyi öğrencilerin bireysel durumlarını dikkate alarak dengeli bir şekilde yaptıkları görülmektedir.

Envanterin geliştirme çalışmasında 404 Kanadalı öğretmenin senaryolara verdikleri cevaplar üzerinden değerlendirme profilleri oluşturulmuştur (Deluca vd., 2016b). Bölüm 2'deki senaryo maddelerinin kategorik yapıda olmasından dolayı, ortalamaya dayalı istatistiksel analizler yerine mod istatistikleri kullanılarak öğretmenlerin değerlendirme profilleri analiz edilmiştir. Pearson Ki-Kare analizi yapılarak her bir değerlendirme temasında en çok tekrar eden alt boyut cinsiyet, yaş ve çalışılan okul türüne değişkenlerine göre belirlenmiştir. Analiz sonuçlarına göre *Öğrenme İçin Değerlendirme*, *Tasarım ve İletişim*, *Farklılaştırılmış Yaklaşım ve Dengeli* yaklaşım en çok tekrar eden alt boyutlardır. Ki kare analiz sonuçlarına göre öğretmenlerin değerlendirme profilleri ile demografik bilgileri (cinsiyet, yaş, çalışılan okul türü) arasında önemli farklılıklar bulunmamıştır.

Bölüm 2. Sınıf İçi Değerlendirmede Güven Ölçeği

Bu bölümde öğretmenlerin sınıf içi değerlendirmede kendilerine duydukları güven beşli likert tipi (1=Henüz çok acemiyim, 5=Bu konuda uzmanım) 12 madde ile ölçülmektedir. Maddeler Sınıf İçi Değerlendirme Standartları'nın (Klinger ve ark., 2015) *Temeller, Kalite ve Kullanım* başlıklarına göre yapılandırılmıştır. Ölçek 12 madde ve iki alt boyuttan oluşmaktadır. Yazarlar tarafından birinci alt boyut *Değerlendirme Tasarımı, Uygulama ve Geribildirim (DTUG)* olarak isimlendirilmiştir. İkinci alt boyuta ise *İzleme, Analiz Etme ve Değerlendirme Sonuçlarını İletme (İADSI)* adı verilmiştir.

Envanterin geliştirilme çalışmasında alt boyutlar arasındaki korelasyon değeri $r=,71$ olarak bulunmuştur. Alt boyutlar için ise Cronbach α güvenilirlik katsayısı kestirilmiştir. Alt boyutlar için güvenilirlik değerleri sırasıyla $\alpha=,89$ ve $\alpha=,90$ 'dır.

Bölüm 3. Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri ve Tercihleri Ölçeği

Envanterin bu bölümü *Değerlendirme Eğitimi ve Öğretmenlerin Değerlendirme Konusunda Desteklenmesi* temasına uygun olarak iki alt bölüm şeklinde yapılandırılmıştır.

Bölüm 3A. Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri

Ölçek beşli likert tipi 10 maddeden oluşmaktadır. Öğretmenlerin her bir maddeyi yapma olasılıklarını birden beşe kadar derecelendirmeleri istenmektedir. Ölçek *Değerlendirme Uygulamalarının Bütünleştirilmesi/İletimi (DUBİ)* ve *Mevcut Değerlendirme Teorisi, İlkeleri ve Uygulamalarla Uyum (MDTİU)* ismi verilen iki alt boyuttan oluşmaktadır. Alt boyutlar arasındaki korelasyon değeri $r=,50$ ve Cronbach α güvenilirlik katsayısı ise sırasıyla $,83$ ve $,85$ 'tir.

Bölüm 3B. Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri

Ölçek beşli likert tipi (1=Tercih etmem, 5=Çok Tercih ederim) 11 maddeden oluşmaktadır. Ölçek *Çevrimiçi Öğrenme, Yüz-Yüze (Grup) Öğrenme ve Bire-Bir Öğrenme* adı verilen üç alt boyuttan oluşmaktadır.

Alt boyutlar arasındaki korelasyon değerleri $r_{\text{çevrimiçi-yüzyüze}}=-,10$; $r_{\text{çevrimiçi-birebir}}=,04$; $r_{\text{yüzyüze-birebir}}=,42$ 'dir. Alt boyutlar için hesaplanan Cronbach α güvenilirlik katsayıları ise $\alpha_{\text{çevrimiçi}}=,86$; $\alpha_{\text{yüzyüze}}=,74$ ve $\alpha_{\text{birebir}}=,92$ 'dir.

3.3.1. Envanterin Uyarlanma Adımları

Bu arařtırmada Hambleton ve Patsula (1999) ve ITC (2017) tarafından hazırlanan ölçek uyarlanma adımları takip edilerek Sınıf İçi Deęerlendirme Yaklařımları Envanteri'nin Türkçe'ye uyarlanması yapılmıřtır. Uyarlamada izlenen adımlar sırasıyla verilmiřtir.

1. Ölçme aracının uyarlanmasının yapılabilmesi için gerekli izinler alınmıřtır (Ek 1).
2. Yapısal eřitlięin saęlanabilmesi amacıyla alan uzmanlarından görüş alınmıřtır.
3. Ölçme aracının çevirisinde her iki dile hakim iki çevirmen ve iki ölçme deęerlendirme uzmanından faydalanılmıřtır.
4. İleriye doęru çeviri yöntemi kullanılmıřtır. Geriye doęru çeviri yapılmamıřtır.
5. Çevirisi yapılan ölçme aracı için uzman görüşleri alınmıř ve gerekli düzenlemeler yapılmıřtır.
6. Dil eřdeęerlięi için ölçme aracının İngilizce ve Türkçe formları 15 İngilizce öęretmenine uygulanmıř ve alınan puanlar arasındaki korelasyon deęerleri hesaplanmıřtır.
7. Envanterin deneme uygulaması 26 öęretmenin katılımıyla gerçekleştirilmiřtir. Katılımcılardan elde edilen dönütlere göre ölçme aracında deęiřiklikler yapılmıř, envanterin ortalama cevaplandırılma süresi hesaplanmıřtır.
8. Envanter 405 öęretmene uygulanmıřtır.
9. Envanter ile ilgili farklı güvenilirlik belirleme yöntemleri kullanılmıřtır.
10. DFA yapılarak ölçeklerin orjinal yapısı incelenmiřtir.
11. Envanter için kullanıcı rehberi hazırlanmıřtır (Ek 11).

Alanyazın incelendięinde deęerlendirme okuryazarlıęının cinsiyete göre anlamlı olarak farklılařtıęı (Akpolat & Genç, 2022; Atasoy, 2022; Tünkler, 2019) veya farklılařmadıęı (Ergül, 2019) arařtırmalar mevcuttur. Bu arařtırmalarda kullanılan veri toplama aracına iliřkin (Bütüner vd., 2010; *Ölçme Deęerlendirme Okuryazarlıęı Envanteri*) cinsiyete göre ölçme deęiřmezlięi analizleri ise bulunmamaktadır. Ölçme deęiřmezlięinin gruplar arası karşılařtırma yapabilmek için bir önkořul olmasından

dolayı (Horn & Mcardle, 1992), bu arařtırmalarda elde edilen bulguların geerliđi üzerinde řüphe oluřmaktadır. Dolayısıyla cinsiyet gruplarında deđerlendirme okuryazarlıđının dızeylelerinin karřılařtırılmasında, olme deđiřmezliđi kanıtları elde edilmiř olme araları kullanılmalıdır. Bu nedenle Sınıf İi Deđerlendirme Yaklařımları Envanteri'nin Trke'ye uyarlanma srecinden sonra, envanterde yer alan oleklerin cinsiyete gre olme deđiřmezliđi incelenmiřtir. Olme deđiřmezliđi GDFA tekniđiyle gerekleřtirilmiřtir. Biimsel, metrik, olek ve katı deđiřmezlik ařamaları sırasıyla takip edilmiřtir.

3.4. Verilerin Toplanması

Uyarlama iřlemine ardından veriler katılımcıların gnlllk esasına gre Google Formlar uygulaması kullanılarak internet zerinden ve katılımcılara kâđıt kalem uygulaması yapılarak toplanmıřtır.

3.5. Verilerin Analizi

Arařtırmada alıřma grubuna ait betimsel istatistikler SPSS 23 programı kullanılarak elde edilmiřtir. Senaryo maddelerin kategorik olmasından dolayı Pearson's Ki-Kare testi kullanılarak cinsiyet ve mesleki kıdem deđiřkenleri ile katılımcıların deđerlendirme yaklařımları arasındaki iliřki incelenmiřtir.

Dil eřdeđerliđi iin beři Olme ve Deđerlendirme alanından biri ise Eđitim Bilimleri alanından olmak zere altı đretim yesinden uzman grřleri alınmıřtır. Uzman grřleri arasında ise uyum indeksileri hesaplanmıřtır. Dil eřdeđerliđi iin 15 İngilizce đretmenine envanterin Trke ve İngilizce formları iki hafta arayla uygulanmıř ve formlardan alınan puanlar arasındaki iliřki Spearman Brown Sıra Farkları korelasyonu ile incelenmiřtir.

3.5.1. Varsayımların Test Edilmesi

Dil eřdeđerliđinin sađlanmasının ardından ise DFA ve olme deđiřmezliđi analizleri iin veri setinin gerekli varsayımları karřılayıp karřılamadıđı test edilmiřtir. ok deđiřkenli istatistiksel analizler yapılmadan nce kullanılacak veri setine iliřkin kayıp deđerler, u deđerler, normallik, oklu bađıntı ve eřvaryanslılık incelenmesi

gereken temel varsayımlardır (Çokluk vd., 2021). Tabachnick ve Fidell (2007) ise eşvaryanslılık varsayımının çok değişkenli istatistiksel işlemlerde kritik bir varsayım olmadığını, çok değişkenli normalliğin olmadığı veri setlerinde eşvaryanslılığın sağlanamayacağını ifade etmiştir. Bu nedenle eşvaryanslılık varsayımı test edilmemiştir.

Kayıp Değerler: Araştırmada çeşitli sebeplerden ötürü kayıp veriler oluşabilir. Cevaplanmayan maddeler, test kitapçığındaki hatalı basımlar, internet tabanlı uygulamalarda karşılaşılan yazılımsal sıkıntılar başlıca kayıp değer nedenleridir. Araştırmalarda kayıp değerlerin toplam veri setinin yüzde beşinden daha az olması ve belirli bir örüntü göstermemesi istenir (Kline, 2016). Bu araştırmada ise kayıp değerlerin toplam verilerin %1,84'ünü oluşturmaktadır. Kayıp veri analizinde ise 330, 336 ve 348 numaralı katılımcılara ait kayıp verilerde örüntü olduğu görülmüş, bu nedenle bu katılımcılara ait veriler analiz dışı bırakılmıştır. Diğer kayıp verilere ise SPSS 23 programı kullanarak ortalama değer atanmıştır.

Uç Değerler: Bir örnekleme ait dağılımının uç noktalarında yer alan, ortalamadan aşırı derecede farklılaşan değerlere uç değer denir (Çokluk vd., 2021). Uç değerler tek değişkenli ve çok değişkenli uç değer olmak üzere ikiye ayrılır. İstatistiksel analizlerin anlamlı bir şekilde gerçekleştirilebilmesi için analiz öncesinde uç değerler dikkatli bir şekilde gözden geçirilmelidir. SPSS programı kullanılarak tek değişkenli uç değerler için kutu grafiği, çok değişkenli uç değerler için Mahalonobis uzaklıkları incelenebilir (Tabachnick & Fidell, 2007). Envanterin ikinci ve üçüncü bölümlerinde yer alan ölçekler için uç değer analizleri SPSS 23 programı ile gerçekleştirilmiştir. Tek değişkenli uç değerlerin tespiti için kutu grafikleri incelenmiş, envanterin ikinci bölümünde 284 nolu katılımcının tek değişkenli uç değer olduğu ve envanterin üçüncü bölümünde ise tek değişkenli uç değerlerin olmadığı belirlenmiştir. Çok değişkenli uç değer analizi için ise Mahalonobis uzaklıkları hesaplanmıştır. 132, 204, 263 ve 400 numaralı katılımcıların çok değişkenli uç değer oldukları ($p < ,001$) tespit edilmiştir. Uç değerlerin analiz dışında bırakılmasına karar verilmiştir.

Normallik: Tek değişkenli normalliğin belirlenebilmesi için çarpıklık ve basıklık değerleri incelenebilir, histogram grafiğinin normal eğrisiyle uygunluğuna ya da Q-Q saçılma grafiğinde gözlenen ve beklenen değerlerin dağılımına bakılabilir (Çokluk ve ark., 2021). Çarpıklık ve basıklık değerlerinin +1,5 ile -1,5 arasında yer

alması tek değişkenli normallik varsayımının sağlandığını gösterir (Tabachnick & Fidell, 2007). Envanterin 2. ve 3. bölümlerinde yer alan ölçeklere ait hesaplanan çarpıklık ve basıklık değerleri Tablo 12’de verilmiştir.

Tablo 12. *Tek değişkenli normallik analizi sonuçları*

Ölçek Adı	Kadın		Erkek	
	Çarpıklık	Basıklık	Çarpıklık	Basıklık
Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği	-,73	,90	-,45	-,59
Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği	-,41	-,52	-,32	-,48
Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği	-,48	-,32	-,50	-,10

Tablo 12’de verilen analiz sonuçlarına göre ölçeklere ait çarpıklık ve basıklık değerlerinin -1 ile +1 arasında değişmektedir. Dolayısıyla tek değişkenli normallik varsayımı sağlanmıştır (Tabachnick & Fidell, 2007). Çok değişkenli normalliğin sağlanabilmesi için tek değişkenli normallik varsayımı önkoşuldur (Kline, 2016). Tek değişkenli normalliğin sağlanmasının ardından, Mplus 8.3 programı kullanılarak çok değişkenli çarpıklık ve basıklık testleri gerçekleştirilmiştir. Tablo 13’te analiz sonuçları verilmiştir.

Tablo 13. *Çok değişkenli normallik analizi sonuçları*

	Çok Değişkenli Çarpıklık				Çok Değişkenli Basıklık			
	Örneklem Değeri	\bar{x}	S	p	Örneklem Değeri	\bar{x}	S	p
Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği	25,31	5,46	,42	,00	238,89	167,2	1,76	,00
Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği	21,45	3,33	,34	,00	193,25	119,49	1,53	,00
Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği	24,56	4,27	,36	,00	231	142,34	1,66	,00

Tablo 13’teki analiz sonuçlarının anlamlı bulunması ($p < ,05$) çok değişkenli normalliğin sağlanmadığını göstermektedir (Şen, 2020).

Çoklu Bağntı: Bağımsız değişkenler arasındaki güçlü ilişkiler ($r > ,90$) çoklu bağntı olarak adlandırılır (Çokluk vd., 2021). Değişkenler arasındaki VIF (varyans artış faktörü) veya tolerans değerlerine bakılarak çoklu bağntı problemi test edilebilir (Tabachnick & Fidell, 2007). VIF değerinin 10'un üzerinde bulunması ve tolerans değerinin ise ,10 dan küçük olması çoklu bağntı problemine işaret eder (Çokluk vd., 2021). SPSS 23 programı kullanılarak envantere yer alan ölçek maddelerine ait VIF ve tolerans değerleri hesaplanmıştır. Analizler sonucunda değişkenler arasında çoklu bağntı probleminin olmadığı görülmüştür. VIF ve tolerans değerlerine ilişkin bilgiler Ek 3'te verilmiştir.

Varsayımların test edilmesinden sonra ise envantere güvenirlilik analizleri için Cronbach α ve McDonald's ω katsayıları JASP 0.16.3.0 programı kullanılarak kestirilmiştir. Mplus 8.3 programı ile DFA ve ölçme değişmezliği analizleri gerçekleştirilmiştir. DFA analizi ile envantere ikinci ve üçüncü bölümlerinde yer alan ölçeklerin orjinal faktör yapılarının Türkçe formunda aynı olup olmadığı test edilmiştir. Envantere cinsiyete göre ölçme değişmezliği analizi ise ÇGDFA tekniği ile yapılmıştır. Bu analizde modelde parametreler aşamalı olarak sınırlandırılmış ve yapısal, metrik, ölçek ve katı değişmelik aşamaları sırasıyla takip edilmiştir. Ölçme değişmezliğinin değerlendirilmesinde $\Delta CFI \leq ,01$ ve $\Delta RMSEA \leq ,015$ değerleri karar ölçütü olarak kullanılmıştır (Chen, 2007).

BÖLÜM IV

BULGULAR

Bu bölümde envanterin uyarlama ve uygulama aşamalarında elde edilen bulgular sırasıyla sunulmuştur.

4.1. Dil Eşdeğerliği ve Deneme Uygulaması

Envanterin İngilizce'den Türkçe'ye çevirisinde dil eşdeğerliğinin sağlanması amacıyla öncelikle uzman görüşleri alınmıştır. Uzman görüşleri doğrultusunda envantere son hali verilmiş, daha sonra ise envanterin sırasıyla İngilizce ve Türkçe formu 15 İngilizce öğretmenine iki hafta arayla uygulanmıştır. Dil eşdeğerliğin sağlanması amacıyla uzmanlar arasındaki uyum indeksleri ve öğretmenlerin envanterin İngilizce ve Türkçe formlarına verdikleri cevaplar arasındaki korelasyon değerleri hesaplanmıştır. Dil eşdeğerliğinin sağlanmasının ardından envanterin ön deneme uygulaması 26 öğretmenin katılımıyla gerçekleştirilmiştir.

Çevirisi yapılan envantere ilişkin uzman görüş formu hazırlanarak beşi Ölçme ve Değerlendirme alanından biri ise Eğitim Bilimleri alanından olmak üzere altı öğretim üyesinin görüşlerine başvurulmuştur. Çevirideki dil bilgisi hataları (anlamsal eşdeğerlik), her bir maddenin her iki kültürde aynı anlama gelip gelmedikleri (deyimsel eşdeğerlik), maddelerin hedef gruba uygulanabilirliği (deneysel eşdeğerlik) ve maddelerin kavramsal olarak kültürler arasındaki eşdeğerliği (kavramsal eşdeğerlik) uzman görüşleri dikkate alınarak sağlanmıştır (Borsa vd., 2012 akt. Çapık vd., 2018). Uzman görüşleri arasında uyum indeksleri hesaplanmış ve maddelere son hali verilmiştir. Tablo 14 ve 15'te envantere ilişkin madde ve bölüm düzeyinde hesaplanan uyum indeksleri sunulmuştur.

Tablo 14. *Uzman görüşleri arasındaki uyum indeksleri (senaryolar)*

Bölüm 1: Senaryolar							
Senaryo 1		Senaryo 2		Senaryo 3		Senaryo 4	
Madde No	Uyum İndeksi	Madde No	Uyum İndeksi	Madde No	Uyum İndeksi	Madde No	Uyum İndeksi
1	83,33	1	100	1	50	1	50
2	66,67	2	83,33	2	66,67	2	100
3	66,67	3	50	3	83,33	3	100
4	100	4	100	4	100	4	100
5	83,33	5	66,67	5	66,67	5	50
6	83,33	6	100	6	83,33	6	66,67
7	83,33	7	66,67	7	100	7	66,67
8	16,67	8	66,67	8	50	8	83,33
9	83,33	9	83,33	9	100	9	100
10	66,67	10	83,33	10	100	10	50
11	66,67	11	83,33	11	100	11	100
12	66,67	12	100	12	83,33	12	50
Genel Uyum İndeksi		Genel Uyum İndeksi		Genel Uyum İndeksi		Genel Uyum İndeksi	
72,22		81,94		80,56		70,83	

Tablo 14 incelendiğinde, senaryo maddeleri arasındaki uyum indekslerinin 16,67 ile 100 arasında değiştiği görülmektedir. Genel uyum indeksleri ise 70,83 ile 81,94 arasında değişmektedir. Uyum indekslerinin düşük olduğu maddeler uzmanların önerileri doğrultusunda düzenlenmiştir. Uzman görüşleri doğrultusunda çeviriden kaynaklanan dil bilgisi hataları düzeltilmiş, *rubrik ve müfredat* kavramları yerine *puanlama ve öğretim programı* kavramları kullanılmıştır. Deyimsel eşdeğerliğin sağlanması için “*professional judgement*” ifadesinin çevirisi “*profesyonel yargı*” olarak yapılmış ve ilgili madde düzenlenmiştir.

Tablo 15. *Uzman görüşleri arasındaki uyum indeksleri (ölçekler)*

Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği		Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği		Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği	
Madde No	Uyum İndeksi	Madde No	Uyum İndeksi	Madde No	Uyum İndeksi
1	100	1	83,33	1	100
2	83,33	2	100	2	83,33
3	100	3	100	3	100
4	100	4	100	4	100
5	83,33	5	100	5	100
6	100	6	100	6	100
7	83,33	7	66,67	7	83,33
8	83,33	8	83,33	8	50
9	33,33	9	100	9	100
10	83,33	10	100	10	100
11	100			11	100
12	100				
Genel Uyum İndeksi 87,5		Genel Uyum İndeksi 94,44		Genel Uyum İndeksi 92,42	

Tablo 15'teki analiz sonuçlarına göre ölçek maddelerine ilişkin uzman görüşleri arasındaki uyum indekslerinin ise 50 ile 100 arasında değiştiği, genel uyum indekslerinin ise 87,5, 92,42 ve 94,44 olduğu görülmektedir. Uzman görüşleri doğrultusunda maddelere ilişkin dil bilgisi hataları giderilmiş, Bölüm 3A madde 7'de "eşit değerlendirme koşulları" ifadesi yerine "uygun değerlendirme koşulları" ifadesi kullanılmıştır.

Uzman görüşlerinin alınmasının ardından 15 İngilizce öğretmenine envanterin önce İngilizce ve daha sonra ise Türkçe formu iki hafta ara ile uygulanmıştır. Öğretmenlerin her iki forma verdikleri cevaplara ait betimsel istatistikler JASP (versiyon 0.16.3) programı ile hesaplanmıştır. Verilere ilişkin normallik analizi sonuçları Ek 4'te verilmiştir. Analiz sonuçlarına göre; -1,5 ile +1,5 sınırları ile uyumlu olmayan çarpıklık ve basıklık değerlerinin olduğu, dolayısıyla verilerin normal dağılım göstermediği görülmüştür (Tabachnick & Fidell, 2007). Normal dağılımın olmaması nedeniyle, öğretmenlerin Türkçe ve İngilizce formuna verdikleri cevaplar arasındaki ilişki Spearman Sıra Farkları korelasyon analizi yapılarak incelenmiştir. Analiz sonuçları Tablo 16'da sunulmuştur.

Tablo 16. *Dil eşdeğerliği korelasyon değerleri*

Bölüm 1: Senaryolar		
Temalar	Alt Boyutlar	Spearman's rho
Değerlendirme Amaçları	Öğrenmenin Değerlendirilmesi	,58
	Öğrenme İçin Değerlendirme	,73
	Öğrenirken Değerlendirme	,74
Değerlendirme Süreçleri	Tasarım	,56
	Kullanım/Puanlandırma	,47
	İletişim	,75
Değerlendirme Adilliği	Standart Yaklaşım	,56
	Eşitlikçi Yaklaşım	,37
	Farklılaştırılmış Yaklaşım	,61
Değerlendirme Teorisi	Tutarlı	,62
	İçeriksel	,83
	Dengeli	,62
Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği		,48
Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği		,72
Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği		,67

Tablo 16'da envanterin ilk bölümünde yer alan senaryolar için alt boyutlarda, envanterin ikinci ve üçüncü bölümlerinde ise ölçeklerden elde edilen puanlar arasında hesaplanan Spearman sıra farkları korelasyon katsayıları verilmiştir. Elde edilen bulgular sonucunda korelasyon değerlerinin ,37 ile ,75 arasında değiştiği, envanterin

Türkçe ve İngilizce formları arasında pozitif yönde, orta ve yüksek düzeyde ilişkiler bulunduğu görülmüştür. Dolayısıyla formlar arasında dil eşdeğerliğinin sağlandığını ifade edilebilir.

Dil eşdeğerliği için kanıtlar elde edildikten sonra envanterin deneme uygulaması 26 öğretmenin katılımıyla gerçekleştirilmiştir. Deneme uygulaması sonucunda envanterin cevaplanma süresinin ortalama 20 dakika olduğu, maddelerde anlaşılmayan ifadeler bulunmadığı görülmüştür. Deneme uygulamasına ait betimsel istatistikler Ek 5'te verilmiştir.

Dil eşdeğerliği analizleri ve deneme uygulaması sonrasında envanter 405 öğretmenin katılımıyla uygulanmıştır. Bu uygulama sonucunda elde edilen verilerle sırasıyla madde analizleri, ki-kare analizi, DFA ve ölçme değişmezliği analizleri gerçekleştirilmiştir. Bu analizlere ait bulgular sırasıyla verilmiştir.

4.2. Madde Analizleri

Envanterde yer alan ölçeklere ait madde analizleri gerçekleştirilmiştir. SPSS 23 programı kullanılarak betimsel istatistikler hesaplanmış ve Ek 6'da verilmiştir. Betimsel analiz sonuçlarına göre maddelere ait çarpıklık ve basıklık değerlerinin -1,5 ile +1,5 arasında değiştiği, maddelerin normal dağılım gösterdikleri görülmüştür (Tabachnick & Fidell, 2007). Ölçek maddelerine ilişkin madde-toplam test korelasyonları ise Tablo 17'de verilmiştir.

Tablo 17. *Madde-toplam test korelasyonları*

Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği		Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği		Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği	
Madde No	Madde-Toplam Test Korelasyonu	Madde No	Madde-Toplam Test Korelasyonu	Madde No	Madde-Toplam Test Korelasyonu
m1	,69	m1	,74	m1	,8
m2	,66	m2	,82	m2	,81
m3	,7	m3	,82	m3	,77
m4	,63	m4	,82	m4	,87
m5	,78	m5	,82	m5	,82
m6	,75	m6	,7	m6	,85
m7	,69	m7	,62	m7	,77
m8	,71	m8	,72	m8	,82
m9	,78	m9	,79	m9	,71
m10	,76	m10	,78	m10	,69
m11	,72			m11	,67
m12	,76				

Tablo 17'deki madde-toplam test korelasyonları ,62 ile ,868 arasında değişmektedir. Korelasyon değerinin ,30 ve üzeri olması maddenin çok iyi bir ayırt edici olduğu anlamına gelir (İlhan, 2022). Dolayısıyla ölçeklerde yer alan maddelerin çok iyi ayırt edici oldukları ve madde düzeyinde geçerliğin sağlandığı ifade edilebilir (Erkuş, 2021). Madde ayırt ediciliklerinin $r_{jx} \geq ,30$ üzerinde olması sebebiyle, uyarılama aşamasında ölçeklerden herhangi bir madde çıkarılmamıştır.

4.3. Güvenirlik Analizi

Envanterde yer alan ölçeklere ait güvenirlik analizleri için JASP 0.16.3.0 programı ile Cronbach α ve McDonald's ω katsayıları kestirilmiştir. Konjenerik maddelere ait (farklı faktör yüklerine sahip maddeler) güvenirlik analizinde Cronbach α değerinin yanı sıra sonuçlar vermesi nedeniyle McDonald's ω katsayısı da hesaplanmıştır (Yurdugül, 2006). Tablo 18'de güvenirlik analizi sonuçları verilmiştir.

Tablo 18. *Güvenirlik analizi sonuçları*

Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği	\bar{x}	S	Cronbach's α	McDonald's ω
Tüm Ölçek	48,91	7,68	,91	,91
Alt Boyut 1 (DTUG)	27,7	5,22	,88	,88
Alt Boyut 2 (İADSI)	21,21	3,04	,84	,83
Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği				
Tüm Ölçek	41,1	6,42	,92	,91
Alt Boyut 1 (DUBİ)	24,65	3,94	,85	,85
Alt Boyut 2 (MDTİU)	16,45	2,97	,93	,93
Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği				
Tüm Ölçek	41,43	9,57	,94	,94
Alt Boyut 1 (Çevrimiçi)	18,55	4,88	,91	,92
Alt Boyut 2 (Yüz Yüze)	14,96	3,85	,89	,88
Alt Boyut 3 (Bire Bir)	7,92	1,95	,79	,79

Analiz sonuçlarına göre güvenirlik katsayılarının ,79 ile ,92 arasında değiştiği görülmektedir. Elde edilen değerlerin ,70 ve üzerinde olması yapılan ölçümlerin yeterli düzeyde güvenilir olduğunu göstermektedir (Büyüköztürk, 2021).

4.4. Ki Kare Analizi

Envanterin ilk bölümünde yer alan senaryoların maddeleri değerlendirme temalarına göre kategorik olarak yapılandırılmıştır. Dolayısıyla senaryo maddelerine verilen cevapların analizinde Pearson Ki kare testi kullanılmış, cevaplar ile cinsiyet ve mesleki kıdem değişkenleri arasındaki ilişkiler incelenmiştir. Tablo 19’da ki kare analizi sonuçları verilmiştir.

Tablo 19. *Ki kare analizi sonuçları*

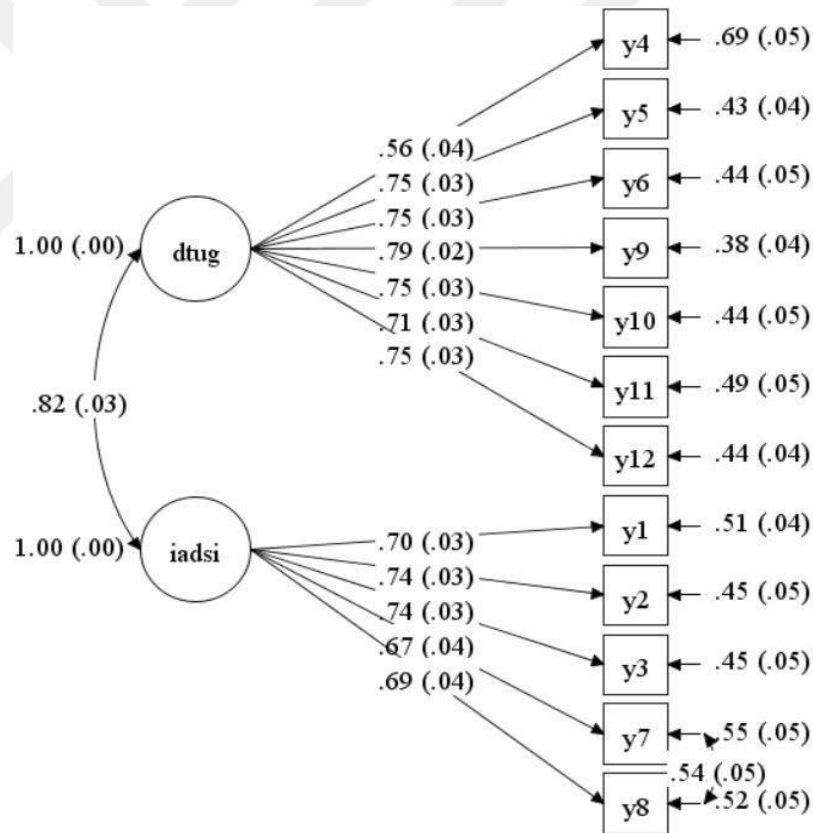
Değerlendirme Teması	Alt Boyutlar	Cinsiyet		Mesleki Kıdem (Yıl)						
		Kadın	Erkek	1-5	6-10	11-15	16-20	21-25	26-30	31-35
Değerlendirme Amacı	Öğrenmenin Değerlendirmesi	21,99	19,42	25,86	17,78	24,62	18,33	16,13	15,38	0
		p=,3		p=,55						
	Öğrenme İçin Değerlendirme	44,81	37,41	34,48	44,45	47,69	38,33	51,62	53,85	60
		p=,07		p=,18						
Değerlendirme Süreçleri	Öğrenirken Değerlendirme	33,2	43,17	39,66	37,78	27,69	43,33	32,26	30,77	40
		p=,03		p=,35						
	Tasarım	45,38	46,21	40,83	46,94	47,3	53,13	39,39	50	50
		p=,44		p=,67						
Değerlendirme Süreçleri	Kullanım/ Puanlandırma	8,46	7,59	6,67	9,18	10,81	7,81	9,09	0	0
		p=,41		p=,81						
	İletişim	46,15	46,21	52,5	43,88	41,89	39,06	51,52	50	50
	p=,38		p=,46							
Değerlendirme Adilliği	Standart Yaklaşım	8,98	16,67	8,49	9,57	18,18	9,52	14,71	12,5	50
		p=,03		p=,06						
	Eşitlikçi Yaklaşım	45,71	50,72	51,89	53,19	42,42	44,44	41,18	37,5	25
		p=,28		p=,59						
Değerlendirme Teorisi	Farklılaştırılmış Yaklaşım	45,31	32,61	39,62	37,23	39,39	46,03	44,12	50	25
		p=,03		p=,35						
	Tutarlı	17,62	17,36	17,43	18,68	19,44	14,52	19,44	14,29	0
	p=,56		p=,9							
Değerlendirme Teorisi	İçeriksel	62,3	61,11	63,3	58,24	65,28	61,29	63,89	57,14	50
		p=,51		p=,45						
	Dengeli	20,08	21,53	19,27	23,08	15,28	24,19	16,67	28,57	50
	p=,39		p=,42							

p<,05

Tablo 19’deki ki kare analizi sonuçlarına göre sadece cinsiyet değişkeni ile öğrenirken değerlendirme, standart yaklaşım ve farklılaştırılmış yaklaşım alt boyutlarında anlamlı ilişkilerin olduğu belirlenmiştir (p<,05). Öğrenirken değerlendirme ve standart yaklaşım alt boyutlarında yer alan maddeleri erkek öğretmenlerin kadın öğretmenlere göre daha çok tercih ettiği, farklılaştırılmış yaklaşım boyutunda yer alan maddeleri ise kadın öğretmenlerin erkek öğretmenlere göre daha çok tercih ettiği görülmüştür. Mesleki kıdem değişkeni ile temalar arasında anlamlı bir ilişki ise bulunmamıştır (p>,05).

4.5. DFA Bulguları

DFA ile envanterde yer alan ölçeklerin orijinal faktör yapılarının Türkçe'ye uyarlaması yapılan formda aynı olup olmadığı Mplus 8.3 programı ile test edilmiştir. Çok değişkenli normallik varsayımının sağlanamaması nedeniyle kestirim yöntemi olarak MLM (güçlü en çok olabilirlik yöntemi) kullanılmıştır (Şen, 2020). Uyum iyiliğinin değerlendirilmesinde RMSEA, CFI, TLI (NNFI) ve SRMR uyum indeksleri yorumlanmıştır. Ki kare uyum indeksinin kullanılacağı veri setinin çoklu normallığı sağlaması gerektiğinden (West vd., 2012), ki kare uyum indeksi model-veri uyumunda ölçüt olarak kullanılmamıştır. Envanterde yer alan ölçeklere ilişkin öncelikle DFA modelleri verilmiş, daha sonra ise modellere ait uyum indeksleri yorumlanmıştır. Standart hata değerleri ise parantez içerisinde verilmiştir.



Şekil 9. Sınıf içi değerlendirmede güven ölçeği DFA modeli

Şekil 9'daki Sınıf İçi Değerlendirmede Güven Ölçeği'ne ait DFA modeli incelendiğinde tüm yolların $p < .001$ düzeyinde manidar olduğu, standartlaştırılmış faktör yüklerinin ,56 ile ,79 arasında ve hata varyanslarının ise ,38 ile ,69 arasında

değiştirdiği görülmektedir. Faktörler arasındaki korelasyon değeri ise ,82'dir. Yedi ve sekiz numaralı gözlenen değişkenler arasında ise modifikasyon işlemi (MI) yapılmıştır. İlgili maddeler sırasıyla verilmiştir.

m7. *Öğrencilerimin öğrenmelerini geliştirmek için onlara anlık geribildirimler veririm.*

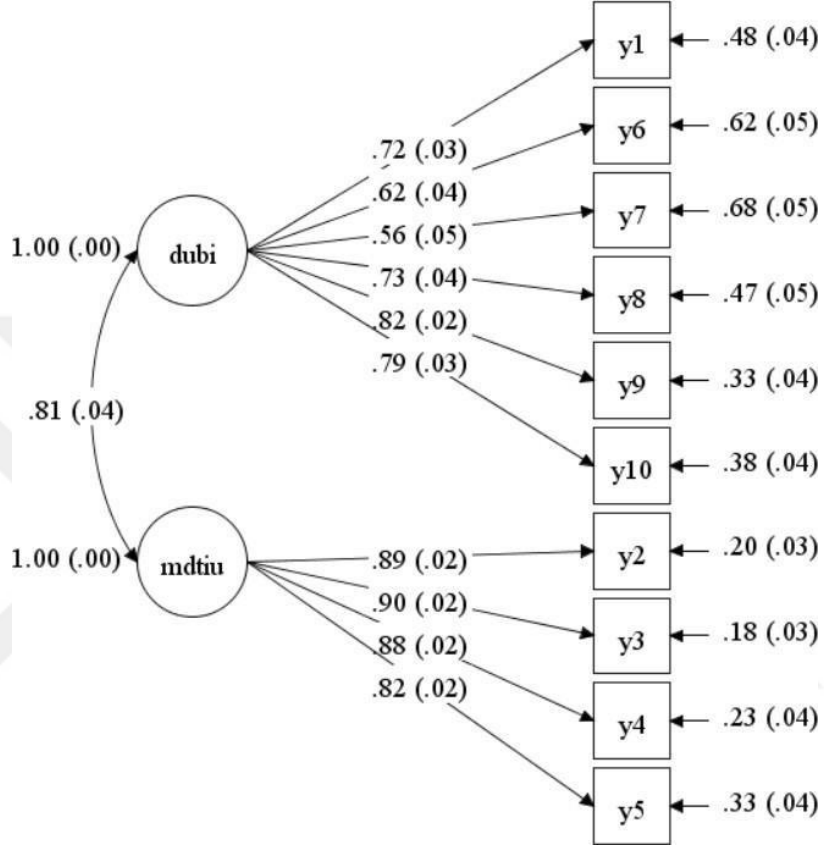
m8. *Öğrencilerimin öğrenmelerini geliştirmek için onlara faydalı geribildirimler veririm.*

Maddeler incelendiğinde maddeler arasında pozitif yönlü, yüksek düzeyde ilişki ($r=,751$; $p<,001$) olduğu görülmüştür. Maddelere ait VIF ($m_7=2,79$; $m_8=2,58$) ve tolerans değerleri ($m_7=,37$; $m_8=,36$) ise çoklu bağıntı probleminin olmadığını göstermektedir (Çokluk vd., 2021). Madde ayırt edicilik değerlerinin ise m7 için ,69 ve m8 için ,71'dir. ($r_{jx}\geq,30$; İlhan, 2022). Elde edilen bulgular doğrultusunda istatistiksel olarak maddelerden herhangi birinin ölçekten çıkartılmamasına karar verilmiştir. Madde içerikleri incelendiğinde ise m7'nin geribildirim zamanı, m8'in ise geribildirim verilerek öğrenci öğrenmesini desteklenmesi ile ilgili olduğu görülmüştür. Dolayısıyla maddeler arasında modifikasyon işlemi (MI) yapılmış ve DFA analizi gerçekleştirilmiştir.

DFA sonuçlarına göre faktör yüklerinin ,45'den yüksek olduğu (Büyüköztürk, 2021) ve faktörler arasında ayırt edici geçerliğin ($r<,90$) sağlandığı görülmektedir (Kline, 2016). Analiz sonucunda hesaplanan uyum indeksleri ise RMSEA=,06 (,05-,07), CFI=,96, TLI(NNFI)=,95 ve SRMR=,04'tür. RMSEA uyum indeksinin kabul edilebilir seviyede olduğu, CFI, TLI(NNFI) ve SRMR indekslerinin ise mükemmel seviyede olduğu (Çokluk vd., 2021; Hu & Bentler, 1999; Kline, 2016; Şen, 2020), dolayısıyla ölçeğin faktör yapısının doğrulandığı sonucuna varılmıştır.

Şekil 10'daki Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği'ne ait DFA modeli incelendiğinde tüm yolların $p<,001$ düzeyinde manidar olduğu, standartlaştırılmış faktör yüklerinin ,56 ile ,9 arasında, hata varyanslarının ise ,18 ile ,68 arasında değiştiği görülmektedir. Faktörler arasındaki korelasyon değeri ise ,81'dir. Elde edilen değerlere göre maddelere ilişkin faktör yüklerinin ,45'ten büyük olduğu (Büyüköztürk, 2021) ve faktörler arasında ayırt edici geçerliğin ($r<,90$) sağlandığı ifade edilebilir (Kline, 2016). Uyum indeksleri ise RMSEA=,07 (,06-,09), CFI=,96, TLI(NNFI)=,94 ve SRMR=,04'tür, RMSEA ve TLI(NNFI)'nin kabul edilebilir

seviyede olduğu, CFI ve SRMR değerlerinin ise mükemmel seviyede uyum gösterdiği görülmektedir (Çokluk vd., 2021; Hu & Bentler, 1999; Kline, 2016; Şen, 2020). Analiz sonuçlarına göre ölçeğin faktör yapısı doğrulanmıştır.



Şekil 10. Değerlendirmeye ilişkin mesleki öğrenme öncelikleri ölçeği DFA modeli

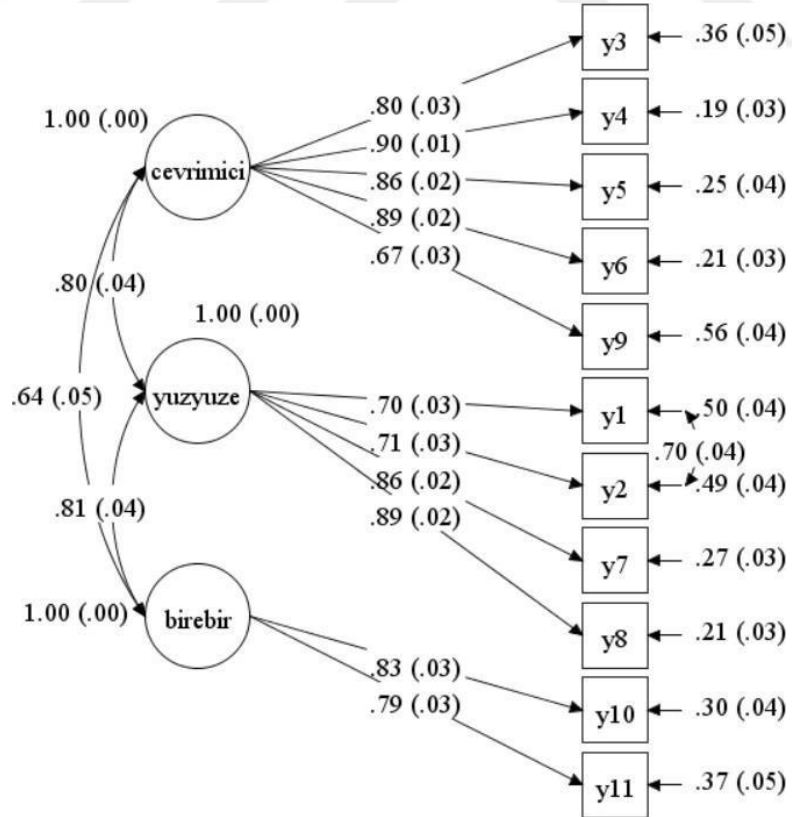
Şekil 11'deki Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği'ne ait DFA modeli incelendiğinde tüm yolların $p < .001$ düzeyinde manidar olduğu, standartlaştırılmış faktör yüklerinin ,67 ile ,9 arasında, hata varyanslarının ise ,19 ile ,56 arasında değiştiği görülmektedir. Faktör yüklerinin ,45'ten büyük olduğu (Büyüköztürk, 2021), faktörler arasındaki korelasyon değerlerinin ise ,9'dan küçük olduğu görülmektedir. Faktörler arasında ayırt edici geçerlik sağlanmıştır ve faktör yükleri yüksek seviyededir (Kline, 2016).

Bir ve iki numaralı gözlenen değişkenler arasında modifikasyon işlemi (MI) yapılmıştır. İlgili maddeler sırasıyla verilmiştir.

m1. *Alanda çalışan diğer öğretmenlerle birlikte yüz yüze tam zamanlı değerlendirme kurslarına katılıyorum.*

m2. *Alanda çalışan diğer öğretmenlerle birlikte yüz yüze kısa zamanlı değerlendirme modüllerine katılıyorum.*

Maddeler incelendiğinde maddeler arasında pozitif yönlü ve yüksek düzeyde ilişki ($r=,848$; $p<,001$) olduğu görülmüştür. Maddelere ait VIF ($m_1=4,04$; $m_2=4,01$) ve tolerans değerleri ($m_1=,25$; $m_2=,24$) ise çoklu bağıntı probleminin olmadığını göstermektedir (Çokluk vd., 2021). Madde ayırt edicilik değerlerinin ise m1 için ,8 ve m2 için ,81'dir. ($r_{jx} \geq ,30$; İlhan, 2022). Elde edilen bulgular doğrultusunda istatistiksel olarak herhangi bir maddelinin ölçekten çıkartılmamasına karar verilmiştir. Madde içerikleri incelendiğinde ise maddelerin değerlendirme ile ilgili verilen eğitimlerin süreleriyle ilgili olduğu görülmüştür. Dolayısıyla maddeler arasında modifikasyon işlemi (MI) yapılmış ve DFA analizi gerçekleştirilmiştir.



Şekil 11. Değerlendirmeye ilişkin mesleki öğrenme tercihleri ölçeği DFA modeli

DFA sonucu elde edilen uyum indeksleri RMSEA=,07 (,05-,08), CFI=,97, TLI(NNFI)=,95 ve SRMR=,06'dır. Analiz sonuçlarına göre RMSEA ve SRMR değerlerinin kabul edilebilir seviyede olduğu, CFI ve TLI(NNFI)'nın değerlerinin ise mükemmel uyum gösterdiği görülmüştür (Çokluk vd., 2021; Hu & Bentler, 1999; Kline, 2016; Şen, 2020). Analiz sonuçlarına göre ölçeğin faktör yapısı doğrulanmıştır. Envanterde yer alan ölçeklerin faktör yapılarının doğrulanmasının ardından ise envanterin cinsiyete göre ölçme değişmezliği incelenmiştir.

4.6. Ölçme Değişmezliği Bulguları

Meredith (1993) ölçme değişmezliğinin yapısal, metrik, ölçek ve katı değişmezlik olmak üzere dört aşamada incelenmesini önermiştir. Yapısal değişmezlik aşamasında ölçek maddelerinin farklı gruplarda aynı yapıyı ölçüp ölçmediği test edilir. Bir sonraki aşamada ise parametrelere ek sınırlamalar getirilerek model-veri uyumu incelenir ve bir önceki modelle karşılaştırılır. Uyum indekslerindeki değişimler incelenerek ölçme değişmezliğinin sağlanıp sağlanmadığına karar verilir. Ölçme değişmezliğinin sağlanmadığı durumda bir sonraki aşama test edilmez.

Bu araştırmada ise hiyerarşik olarak kurulan modellerle, öncelikle yapısal değişmezlik analizi yapılarak ölçek maddelerin farklı gruplarda aynı yapıyı ölçüp ölçmedikleri test edilmiştir. Daha sonra ise sırasıyla metrik, ölçek ve katı değişmezlik aşamaları test edilmiştir. Her bir aşamada kurulan modellerin veriye uyum düzeyinin değerlendirilmesinde CFI, RMSEA ve TLI(NNFI) değerleri analiz edilmiştir. Ölçme değişmezliğinin testinde ise her bir değişmezlik aşamasında belirlenen uyum indeksleri ile yapısal değişmezlik aşamasında belirlenen CFI ve RMSEA uyum indeksleri arasındaki farklar incelenmiştir.

Veri setinin çoklu normalliği sağlamaması nedeniyle modeller arasındaki $\Delta\chi^2$ fark değerleri ölçme değişmezliğinde karar ölçütü olarak kullanılmamıştır. Brown (2006) ise çoklu normallik varsayımının sağlanmadığı durumlarda model uyumun değerlendirilmesinde $\Delta\chi^2$ farkının kullanımının uygun olmadığını belirtmiştir (akt. Önen, 2009). Çalışma grubunun 405 kişiden ($n>300$) oluşması nedeniyle $\Delta CFI \leq ,01$ ve $\Delta RMSEA \leq ,015$ değerleri ölçme değişmezliğinde karar ölçütü olarak kullanılmıştır (Chen, 2007). Tablo 20'de envanterde yer alan ölçeklere ilişkin cinsiyete göre ölçme değişmezliği analizi bulguları verilmiştir.

Tablo 20. Cinsiyete göre ölçme değişmezliği bulguları

Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği								
Aşamalar	χ^2 (p<,05)	sd	CFI	RMSEA	TLI (NNFI)	Δ CFI	Δ RMSEA	
Yapısal Değişmezlik	228,68	104	,94	,08 (.06 - ,09)	,92	-	-	
Metrik Değişmezlik	238,25	114	,94	,07 (.06 - ,09)	,93	0	-,01	
Ölçek Değişmezliği	247,44	124	,94	,07 (.06 - ,08)	,93	0	-,01	
Katı Değişmezlik	255,65	136	,94	,07 (.05 - ,08)	,94	0	-,01	

Tablo 21. (Devamı)

Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği								
Aşamalar	χ^2 (p<,05)	sd	CFI	RMSEA	TLI (NNFI)	Δ CFI	Δ RMSEA	
Yapısal Değişmezlik	165,84	68	,95	,08 (.07 - ,1)	,93	-	-	
Metrik Değişmezlik	180,79	76	,94	,08 (.07 - ,1)	,93	-,01	0	
Ölçek Değişmezliği	189,21	84	,94	,08 (.06 - ,09)	,94	-,01	0	
Katı Değişmezlik	187,06	94	,95	,07 (.06 - ,09)	,95	0	-,01	

Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği

Aşamalar	χ^2 (p<,05)	sd	CFI	RMSEA	TLI (NNFI)	Δ CFI	Δ RMSEA	
Yapısal Değişmezlik	164,49	80	,96	,07 (.06 - ,09)	,95	-	-	
Metrik Değişmezlik	175,31	88	,96	,07 (.06 - ,09)	,95	0	0	
Ölçek Değişmezliği	195,08	96	,96	,07 (.06 - ,09)	,95	0	0	
Katı Değişmezlik	217,31	107	,95	,07 (.06 - ,09)	,95	-,01	0	

Tablo 20'deki analiz sonuçlarına göre ölçeklerin yapısal değişmezlik aşamasında model veri uyumunun kabul edilebilir ve mükemmel seviyede olduğu görülmektedir ($,95 \geq CFI \geq ,90$; $RMSEA \geq ,08$; $,95 \geq TLI(NNFI) \geq ,90$). Bu sonuçlara göre ölçeklerin kadın ve erkek gruplarında aynı yapıyı ölçtükleri söylenebilir. Yapısal değişmezliğin sağlanması ölçek maddelerinin cevaplanmasında grupların aynı kavramsal bakış açılarına sahip olduğunu ve yapısal değişmezlik modelinin temel alınarak diğer değişmezlik testlerinin yapılabileceği anlamına gelir (Önen, 2009). Bu nedenle yapısal değişmezliğin sağlanmasının ardından metrik değişmezlik aşamasına geçilmiştir.

Metrik değişmezlik aşamasında faktör yüklerinin gruplar arasında eşit olduğu, diğer parametrelerin ise serbest bırakıldığı model test edilir. Ölçme aracındaki birimsel eşitliğin gruplar arasında test edilmesi söz konusudur. Tablo 21'deki analiz sonuçları incelendiğinde metrik değişmezlik aşamasında model-veri uyumunun kabul edilebilir ve mükemmel seviyede olduğu görülmektedir ($,95 \geq CFI \geq ,90$; $RMSEA \geq ,08$; $,95 \geq TLI(NNFI) \geq ,90$). Metrik değişmezliğin test edilmesinde ise yapısal ve metrik değişmezlik aşamalarında elde edilen CFI ve RMSEA uyum indekslerindeki değişimler hesaplanmıştır. Analiz sonucunda ölçeklerin metrik değişmezlik aşamasını sağladıkları görülmüştür ($\Delta CFI \leq ,01$ ve $\Delta RMSEA \leq ,015$). Dolayısıyla kadın ve erkek gruplarının

ölçek maddelerini aynı şekilde anlayıp, yorumladıkları söylenebilir (Şen, 2020). Metrik değişmezliğin sağlanmasının ardından ölçek değişmezliği aşamasına geçilmiştir.

Ölçek değişmezliği aşamasında faktör yükleri ve gösterge sabitlerinin gruplar arasında eşit olduğu hipotez test edilir. Tablo 21'deki analiz sonuçları incelendiğinde ölçek değişmezlik aşamasında model-veri uyumunun kabul edilebilir ve mükemmel seviyede olduğu görülmektedir ($,95 \geq CFI \geq ,90$; $RMSEA \geq ,08$; $,95 \geq TLI(NNFI) \geq ,90$). Ölçek değişmezliğin test edilmesinde ise yapısal ve ölçek değişmezlik aşamalarında elde edilen CFI ve RMSEA uyum indekslerindeki değişimler hesaplanmıştır. Analiz sonucunda ölçeklerin ölçek değişmezlik aşamasını sağladıkları görülmüştür ($\Delta CFI \leq ,01$ ve $\Delta RMSEA \leq ,015$). Bu sonuçlar ölçeklerden elde edilen puanlarla gruplar arasında karşılaştırma yapılabileceği ve ölçümlere yanlılık karışmayacağı anlamına gelmektedir. Ölçek değişmezliğin sağlanmasından sonra ise katı değişmezlik aşamasına geçilmiştir.

Katı değişmezlik aşamasında gruplar arasında faktör yükleri, gösterge sabitleri ve hata varyansları eşitliği hipotezi test edilir. Tablo 21'deki analiz sonuçları incelendiğinde katı değişmezlik aşamasında model-veri uyumunun kabul edilebilir ve mükemmel seviyede olduğu görülmektedir ($,95 \geq CFI \geq ,90$; $RMSEA \geq ,08$; $,95 \geq TLI(NNFI) \geq ,90$). Katı değişmezlik ve yapısal değişmezlik aşamaları arasındaki CFI ve RMSEA değerlerindeki değişimler incelendiğinde katı değişmezliğin sağlandığı anlaşılmaktadır ($\Delta CFI \leq ,01$; $\Delta RMSEA \leq ,015$). Ölçme değişmezliği analizi sonuçlarına göre Sınıf İçi Değerlendirme Envanterinde yer alan ölçekler kullanılarak kadın ve erkek grupları için yapılacak tüm karşılaştırmalarda güvenilir ve geçerli ölçümler elde edilebileceği ifade edilebilir.

BÖLÜM V

TARTIŞMA

Bu bölümde araştırmada elde edilen bulgular, araştırma problemleri çerçevesinde sırasıyla tartışılacaktır.

5.1. Araştırma Bulgularının Tartışılması

Bu araştırmada cevaplanan ilk araştırma problemi “*Türkçe’ye uyarlaması yapılan Sınıf İçi Değerlendirme Yaklaşımları Envanteri’nin güvenilirlik kanıtları nelerdir?*” şeklinde belirlenmiştir. Güvenirlik kanıtları olarak envanterin her bir bölümüne ait Cronbach α ve McDonald’s ω katsayıları kestirilmiştir. Türkiye’de gerçekleştirilen ölçek geliştirme ve uyarlama çalışmalarında ise çoğunlukla Cronbach α iç tutarlılık güvenilirlik katsayısının hesaplandığı alanyazınından anlaşılmaktadır (Çüm ve Koç, 2013; Güvendir ve Özkan, 2015; Öztürk ve ark., 2015). Bu durum bu çalışmalarda geliştirilen ya da uyarlanan ölçme araçlarındaki maddelerin paralel, eşdeğer ya da eşbiçimli olarak kabul edildiklerini göstermektedir. Maddelerin eşit faktör yüklerine sahip oldukları varsayılmaktadır.

DFA analizi yapılarak maddelerin faktör yüklerinin hesaplanabilmesi ve böylelikle maddelerin konjenerik bir yapıda olup olmadığının sorgulanması mümkündür (Yurdugül, 2006). Konjenerik maddelerin yer aldığı ölçme aracının güvenilirlik analizinde Cronbach α değerinin yanı sıra sonuçlar verdiği, güvenilirlik değerinin gerçek değerinin altında hesaplandığı bildirilmektedir (Yurdugül, 2005). Dolayısıyla ölçme aracında konjenerik maddelerin varlığı durumunda güvenilirlik analizinde Cronbach α katsayısının yanı sıra McDonald’s ω ya da Revelle’s β katsayılarının da hesaplanması önerilmektedir (Yurdugül, 2006). Bu araştırmada ise

alanyazınındaki diğer ölçek uyarlama çalışmalarından farklı olarak Cronbach α ve McDonald's ω katsayıları envantere yer alan ölçekler için alt boyutlarda ve tüm ölçek düzeyinde kestirilmiştir (Çüm ve Koç, 2013; Güvendir ve Özkan, 2015; Öztürk ve ark., 2015). Envanterin orijinal formunda yer alan ölçeklere ait güvenilirlik kestirimlerinde ise sadece Cronbach α iç tutarlılık katsayısının raporlandığı görülmüştür (DeLuca vd., 2016b). Orijinal formdaki ölçeklere ilişkin Cronbach α değerleri 0,74 ile 0,92 arasında değişmektedir. Ölçeklerin uyarlanan formlarına ilişkin hesaplanan Cronbach α ve McDonald's ω değerleri ise ,79 ile ,94 arasındadır. Elde edilen güvenilirlik katsayılarının ,70 ve üzerinde olması yeterli düzeyde güvenilir ölçümler yapıldığını göstermektedir (Büyüköztürk, 2021).

Bu araştırmada cevaplanan ikinci araştırma problemi ise “*Türkçe’ye uyarlaması yapılan Sınıf İçi Değerlendirme Yaklaşımları Envanteri’nin geçerlilik kanıtları nelerdir?*” şeklinde belirlenmiştir. İkinci araştırma problemine ait alt problemler ise “*Sınıf İçi Değerlendirme Yaklaşımları Envanterinden elde edilen puanlar cinsiyete göre ölçme değişmezliği aşamalarından yapısal, metrik, ölçek ve katı değişmezliği(ni) sağlamakta mıdır?*” şeklinde belirlenmiştir.

Uyarlama çalışmasında öncelikle yapısal ve dil eşdeğerliğinin sağlanması amacıyla uzman görüşlerine başvurulmuştur. Envanterde ölçülmek istenen yapının, öğretmen değerlendirme okuryazarlığı, her iki kültürde de var olduğu ve aynı anlama geldiği uzman görüşleri alınarak belirlenmiştir. Envanterin çevirisi yapıldıktan sonra dil eşdeğerliğinin sağlanması amacıyla da uzman görüşlerine başvurulmuştur. Uzman görüşleri arasında uyum indeksleri hesaplanmıştır. Hedef ve kaynak kültürler arasında anlamsal, deyimsel, deneysel ve kavramsal eşdeğerliğin sağlanabilmesi amacıyla uzmanlardan elde edilen geri dönütler incelenmiş ve maddelere son hali verilmiştir.

Envanterin orijinal ve uyarlanmış formları arasındaki dil eşdeğerliğinin sağlanması amacıyla 15 İngilizce öğretmenine iki hafta arayla önce orijinal form daha sonra uyarlanmış form uygulanmıştır. Verilerin analizi aşamasındaki hedef ve kaynak formdan elde edilen puanlar arasındaki korelasyon değerleri analiz edilmiştir. Analiz sonucunda Türkçe ve İngilizce formlarına verilen cevaplar arasında orta ve yüksek düzeyde ilişkiler bulunduğu ve böylece dil eşdeğerliğinin sağlandığı sonucuna varılmıştır.

Dil eşdeğerliğinin sağlanmasının ardından envanterin deneme uygulaması gerçekleştirilmiştir. 26 öğretmenin katılımıyla yapılan uygulama sonucunda ortalama cevaplama süresinin 20 dakika olduğu, envantere anlaşılmayan ifadelerin bulunmadığı görülmüştür. Deneme uygulamasından sonra ise envanterin son uygulaması 405 öğretmenin katılımıyla gerçekleştirilmiştir. Uygulama sonucunda madde düzeyinde ayırt ediciliklerin ,30'tan yüksek olduğu ve böylece madde düzeyinde geçerliğin sağlandığı görülmüştür (Erkuş, 2021).

Envantere yer alan ölçeklerin orijinal faktör yapılarının doğrulanması için ise doğrulayıcı faktör analizleri gerçekleştirilmiş ve yapı geçerliği kanıtları elde edilmiştir. DFA öncesinde kayıp değer, uç değer, normallik ve çoklu bağıntı varsayımlarının kontrolü gerçekleştirilmiştir. Çok değişkenli istatistiksel analizlerde varsayımların kontrolü geçerli ölçümler yapılabilmesi için şarttır (Çokluk vd., 2021). Kılıç ve Koyuncu (2017) araştırmalarında 2006-2016 yılları arasında yayımlanan 170 ölçek uyarlama çalışmasını incelemişler ve bu çalışmalar içerisinde DFA'nın kullanıldığı çalışmaların %90'unda varsayımların kontrol edilmediğini ya da raporlanmadığını belirtmişlerdir. Bu araştırmaların % 89,56'sının 600'den az örneklem büyüklüğüne sahip olduklarını ve araştırmaların çoğunluğunun 252 örneklem büyüklüğüyle gerçekleştirildiğini raporlamışlardır. Çok değişkenli istatistiksel analizlerde örneklem büyüklüğünü kaç olması gerektiği konusunda görüş birliği bulunmamaktadır. Parametre sayısının 10 ya da 20 katı örneklem büyüklüğünü öneren araştırmacıların yanı sıra örneklem büyüklüğünün en az 200 kişiden oluşması gerektiğini öneren araştırmacılar da mevcuttur (Everitt, 1975; Kline, 2016; Şen, 2020) Bu farklı öneriler doğrultusunda envanterin uyarlama çalışmasının yeterli örneklem büyüklüğüyle gerçekleştirildiği söylenebilir.

Varsayımların kontrolünden sonra ise DFA yapılmış, model-veri uyumunun değerlendirilmesinde RMSEA, CFI, TLI(NNFI) ve SRMR uyum indeksleri incelenmiştir. Ki kare uyum indeksinin ise örneklem büyüklüğü artıkça anlamlı çıkması ve veri setinin çoklu normallliği gerektirmesi nedeniyle kullanılmamasına karar verilmiştir (Şen, 2020; West vd., 2012). Parametre kestirim yöntemi olarak ise MLM (güçlü en çok olabilirlik) kullanılmıştır. DFA sonucunda uyarlaması yapılmış ölçeklerin orijinal faktör yapılarının doğrulandığı, uyum indekslerinin kabul edilebilir veya mükemmel seviyede olduğu görülmüştür. Faktörler arasındaki korelasyon

değerleri incelendiğinde ise ayırt edici geçerliğin sağlandığı belirlenmiştir ($r < ,90$) (Kline, 2016). Ayırt edici geçerliğin sağlanması faktörlerin farklı göstergeler tarafından açıklandığı anlamına gelmektedir (Çokluk vd., 2021).

DFA sonucunda yapı geçerliği kanıtları elde edildikten sonra envantere yer alan ölçeklerin cinsiyete göre ölçme değişmezliği analizleri yapılmıştır. ÇGDFA tekniğini kullanılarak envanterin cinsiyete göre ölçme değişmezliği dört aşamada incelenmiştir (Meredith, 1993). Yapısal değişmezlik aşamasında ölçeklerin kadın ve erkek gruplarında aynı yapıyı temsil edip etmedikleri incelenmiştir. Bu aşamada serbest ve sabit faktör yükleri örüntüsü açısından gruplar arası farklılık yoktur (Önen, 2009). Analiz sonucunda ölçeklerin yapısal değişmezliği sağladıkları, kadın ve erkek gruplarında ölçeklerin aynı faktör yapılarına sahip oldukları sonucuna varılmıştır. Yapısal değişmezliğin sağlanması ile birlikte daha ileri düzey ölçme değişmezliği testleri mümkün hale gelmiştir.

Metrik değişmezlik alanyazınında ölçme birimlerinin eşitliği olarak isimlendirilmektedir (Van de Vijver, 1998 akt. Önen, 2009). Eğer ölçeklerden elde edilen puanlar gruplar arasında karşılaştırma amacıyla kullanılacaksa ölçme aracının birimleri gruplar arasında eşit aralıklı olmalıdır (Önen, 2009). Metrik değişmezlikte ise gruplar arasında faktör yükleri eşit olduğu diğer parametreler ise serbestçe kestirildiği hipotez test edilir. Envantere yer alan ölçeklerin metrik değişmezlik testinde ise öncelikle model-veri uyumu incelenmiş ve uyumun kabul edilebilir ve mükemmel seviyede olduğu görülmüştür. Değişmezlik testinde ise yapısal ve metrik değişmezlikte hesaplanan CFI ve RMSEA değerleri arasındaki fark değerlerine bakılmıştır. Fark değerlerinin Chen'in (2007) belirlediği ölçütlere ($\Delta CFI \leq ,01$ ve $\Delta RMSEA \leq ,015$) uyumlu olduğu görülmüş ve ölçeklerin metrik değişmezliği sağladıkları sonucuna varılmıştır. Metrik değişmezliğin sağlanması ölçeklerde yer alan maddelerin kadın ve erkek gruplarında aynı şekilde anlaşılıp yorumlandığını göstermektedir (Şen, 2020). Metrik değişmezliğin sağlanmasının gruplar arasında yansız ölçümler yapıldığı anlamına gelmeyebileceği ifade edilmektedir (Önen, 2009). Bu nedenle metrik değişmezlik aşamasından sonra ölçek değişmezliği de test edilmelidir.

Ölçek değişmezliğinde faktör yüklerinin eşitliğinin yanı sıra gösterge sabitleri de gruplar arasında eşitlendiği, diğer faktörlerin serbestçe kestirildiği hipotez test edilir. Ölçek değişmezliğinde öncelikle envantere yer alan ölçeklerin model-veri uyumu değerlendirilmiş, uyumun kabul edilebilir ve mükemmel seviyede olduğu görülmüştür. Değişmezlik testinde ise yapısal ve ölçek değişmezliğinde hesaplanan CFI ve RMSEA değerleri arasındaki fark değerlerine bakılmıştır. Fark değerlerinin Chen'in (2007) belirlediği ölçütlere ($\Delta CFI \leq 0,01$ ve $\Delta RMSEA \leq 0,015$) uyumlu olduğu görülmüş ve ölçek değişmezliğinin sağlandığı sonucuna varılmıştır. Ölçek değişmezliğinin sağlanması ölçeklerden elde edilen puanlarla gruplar arası karşılaştırmalar yapılabileceğini ve ölçümlere cinsiyetten kaynaklanan bir yanlılık karışmayacağını göstermektedir.

Ölçme değişmezliğinin son aşamasında ise katı değişmezlik testleri yapılmıştır. Katı değişmezlikte faktör, gösterge ve hata varyanslarının eşitliği hipotezi test edilir. Katı değişmezlik aşamasında envantere yer alan ölçeklerin model veri uyumlarının kabul edilebilir ve mükemmel seviyede oldukları görülmüştür. Katı değişmezliğinin testinde ise yapısal ve katı değişmezlik aşamalarında hesaplanan CFI ve RMSEA değerleri farklarına bakılmış, farkın Chen'in (2007) belirlediği ölçütlerle ($\Delta CFI \leq 0,01$ ve $\Delta RMSEA \leq 0,015$) uyumlu olduğu görülmüştür. Katı değişmezliğin sağlanması ile birlikte envantere yer alan ölçeklerin kullanılarak yapılan ölçümlerde, cinsiyet gruplarına göre geçerli ölçümler elde edilebileceği ifade edilebilir.

Alanyazında değerlendirme okuryazarlığı ile ilgili geliştirilen ölçme araçları (Campbell vd.,2002; Mertler & Campbell, 2005; Gotch & French, 2013; Mertler, 2003; Plake vd., 1993; Tao, 2014) kullanılarak öğretmen adaylarının ve öğretmenlerin değerlendirme okuryazarlığı düzeylerinin; cinsiyet, öğrenim düzeyi, hizmet içi eğitim alma durumu, mesleki kıdem, branş (Akpolat & Genç, 2022; Atasoy, 2022; Erdoğan & Kurt, 2012; Ergül, 2019; Tünkler, 2019) gibi çeşitli değişkenler açısından karşılaştırıldığı görülmektedir. Bu araştırmaların elde edilen bulguların geçerliğinin sağlanması ise kullanılan veri toplama araçlarına ilişkin ölçme değişmezliği analizlerinin yapılması ile mümkündür. Gruplar arası karşılaştırma yapılabilmesi için önkoşul olarak bu ölçme araçlarına ait ölçme değişmezliği analizleri yapılmalı (Horn & Mcardle, 1992) ve geçerlik kanıtları sunulmalıdır. Ölçme değişmezliğinin

sağlanamadığı durumlarda ise gruplar arasında niceliksel karşılaştırmalar yapmak yerine niteliksel yorumların yapılması önerilmektedir (Başusta ve Gelbal, 2015).

Ulusal düzeyde öğretmen ve öğretmen adaylarının değerlendirme okuryazarlığı düzeylerinin ölçülmesinde en sık kullanılan ölçme aracı Bütüner vd. (2010) tarafından Türkçe'ye uyarlanan *Ölçme Değerlendirme Okuryazarlığı Envanteri*'dir. Bu envantere ait ölçme değişmezliği analizlerinin yapıldığı bir çalışmaya ise rastlanılmamıştır. Bu araştırmada ise *Sınıf İçi Değerlendirme Yaklaşımları Envanteri*'nin cinsiyete gruplarında ölçme değişmezliği analizleri yapılmış ve geçerlik kanıtları elde edilerek, cinsiyete göre değerlendirme okuryazarlığının karşılaştırılabilmesi için kullanılabilir bir ölçme aracı alanyazına kazandırılmıştır.

BÖLÜM VI

SONUÇ VE ÖNERİLER

Bu arařtırmada DeLuca vd. (2016b) tarafından geliřtirilen Sınıf İçi Deęerlendirme Yaklařımları Envanterini (*Approaches to Classroom Assessment Inventory, ACAI*) Türkçe'ye uyarlamak ve envanterin cinsiyete göre ölçme deęiřmezlięini incelemek amaçlanmıřtır. Bu amaç doęrultusunda envanterin öncelikle Türkçe'ye uyarlaması yapılmıř ve uyarlama iřleminden sonra ise envanterin cinsiyete göre ölçme deęiřmezlięi analizleri gerçekleştirilmiřtir. Arařtırmada elde edilen sonuçlar maddeler halinde özetlenmiř, uygulayıcılar ve arařtırmacılar için öneriler sunulmuřtur.

6.1. Sonuçlar

1. Envanterin Türkçe'ye çevirisinin yapılmasının ardından dil eřdeęerlięinin sağlanması amacıyla öncelikle uzman görüşlerine bařvurulmuřtur. Uzman görüşleri arasında uyum indeksleri hesaplanmıř, uyumun düşük olduęu maddeler yeniden düzenlenmiřtir. Uzmanların önerisi doęrultusunda envanterin İngilizce ve Türkçe formları arasında anlamsal, deyimsel, deneysel ve kavramsal eřdeęerlik sağlanmıřtır. İngilizce ve Türkçe formlar 15 İngilizce öğretmenine iki hafta arayla uygulanmıř, formlardan elde edilen puanlar arasındaki korelasyon deęerleri Spearman Sıra Farkları Korelasyon Testi ile hesaplanmıřtır. Türkçe ve İngilizce formları arasında korelasyon deęerlerinin pozitif yönde, orta ve yüksek düzeyde iliřkiler bulunduęu görölmüřtür. Yapılan analizler sonucunda envanterin dil eřdeęerlięi sağlanmıřtır.

2. Envanterin deneme uygulaması 26 öğretmenin katılımıyla gerçekleştirilmiştir. Uygulama sonucunda envanterin ortalama cevaplama süresinin 20 dakika olduğu ve maddelerde anlaşılmayan ifadeler bulunmadığı görülmüştür. Deneme uygulamasından sonra ise envanter 405 öğretmene uygulanmıştır.
3. Uygulama sonucunda öncelikle varsayımların analizi gerçekleştirilmiştir. Kayıp değerler, uç değerler, normallik ve çoklu bağıntı varsayımları test edilmiş ve veri setinde gerekli düzenlemeler yapılmıştır. Varsayımların testinden sonra ise ölçek maddelerinin analizi gerçekleştirilmiştir.
4. Envanterde yer alan ölçeklerin güvenirlikleri Cronbach α ve McDonald's ω katsayıları hesaplanarak belirlenmiştir. Analiz sonuçlarına göre ölçümlerin yeterli düzeyde güvenilir ($r \geq,70$) olduğu sonucuna varılmıştır (Büyüköztürk, 2021).
5. Senaryo maddelerinin kategorik olarak yapılandırılmasından dolayı, senaryo maddelerine verilen yanıtların analizinde Pearson Ki kare testi kullanılmıştır. Öğretmenlerin senaryo maddelerine verdikleri cevaplar ile cinsiyet ve mesleki kıdem değişkenleri arasındaki ilişkiler incelenmiştir. Analiz sonuçlarına göre sadece cinsiyet değişkeni ile öğrenirken değerlendirme, standart yaklaşım ve farklılaştırılmış yaklaşım alt boyutlarında anlamlı ilişkilerin olduğu belirlenmiştir ($p <,05$). Mesleki kıdem değişkeni ile temalar arasında ise anlamlı bir ilişki bulunmamıştır ($p >,05$).
6. Envanterde yer alan ölçeklerin orijinal faktör yapılarının doğrulanması amacıyla doğrulayıcı faktör analizleri yapılmıştır. Verilerin çoklu normallik varsayımını sağlamaması nedeniyle uyum indekslerinin hesaplanmasında MLM (güçlü en çok olabilirlik) kestirim yöntemi kullanılmıştır. Model veri uyumunun değerlendirilmesinde ise RMSEA, CFI, TLI(NNFI) ve SRMR uyum indeksleri raporlanmıştır. DFA sonuçlarına göre uyum indekslerinin kabul edilebilir ve mükemmel seviyede olduğu görülmüştür. Dolayısıyla ölçeklerin orijinal faktör yapıları doğrulanmıştır.
7. Uyarılama işlemlerinin ardından ise envanterde yer alan ölçeklerin cinsiyete göre ölçme değişmezliği incelenmiştir. Ölçme değişmezliği ÇGDFA tekniğiyle gerçekleştirilmiştir. Yapısal, metrik, ölçek ve katı değişmezlik

aşamaları sırasıyla takip edilmiştir (Meredith, 1993). Ölçme değişmezliğinin değerlendirilmesinde ise karar ölçütü olarak $\Delta CFI \leq 0,01$ ve $\Delta RMSEA \leq 0,015$ fark değerleri kullanılmıştır (Chen, 2007). Analizler sonucunda ölçeklerin tüm ölçme değişmezliği aşamalarını sağladıkları görülmüştür. Dolayısıyla envanterde yer alan ölçekler kullanılarak kadın ve erkek grupları için yapılacak tüm karşılaştırmalarda güvenilir ve geçerli ölçümler elde edilebileceği sonucuna varılmıştır.

6.2. Öneriler

Araştırma sonuçları doğrultusunda uygulayıcılara ve araştırmacılara dönük çeşitli öneriler getirilmiştir.

Uygulayıcılara Dönük Öneriler

1. Öğretmenlerin değerlendirme okuryazarlıkları profilleri incelenerek, değerlendirme konusunda eksik kaldıkları alanlar belirlenebilir. Öğretmenlerin değerlendirmeye ilişkin mesleki öğrenme öncelikleri ve tercihleri göz önünde bulundurularak, öğretmenlerin değerlendirme okuryazarlıklarının desteklenmesi ve geliştirilmesi hususunda eğitimler planlanabilir.
2. Envanter sonuçları kullanılarak verilecek kararlarda öncelikle örnekleme ait özellikler dikkate alınmalı ve ölçme değişmezliği çalışmaları yapılmalı, elde edilen bulgular doğrultusunda kararlar alınmalıdır.

Araştırmacılara Dönük Öneriler

3. Uyarlanan envantere ilişkin kararlılık boyutunda güvenilirlik analizleri gerçekleştirilebilir.
4. Ülkemizde bir milyonun üzerinde öğretmen olduğu gerçeğinden yola çıkarak, farklı örneklerde araştırma yinelenebilir ve sonuçlar analiz edilebilir. Örneklem büyüklüğünün modele etkisi incelenebilir.
5. Yapılacak araştırmalarda ölçüt geçerliği ile ilgili analizler de gerçekleştirilebilir.

KAYNAKLAR

- Abell, S. K. & Siegel, M. A. (2011). Assessment Literacy: What Science Teachers Need to Know and Be Able to Do. In D. Corrigan, J. Dillon & Richard Gunstone (Eds), *The Professional Knowledge Base of Science Teaching* (pp. 205-221). Springer https://doi.org/10.1007/978-90-481-3927-9_12
- Acar Güvendir, M., & Özer Özkan, Y. (2015). Türkiye'deki Eğitim Alanında Yayımlanan Bilimsel Dergilerde Ölçek Geliştirme ve Uyarlama Konulu Makalelerin İncelenmesi. *Electronic Journal of Social Sciences*, 14(52). <https://doi.org/10.17755/esosder.54872>
- Acar Güvendir, M., & Özer Özkan, Y. (2016). *Ölçme Değerlendirme Dersinin Uygulamadaki Yeri*. 5. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi.
- Acar Güvendir, M. & Özkan Özer, Y. (2021). The role of achievement in the correlation between teacher candidates' measurement and evaluation common competency perceptions and assessment literacy. *Uluslararası İnovatif Eğitim Araştırmacısı*, 1(2), 107-117. <https://doi.org/10.29228/iedres.55000>
- Adamson, C. (2020). New teacher assessment literacy: Determining and narrowing the gaps. *Emerging Perspectives*, 4(2), 89-105.
- Akpolat, T. & Genç, T. (2022). KPSS Kursuna Devam Eden Öğretmen Adaylarının Epistemolojik İnançları İle Ölçme Ve Değerlendirme Okuryazarlığı Arasındaki İlişkiler, *Trakya Eğitim Dergisi*, 12 (2), 647-662 <https://doi.org/10.24315/tred.917531>
- Alkharusi, H. (2011). Psychometric properties of the teacher assessment literacy questionnaire for preservice teachers in Oman. *Procedia-Social and Behavioral Sciences*, 29, 1614-1624. <https://doi.org/10.1016/j.sbspro.2011.11.404>
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association (1990). The Standards for Teacher competence in the educational assessment of students. <http://files.eric.ed.gov/fulltext/ED323186.pdf>
- Aşıcı, M. (2009). Kişisel ve Sosyal Bir Değer Olarak Okuryazarlık. *Değerler Eğitimi Dergisi*, 7(17), 9-26.
- Atasoy, B. (2022). *Kimya Öğretmenlerinin ve Öğretmen Adaylarının Ölçme Değerlendirme Okuryazarlıkları ve Ölçme Değerlendirmeye İlişkin Tutumları* (Yayımlanmamış yüksek lisans tezi). Gazi Üniversitesi.

- Atjonen P., Pöntinen S., Kontkanen, S. & Ruotsalainen, P. (2022). In Enhancing Preservice Teachers' Assessment Literacy: Focus on Knowledge Base, Conceptions of Assessment, and Teacher Learning. *Frontiers in Education*, 7, 1-12. <https://doi.org/10.3389/feduc.2022.891391>
- Azrak, Y. & Yalçınkaya, E. (2019). Investigation of the Assessment Literacy Levels of Social Studies Teacher Candidates in Terms of Various Variables. *Anadolu Journal of Educational Sciences International*, 9(1), 27-55. <https://doi.org/10.18039/ajesi.520816>
- Başusta, N. B. (2010). Ölçme Eşdeğerliği. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 58-64.
- Başusta, N. B. & Gelbal, S. (2015). Gruplararası Karşılaştırmalarda Ölçme Değişmezliğinin Test Edilmesi: PISA Öğrenci Anketi Örneği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90.
- Baykul, Y. (2015). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması* (3. baskı). Pegem Akademi.
- Bollen, K. A. & Long, J. S. (Eds.) (1993). *Testing Structural Equation Models* (First Edition). Sage Publications.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Brown, T. A. & Moore, T. M. (2012). Confirmatory Factor Analysis. In Rick H. Hoyle (Ed), *Handbook of Structural Equation Modeling* (pp. 361-379). The Guilford Press.
- Bütünler, S. Ö., Yiğit N. & Çimer S.O. (2010). Ölçme ve Değerlendirme Okuryazarlığı Envanterinin Türkçeye Uyarlanması. *e-Journal of New World Sciences Academy Education Sciences*, 5(3), 792-809.
- Büyüköztürk, Ş. (2021). *Sosyal Bilimler İçin Veri Analizi El Kitabı* (29. baskı). Pegem Akademi.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Erkan Akgün, Ö., Karadeniz, Ş. & Demirel, F. (2022). *Eğitimde Bilimsel Araştırma Yöntemleri* (32. baskı). Pegem Akademi.
- Byrne, B. M. & Watkins, D. (2003). The Issue Of Measurement Invariance Revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155-175. <https://doi.org/10.1177/0022022102250225>
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002). *Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers*. Paper presented at the MidWestern Educational Research Association.

- Chan, C. K. Y. & Luk, L. Y. Y. (2022) A four-dimensional framework for teacher assessment literacy in holistic competencies. *Assessment & Evaluation in Higher Education*, 47(5), 755-769.
<https://doi.org/10.1080/02602938.2021.1962806>
- 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>
- Clark, D. A. & Donnellan, M. B. (2021). What if apples become oranges? A primer on measurement invariance in repeated measures research. In J. F. Rauthmann (Ed.), *The Handbook of Personality Dynamics and Processes* (pp. 837-854). Academic Press. <https://doi.org/10.1016/B978-0-12-813995-0.00032-7>
- Coble, C. R. & Azordegan, J. M. (2004). The Challenges and Opportunities of the No Child Left behind Act: Seven Strategies for Teacher Educators. *Action in Teacher Education*, 26(2), 2-14.
<https://doi.org/10.1080/01626620.2004.10463318>
- Çambay, Ö. & Kazanç, S. (2018). Farklı Branşlardaki Öğretmen Adaylarının Değerlendirme Okuryazarlığı ve Değerlendirme Okuryazarlığına İlişkin Öz-Yeterliliklerinin İncelenmesi. *Fırat Üniversitesi Eğitim Bilimleri Enstitüsü Dergisi*, 5(3), 140-165. <https://doi.org/10.33907/turkjes.461609>
- Çapık, C., Gözüm, S. & Aksayan, S. (2018). Kültürlerarası ölçek uyarlama aşamaları, dil ve kültür uyarlaması: Güncellenmiş rehber. *FNJN Florence Nightingale Journal of Nursing*, 26(3), 199-210. <https://doi.org/10.26650/FNJN397481>
- Çokluk, Ö., Şekercioğlu, G. & Büyüköztürk, Ş. (2021). *Sosyal Bilimler İçin Çok Değişkenli İstatistik SPSS ve LISREL Uygulamaları* (6. baskı). Pegem Akademi.
- Çüm, S., & Koç, N. (2013). Türkiye’de Psikoloji ve Eğitim Bilimleri Dergilerinde Yayımlanan Ölçek Geliştirme ve Uyarlama Çalışmalarının İncelenmesi. *Journal of Educational Sciences & Practices*, 12(24).
- Davidheiser, S. A. (2013). *Identifying Areas for High School Teacher Development: A Study of Assessment Literacy in the Central Bucks School District* (Doctoral dissertation). Drexel University.
- DeLuca, C. & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates’ learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
<https://doi.org/10.1080/0969594X.2010.516643>

- DeLuca, C., LaPointe-McEwan, D. & Luhanga, U. (2016a). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251-272.
<https://doi.org/10.1007/s11092-015-9233-6>
- DeLuca, C., LaPointe-McEwan, D. & Luhanga, U. (2016b). Approaches to Classroom Assessment Inventory: A New Instrument to Support Teacher Assessment Literacy. *Educational Assessment*, 21(4), 248-266.
<https://doi.org/10.1080/10627197.2016.1236677>
- DeLuca, C. Rickey, N. & Coombs, A. (2021). Exploring assessment across cultures: Teachers' approaches to assessment in the U.S., China, and Canada, *Cogent Education*, 8(1), 1-26. <https://doi.org/10.1080/2331186X.2021.1921903>
- Demirdöğen, B. & Korkut, H. M. (2021). Öğretmen eğitimi fark yaratıyor mu? Eğitim ve fen-edebiyat fakültesi mezunu fen bilimleri öğretmenlerinin ölçme-değerlendirme okuryazarlıklarının karşılaştırılması. *Journal of Qualitative Research in Education*, 26, 23-52. <https://doi.org/10.14689/enad.26.2>
- Deniz, K. Z. (2007). Psikolojik ölçme aracı uyarlama. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 40(1), 1-16.
https://doi.org/10.1501/Egifak_0000000145
- Doğan, N. (Ed.). (2019). *Eğitimde Ölçme ve Değerlendirme* (2. baskı). Pegem Akademi. <https://doi.org/10.14527/9786052415825.01>
- Erdoğdu, M. Y. & Kurt, F. (2012). Öğretmenlerin Ölçme Değerlendirme Yeterlilik Algılarının Bazı Değişkenler Açısından İncelenmesi. *Elektronik Eğitim Bilimleri Dergisi*. 1(2), 23-36.
- Ergül, A. (2019). *Öğretmenlerin Ölçme ve Değerlendirme Okuryazarlık Düzeylerinin Bazı Değişkenler Açısından İncelenmesi*. (Yayımlanmamış yüksek lisans tezi). Hacettepe Üniversitesi.
- Erkuş, A. (2007). Ölçek Geliştirme ve Uyarlama Çalışmalarında Karşılaşılan Sorunlar, *Türk psikoloji Bülteni*, 13(40), 17-25.
- Erkuş, A. & Selvi, H. (2019). *Psikolojide Ölçme ve Ölçek Geliştirme III: Ölçek Uyarlama ve Norm Geliştirme* (1. Baskı). Pegem Akademi.
- Erkuş, A. (2021). *Psikolojide Ölçme ve Ölçek Geliştirme I: Temel Kavramlar ve İşlemler* (5. baskı). Pegem Akademi.
- Everitt, B. S., (1975), Multivariate analysis: the need for data and other problems, *British Journal of Psychiatry*, 126, 237-240.
<https://doi.org/10.1192/bjp.126.3.237>

- Fitzgerald, C. T. (2004). Test adaptation in a large-scale certification program. In R. K. Hambleton, P. F. Meranda & C. D. Spielbergler (Eds), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 207-224). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410611758-14>
- Gareis, C. R., & Grant, L. W. (2015). Assessment Literacy for Teacher Candidates: A Focused Approach. *Teacher Educators' Journal*, 2015, 4-21.
- Gelbal, S. & Kelecioğlu, H. (2007). Öğretmenlerin Ölçme ve Değerlendirme Yöntemleri Hakkındaki Yeterlik Algıları ve Karşılaştıkları Sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33, 135-145.
- George, D. & Mallery, P. (2019). *IBM SPSS Statistics 25 Step by Step A Simple Guide and Reference* (15th Edition). Taylor & Francis Group. <https://doi.org/10.4324/9781351033909>
- Gotch, C. & French, B. (2014). A Systematic Review of Assessment Literacy Measures. *Educational Measurement: Issues and Practice*, 33(2), 14-18. <https://doi.org/10.1111/emip.12030>
- Güçlü, N. & Bayrakçı, M. (2004). Amerika Birleşik Devletleri Eğitim Sistemi ve Hiçbir Çocuğun Eğitimsiz Kalmaması Reformu. *Gazi Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 5(2), 51-64.
- Gül, E. (2011). *İlköğretim Öğretmen Adaylarının Ölçme-Değerlendirme Okuryazarlığı ve Ölçme-Değerlendirmeye İlişkin Tutumlarının Belirlenmesi*. (Yayımlanmamış yüksek lisans tezi). Fırat Üniversitesi.
- Güler, N. (2015). *Eğitimde Ölçme ve Değerlendirme* (8. baskı). Pegem Akademi. <https://doi.org/10.14527/9786053641247>
- Gürsoy, G. (2017). Ölçme ve Değerlendirme Okuryazarlığı: Kavramsal Bir Analiz. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 6(1), 281-316.
- Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Applied Testing Technology Journal*, 1-13.
- Hambleton, R. K., & De Jong, J. H. (2003). Advances in translating and adapting educational and psychological tests. *Language testing*, 20(2), 127-134. <https://doi.org/10.1191/0265532203lt247xx>
- Horn, J. L. & Mcardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research, *Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process*, 18 (3), 117-144. <https://doi.org/10.1080/03610739208253916>

- Hoyle, R. H. & Panter, A. T. (1995). Writing About Structural Equation Models. In R. H. Hoyle (Ed.), *Structural Equating Modeling* (pp. 158-176). Sage Publications.
- 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>
- Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of consumer psychology*, 20(1), 90-98.
<https://doi.org/10.1016/j.jcps.2009.09.003>
- International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests (Second edition)*.
- İlhan, M. (2022). Madde Analizi ve Madde ile Test İstatistikleri Arasındaki İlişki. B. Çetin (Editör), *Eğitimde Ölçme ve Değerlendirme* (2. Basım, ss. 269-293) içinde. Anı Yayıncılık.
- İzci, K., Çalışkan, G. & Oğuz Aktürk, A. (2018). Öğretmen Adaylarının Sınıf İçi Ölçme-Değerlendirme Okuryazarlıklarına Etki Eden Etmenlerin İncelenmesi: Bir Yapısal Eşitlik Modellemesi. *Eğitim, Bilim ve Teknoloji Araştırmaları Dergisi*, 3(1), 10-22.
- Karaman, P. & Şahin, Ç. (2014). Öğretmen Adaylarının Okuryazarlıklarının Belirlenmesi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi (KEFAD)*, 15(2), 175-189.
- Koç, A. & Bulut, S. (2020). Fen Bilgisi Öğretmen Adaylarının Ölçme Değerlendirme Okuryazarlık Düzeylerinin Çeşitli Değişkenler Açısından İncelenmesi. *Gazi Üniversitesi Eğitim Fakültesi Dergisi*, 40(3), 877-904.
<https://doi.org/10.17152/gefad.733749>
- Kılıç, A. F. & Koyuncu, İ. (2017). Ölçek Uyarlama Çalışmalarının Yapı Geçerliği Açısından İncelenmesi. Ö. Demirel & S. Dinçer (Editörler), *Küreselleşen Dünyada Eğitim* (1. Basım, ss. 415-438) içinde. Pegem Akademi.
<https://doi.org/10.14527/9786053188407.27>
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (Fourth Edition). The Guilford Press.
- Klinger, D.A., McDivitt, P.R., Howard, B.B., Munoz, M.A., Rogers, W.T., & Wylie, E.C. (2015). *The Classroom Assessment Standards for PreK-12 Teachers*. Kindle Direct Press.
<http://www.jcsee.org/the-classroom-assessment-standards-new-standards>

- Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22(3), 255-276.
<https://doi.org/10.1080/10476210.2011.593164>
- McMillan, J. H. (2001). *Essential assessment concepts for teachers and administrators* (1st edition). Corwin Press.
- Meade, A. W. & Lautenschlager, G. J. (2004). A Monte-Carlo Study of Confirmatory Factor Analytic Tests of Measurement Equivalence/Invariance. *Structural Equation Modeling*, 11 (1), 60-72.
https://doi.org/10.1207/S15328007SEM1101_5
- MEB (2006). *Temel Eğitimde Destek Projesi (TEDP) "Öğretmen Eğitimi Bileşeni" Öğretmenlik Mesleği Genel Yeterlilikleri*. Ankara.
- MEB (2017). *Öğretmenlik Mesleği Genel Yeterlilikleri*. Ankara.
- Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika*, 58(4), 525-543.
<https://doi.org/10.1007/BF02294825>
- Mertler, C. A. (2003). *Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference?*. Paper presented at the annual meeting of the Mid- Western Educational Research Association.
- Mertler, C. & Campell, C. (2005). *Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the Assessment Literacy Inventory*, American Educational Research Association.
- Mullis, I. V., & Martin, M. O. (Eds). (2019). *PIRLS 2021 Assessment Frameworks*. International Association for the Evaluation of Educational Achievement (IEA). TIMSS & PIRLS International Study Center. Lynch School of Education. Boston College.
- Natır, Z. (2020). *İlköğretim Matematik Öğretmenlerinin Ölçme ve Değerlendirme Okuryazarlıklarına İlişkin Sınıf İçi Uygulamaları* (Yayımlanmamış yüksek lisans tezi). Trabzon Üniversitesi
- No Child Left Behind Act (2002). *Public Law* (No. 107–10). United States Federal Education Legislation.
- Oakland, T., Poortinga, Y. H., Schlegel, J., & Hambleton, R. K. (2001). International Test Commission: Its history, current status, and future directions. *International Journal of Testing*, 1(1), 3-32. https://doi.org/10.1207/S15327574IJT0101_2
- Ogan-Bekiroglu, F. & Suzuk, E. (2014) Pre-service teachers' assessment literacy and its implementation into practice, *The Curriculum Journal*, 25:3, 344-371
<http://dx.doi.org/10.1080/09585176.2014.899916>

- Önen, E. (2009). *Ölçme Değişmezliğinin Yapısal Eşitlik Modelleme Teknikleri İle İncelenmesi*. (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi.
- Özkan, Y. Ö. (2019). Ölçme ve Değerlendirme İle İlgili Temel Kavramlar. B. Çetin (Ed.), *Eğitimde Ölçme ve Değerlendirme* (1. baskı, s. 1-22) içinde. Anı Yayıncılık.
- Özgüven, İ. E. (2022). *Psikolojik Testler* (17. baskı). Nobel Akademi Yayıncılık.
- Öztürk, N. B., Eroğlu, M. G., & Kelecioğlu, H. (2015). Eğitim alanında yapılan ölçek uyarlama makalelerinin incelenmesi. *Eğitim ve Bilim*, 40(178).
- Pastore, S. & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128-138.
<https://doi.org/10.1016/j.tate.2019.05.003>
- Plake, B.S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-24.
- Plake, B. S., Impara, J. C. & Fager, J. J. (1993). Assessment Competencies of Teachers: A National Survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12. <https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>
- Popham, W. (2004). Why Assessment Illiteracy Is Professional Suicide, *Educational Leadership*, 82-83.
- Popham, J. W. (2011). Assessment Literacy Overlooked: A Teacher Educator's Confession. *The Teacher Educator*, 46(4), 265-273.
<https://doi.org/10.1080/08878730.2011.605048>
- Popham, W. J. (2018). *Assessment literacy for educators in a hurry*. ASCD.
- Quilter, S. M. & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator*, 36(2), 115-131. <https://doi.org/10.1080/08878730009555257>
- Seferoğlu, S. S. (2004). Öğretmen yeterlikleri ve mesleki gelişim. *Bilim ve Aklın Aydınlığında Eğitim*, 58, 40-45.
- Stiggins, R. (1991). Assessment Literacy, *The Phi Delta Kappan*, 72(7), 534-539.
- Stiggins, R. J. (1995). Assessment Literacy for the 21st Century. *Phi Delta Kappa International*, 77(3), 238-245.
- Sur, E. (2022). Okuryazarlık Kavramı ve Türkiye'deki Okuryazarlık Araştırmaları Üzerine Bir İnceleme. *Necmettin Erbakan Üniversitesi Ahmet Keleşoğlu Eğitim Fakültesi(AKEF) Dergisi*, 4(2), 445-467. <https://doi.org/10.38151/akef.2022.27>

- Sümer, N. (2000). Yapısal Eşitlik Modelleri: Temel Kavramlar ve Örnek Uygulamalar. *Türk Psikoloji Yazıları*, 3(6), 49-74.
- Şahin, N. (1994). Psikoloji Araştırmalarında Ölçek Kullanımı. *Türk Psikoloji Dergisi*, 9(33), 19-26.
- Şata, M. (2016). Türk Eğitim Sistemi'nde Sınıf İçi ile Geniş Ölçekli Ölçme ve Değerlendirmeye Genel Bir Bakış. *Curr Res Educ*, 2(1), 53-60.
<https://doi.org/10.14527/9786053184010.01>
- Şen, S. (2020). *Mplus İle Yapısal Eşitlik Modellemesi Uygulamaları* (1. baskı). Nobel Akademi.
- Tao, N. (2014). *Development and Validation of Classroom Assessment Literacy Scales: English as a Foreign Language (EFL) Instructors in a Cambodian Higher Education Setting*. (Unpublished doctoral dissertation). Victoria University.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics* (Fifth Edition). Pearson Education.
- Tünkler, V., (2019). Sosyal Bilgiler Öğretmen Adaylarının Değerlendirme Okuryazarlık Düzeyi İle Değerlendirme Öz-Yeterlik Algılarının İncelenmesi, *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 37, 319-332.
<https://doi.org/10.30794/pausbed.525176>
- Uysal, İ. & Kılıç, A. F. (2022). Çok Değişkenli Normallik: Testler Ne Kadar Doğru ve Güçlü?. Nayır, F. ve Poyrazlı, Ş. (Editörler), *Eğitim Bilimlerinde Güncel Araştırmalar* (1. Basım, ss. 19-32) içinde. Anı Yayıncılık.
- Uzun, B. & Öğretmen, T. (2010). Fen Başarısı ile İlgili Bazı Değişkenlerin TIMSS-R Türkiye Örneğinde Cinsiyete Göre Ölçme Değişmezliğinin Değerlendirilmesi. *Eğitim ve Bilim Dergisi*, 35(155), 26-35.
- Uzun, Y. & Çelik, G. (2020). Akademisyenlerin Okuryazarlık Algısındaki Değişimler. *Uluslararası Toplum Araştırmaları Dergisi*, 15(22).
<https://doi.org/10.26466/opus.646592>
- Vanderbeg, R. J. & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational research methods*, 3(1), 4-70.
<https://doi.org/10.1177/109442810031002>
- Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European psychologist*, 1(2), 89.
<https://doi.org/10.1027/1016-9040.1.2.89>

- Volante, L., & Fazio, X. (2007). Exploring Teacher Candidates' Assessment Literacy: Implications for Teacher Education Reform and Professional Development. *Canadian Journal of Education*, 30(3), 749-770.
<https://doi.org/10.2307/20466661>
- Webb, N. (2002). Assessment Literacy in a Standards-Based Education Setting, *WCER Working Paper* No. 2002-4.
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In Rick H. Hoyle(Ed), *Handbook of structural equation modeling* (pp. 209-231). The Guilford Press.
- Wylie, C., & Lyon, C. (2017). Supporting teacher assessment literacy: A proposed sequence of learning. *Teachers College Record*, 1-13.
- Xu, Y. & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization, *Teaching and Teacher Education*, 58, 149-162.
<http://dx.doi.org/10.1016/j.tate.2016.05.010>
- Xu, Y. & T. L. Brown, G. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133-158.
- Xu, Y., & He, L. (2019). How pre-service teachers' conceptions of assessment change over practicum: Implications for teacher assessment literacy. *Frontiers in Education*, 4(145), 1-16. <https://doi.org/10.3389/educ.2019.00145>
- Yalvaç, E. (2016). *Fen Bilgisi Öğretmen Adaylarının Ölçme Değerlendirme Okuryazarlık Düzeyleri ile Bilişsel Stilleri Arasındaki İlişkinin İncelenmesi* (Yayımlanmamış Yüksek Lisans Tezi). Gazi Üniversitesi
- Yandı, A., Köse, İ. A. & Uysal, Ö. (2017). Farklı Yöntemlerle Ölçme Değişmezliğinin İncelenmesi: Pisa 2012 Örneği. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 13 (1), 243-253. <https://doi.org/10.17860/mersinefd.305952>
- Yeşil, M. (2017). Mantık Biliminin Teşekkülü ile İlişkisi Bakımından Herakleitos'un Akış Öğretisi. *Researcher*, 5(4), 532-547. <https://doi.org/10.18301/rss.465>
- Yurdugül, H. (2005). The Congeneric Test Theory and The Congeneric Item Analysis: An Application for Unidimensional Multiple Choice Tests. *Ankara University Journal of Faculty of Educational Sciences*, 38(2), 21-47.
https://doi.org/10.1501/Egifak_0000000116
- Yurdugül, H. (2006). The Comparison of Reliability Coefficients in Parallel, Tau-Equivalent, and Congeneric Measurements. *Ankara University Journal of Faculty of Educational Sciences*, 39(1), 15-37.

EKLER

Ek 1. Uyarılama İçin Alınan İzin Örneği

Yeşim ÖZER ÖZKAN

Wed, Dec 18, 2019 at 9:04 AM

To:

Dear Christopher DeLuca

I'm Yeşim Özer Özkan. I am working as an Assoc. Prof. Dr. in Faculty of Education Department of Educational Measurement and Evaluation at Gaziantep University in Turkey

I would like to use the "Approaches to Classroom Assessment Inventory (ACAI)" that you and your friends developed, in my study. I would like to adapt your survey to Turkish if you have permission.

Thank you in advance for your interest and support.

Kindest regards

Associate. Prof. Dr. Yeşim ÖZER ÖZKAN
Gaziantep University
Faculty of Education
Department of Educational Measurement and Evaluation

*DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018b). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, 25(4), 355–375.

Christopher DeLuca

Wed, Dec 18, 2019 at 12:38 PM

To: Yeşim ÖZER ÖZKAN

Cc: Andrew Coombs

Dear Professor Yeşim ÖZER ÖZKAN,

Thank you for your message and interest in using the ACAI. You are welcome to use the instrument. I've copied my research assistant, Andrew Coombs, to this message who will be able to share with you the background document with further information about its development.

I ask that as you publish results from the instrument that you cite our related work, which can be found at: <https://www.cdeluca.com/teacher-assessment-literacy>

I look forward to reading more about your scholarship and the outcomes of your work with the ACAI.

All the best,
Chris

Christopher DeLuca, PhD, Associate Professor
Chair, AERA Classroom Assessment SIG

Ek 2. Çalışma Grubu

Branş	Cinsiyet	Mesleki Kıdem (Yıl)							Toplam
		1-5	6-10	11-15	16-20	21-25	26-30	31-35	
Almanca	Kadın	1	2	0	0	0	0	0	3
Arapça	Kadın	2	0	0	0	0	0	0	2
	Erkek	1	0	0	0	0	0	0	1
Beden Eğitimi	Kadın	1	0	0	0	0	0	0	1
	Erkek	1	2	1	0	0	0	0	4
Bilişim Teknolojileri	Kadın	0	4	1	0	0	0	0	5
	Erkek	1	4	0	1	0	0	0	6
Biyoloji	Kadın	0	2	0	1	0	0	0	3
	Erkek	0	0	0	0	0	0	1	1
Çocuk Gelişimi	Kadın	0	1	0	0	0	0	0	1
Coğrafya	Kadın	0	0	1	0	0	0	0	1
Din Kültürü ve Ahlak Bilgisi	Kadın	7	4	1	0	0	0	0	12
	Erkek	7	0	0	0	0	0	1	8
Türk Dili ve Edebiyatı	Kadın	4	0	1	0	0	0	0	5
	Erkek	0	0	1	3	0	0	0	4
El Sanatları	Kadın	0	0	0	1	0	0	0	1
Felsefe	Kadın	0	0	0	0	1	0	0	1
Fen Bilimleri	Kadın	7	4	5	4	1	0	1	22
	Erkek	1	3	4	3	1	0	0	12
Fizik	Kadın	0	0	0	0	1	0	0	1
	Erkek	0	0	0	1	0	0	0	1
Fransızca	Kadın	0	1	0	0	0	0	0	1
Görsel Sanatlar	Kadın	1	1	0	1	0	0	0	3
	Erkek	1	0	0	0	0	0	0	1
Grafik Tasarım	Kadın	0	1	0	0	0	0	0	1
İlköğretim	Kadın	11	9	2	0	0	0	0	22
Matematik	Erkek	1	3	3	1	0	0	0	8
İmam Hatip Meslek Dersleri	Kadın	1	0	0	0	0	0	0	1
İngilizce	Kadın	21	9	6	5	1	1	0	43
	Erkek	9	3	0	0	0	1	0	13
Kimya	Kadın	0	0	1	1	0	0	0	2
	Erkek	1	0	0	0	2	0	0	3
Matematik	Kadın	6	6	1	3	0	0	0	16
	Erkek	1	6	4	1	1	1	0	14
Meslek Dersleri	Erkek	0	0	0	0	0	1	0	1
Mobilya ve Tasarım	Kadın	0	0	0	0	1	0	0	1
Moda Tasarım	Kadın	0	1	0	0	0	0	0	1
Motor Teknolojileri	Erkek	0	0	0	0	0	1	0	1
Muhasebe ve Finans	Kadın	0	0	0	1	0	0	0	1
Müzik	Kadın	3	2	0	0	0	0	0	5
Okul Öncesi	Kadın	4	0	3	1	0	0	0	8
	Erkek	0	2	0	0	0	0	0	2
Özel Eğitim	Kadın	2	2	0	0	0	0	0	4
	Erkek	3	0	0	0	0	0	0	3

Ek 2. Devamı

Rehberlik ve Psikolojik Danışmanlık	Kadın	1	0	0	0	0	0	0	1
	Erkek	2	1	0	1	1	0	0	5
Sanat Tasarım	Kadın	1	0	0	0	0	0	0	1
Sınıf Öğretmeni	Kadın	11	11	8	10	10	4	0	54
	Erkek	2	13	14	5	3	2	1	40
Sosyal Bilgiler	Kadın	2	0	0	1	1	0	0	4
	Erkek	1	3	0	2	0	1	0	7
Tarih	Kadın	1	0	0	0	0	0	0	1
	Erkek	0	0	0	1	1	0	0	2
Teknoloji Tasarım	Kadın	0	2	0	1	0	0	0	3
	Erkek	0	1	0	1	1	0	0	3
Türkçe	Kadın	6	7	4	2	1	0	0	20
	Erkek	1	4	6	2	1	0	0	14
		126	114	67	54	28	12	4	405

Ek 3. Ölçek Maddelerine Ait VIF ve Tolerans Değerleri

Madde No	Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği		Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği		Bölüm 3B: Değerlendirmeye İlişkin Mesleki Değerlendirme Tercihleri Ölçeği	
	VIF	Tolerans	VIF	Tolerans	VIF	Tolerans
m1	1,84	,54	2,12	,47	4,04	,25
m2	2,01	,5	4,16	,24	4,01	,24
m3	1,94	,52	4,12	,24	2,67	,37
m4	1,48	,68	3,51	,29	4,19	,24
m5	2,2	,45	2,91	,34	3,42	,29
m6	2,1	,48	1,68	,6	3,82	,26
m7	2,71	,37	1,53	,65	3,22	,31
m8	2,79	,36	1,99	,5	3,56	,28
m9	2,58	,39	2,59	,39	1,85	,54
m10	2,17	,46	2,26	,44	2,25	,44
m11	1,98	,51			2,05	,49
m12	2,15	,47				

Ek 4. Dil Eşdeğerliği Verilerine Ait Normallik Analizi Sonuçları

Bölüm 1: Senaryolar			Cinsiyet			
			Kadın		Erkek	
Temalar	Alt Boyutlar	Form	Çarpıklık	Basıklık	Çarpıklık	Basıklık
Değerlendirme Amaçları	Öğrenmenin Değerlendirilmesi	Türkçe	-1	,73	-,54	-2,94
		İngilizce	,59	-,73	,00	-5,41
	Öğrenme İçin Değerlendirme	Türkçe	,66	-,73	-1,72	3,27
		İngilizce	,25	-1,2	-,23	-3,87
	Öğrenirken Değerlendirme	Türkçe	,06	-1,65	-,48	-1,7
		İngilizce	-,83	-,27	,83	-,04
Değerlendirme Süreçleri	Tasarım	Türkçe	-,18	-,89	,32	-3,03
		İngilizce	,08	-1,28	,00	-4,34
	Kullanım/Puanlandırma	Türkçe	-,76	-1,14	-,75	,34
		İngilizce	-,08	-1,51	,2	-4,86
	İletişim	Türkçe	-,5	-1,3	-,83	-,04
		İngilizce	-,62	-1,31	,46	-,95
Değerlendirme Adilliği	Standart Yaklaşım	Türkçe	1,65	3,25	-,86	-1,29
		İngilizce	1,02	,6	-2	4
	Eşitlikçi Yaklaşım	Türkçe	-,53	-,22	-,48	-2,71
		İngilizce	1,02	-,66	1,77	3,39
	Farklılaştırılmış Yaklaşım	Türkçe	,09	-1,75	,86	-1,29
		İngilizce	-,87	,04	1,77	3,39
Değerlendirme Teorisi	Tutarlı	Türkçe	,7	,96	,42	-,42
		İngilizce	-,09	-,2	-1,44	2,24
	İçeriksel	Türkçe	-,65	-1,1	-,12	-5,29
		İngilizce	-,45	-1,25	-,98	1,7
	Dengeli	Türkçe	-,42	-,28	,86	-1,29
		İngilizce	-,43	-1,61	-1,37	1,5
Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği		Türkçe	,53	-,60	-1,92	3,7
		İngilizce	,1	-1,48	-,56	-1,52
Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği		Türkçe	,55	-,8	-,84	-,13
		İngilizce	-,97	2,65	,81	-,93
Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği		Türkçe	,45	-,47	-1,25	2,35
		İngilizce	,62	-,51	,22	-2,88

Ek 5. Deneme Uygulamasına Ait Betimsel İstatistikler

Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği									
	Ranj	Minimum	Maksimum	Toplam	\bar{x}	S	Varyans	Çarpıklık	Basıklık
m1	2	3	5	112,7	4,33	,73	,53	-,69	-,69
m2	2	3	5	116,4	4,48	,62	,39	-,98	,27
m3	3	2	5	111,1	4,27	,81	,66	-1,13	1,08
m4	4	1	5	99,7	3,83	1,15	1,33	-,82	-,05
m5	3	2	5	115,3	4,44	,68	,47	-1,78	5,2
m6	2	3	5	107,5	4,14	,76	,58	-,29	-1,16
m7	3	2	5	118,6	4,56	,8	,65	-1,96	3,4
m8	2	3	5	116,4	4,48	,69	,47	-1,13	,22
m9	2	3	5	107	4,12	,77	,59	-,2	-1,21
m10	4	1	5	108,7	4,18	1	1,01	-1,7	3,27
m11	3	2	5	108,5	4,17	,78	,61	-,91	,98
m12	3	2	5	107,4	4,13	,99	,98	-1,1	,35
Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği									
	Ranj	Minimum	Maksimum	Toplam	\bar{x}	S	Varyans	Çarpıklık	Basıklık
m1	2	3	5	120	4,62	,64	,41	-1,47	1,19
m2	2	3	5	112	4,31	,68	,46	-,47	-,67
m3	2	3	5	111	4,27	,6	,37	-,17	-,4
m4	2	3	5	108	4,15	,67	,46	-,19	-,65
m5	2	3	5	110	4,23	,59	,35	-,06	-,21
m6	4	1	5	100	3,85	1,22	1,5	-,68	-,6
m7	2	3	5	117	4,5	,58	,34	-,66	-,48
m8	3	2	5	114	4,39	,8	,65	-1,35	1,71
m9	2	3	5	114	4,39	,7	,49	-,7	-,58
m10	3	2	5	111	4,27	,83	,69	-1,02	,7
Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği									
	Ranj	Minimum	Maksimum	Toplam	\bar{x}	S	Varyans	Çarpıklık	Basıklık
m1	4	1	5	91	3,5	,95	,9	-,91	,73
m2	3	2	5	96	3,69	1,05	1,1	-,44	-,9
m3	4	1	5	100	3,85	1,01	1,02	-,94	1,18
m4	3	2	5	97	3,73	,78	,61	-,58	,43
m5	4	1	5	96	3,69	1,01	1,02	-,57	,52
m6	4	1	5	94	3,62	1,06	1,13	-,65	,08
m7	3	2	5	101	3,89	,95	,91	-,66	-,23
m8	3	2	5	99	3,81	0,9	0,8	-,68	,08
m9	3	2	5	102	3,92	,84	,71	-,28	-,58
m10	3	2	5	109,2	4,2	,89	,8	-1,16	1,02
m11	3	2	5	99	3,81	,9	,8	-,68	,08

Deneme uygulamasına ait betimsel istatistikler incelendiğinde bölüm 2 ve 3'deki 5'li likert tipi ölçeklerin ortalama değerlerinin 3,5 ve üzerinde olduğu görülmektedir. Ölçeklere ait çarpıklık ve basıklık katsayıları incelendiğinde ise maddelerin normal dağılım göstermediği görülmektedir (Tabachnick & Fidell, 2007).

Ek 6. Uygulamaya Ait Betimsel İstatistikler

Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği

Madde No	n	Ranj	Minimum	Maksimum	Toplam	\bar{x}	S	Varyans	Çarpıklık	Basıklık
m1	397	4	1	5	1530	3,85	,89	,8	-,52	,11
m2	397	4	1	5	1698	4,28	,76	,57	-,76	,14
m3	397	4	1	5	1679	4,23	,82	,67	-,89	,49
m4	397	4	1	5	1484	3,74	1,15	1,33	-,67	-,35
m5	397	4	1	5	1671	4,21	,82	,67	-,85	,47
m6	397	4	1	5	1599	4,03	,95	,91	-,8	,14
m7	397	4	1	5	1750	4,41	,72	,52	-1,12	1,19
m8	397	4	1	5	1764	4,44	,68	,47	-1,12	1,4
m9	397	4	1	5	1510	3,80	1,06	1,12	-,59	-,3
m10	397	4	1	5	1544	3,89	,99	,98	-,61	-,19
m11	397	4	1	5	1626	4,10	,87	,76	-,83	,52
m12	397	4	1	5	1561	3,93	,98	,95	-,73	,09

Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği

Madde No	n	Ranj	Minimum	Maksimum	Toplam	\bar{x}	S	Varyans	Çarpıklık	Basıklık
m1	398	3	2	5	1696	4,26	,75	,56	-,69	-,2
m2	398	4	1	5	1652	4,15	,81	,65	-,69	,23
m3	398	4	1	5	1650	4,15	,81	,65	-,82	,67
m4	398	4	1	5	1631	4,10	,82	,67	-,66	,19
m5	398	4	1	5	1613	4,05	,85	,73	-,69	,28
m6	398	4	1	5	1542	3,87	,98	,96	-,54	-,35
m7	398	4	1	5	1594	4,01	,97	,93	-,89	,35
m8	398	4	1	5	1699	4,27	,79	,63	-,97	,86
m9	398	4	1	5	1664	4,18	,8	,63	-,73	,17
m10	398	4	1	5	1616	4,06	,9	,8	-,77	,01

Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği

Madde No	n	Ranj	Minimum	Maksimum	Toplam	\bar{x}	S	Varyans	Çarpıklık	Basıklık
m1	398	4	1	5	1437	3,61	1,17	1,38	-,44	-,71
m2	398	4	1	5	1472	3,70	1,14	1,3	-,53	-,63
m3	398	4	1	5	1486	3,73	1,14	1,29	-,67	-,25
m4	398	4	1	5	1439	3,62	1,18	1,4	-,58	-,54
m5	398	4	1	5	1456	3,66	1,13	1,28	-,56	-,44
m6	398	4	1	5	1456	3,66	1,14	1,29	-,58	-,41
m7	398	4	1	5	1559	3,92	1,02	1,04	-,85	,2
m8	398	4	1	5	1488	3,74	1,12	1,25	-,79	,01
m9	398	4	1	5	1545	3,88	1,07	1,16	-,87	,22
m10	398	4	1	5	1634	4,11	1	1	-,1	,28
m11	398	4	1	5	1517	3,81	1,14	1,3	-,7	-,41

Ek 7. Ölçek Maddelerine Ait Korelasyon Değerleri

Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği Maddeleri Korelasyon Değerleri (n=397; p<,001)

	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12
m1	1											
m2	,559	1										
m3	,505	,591	1									
m4	,344	,325	,438	1								
m5	,493	,496	,479	,486	1							
m6	,448	,404	,426	,408	,571	1						
m7	,407	,502	,491	,312	,524	,454	1					
m8	,393	,526	,509	,309	,532	,401	,751	1				
m9	,513	,379	,440	,402	,56	,615	,367	,461	1			
m10	,453	,375	,404	,406	,547	,568	,491	,488	,659	1		
m11	,442	,339	,406	,399	,535	,498	,411	,459	,575	,493	1	
m12	,432	,417	,452	,391	,516	,576	,488	,501	,579	,54	,615	1

Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği Maddeleri Korelasyon Değerleri (n=398; p<,001)

	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
m1	1									
m2	,649	1								
m3	,613	,833	1							
m4	,592	,77	,782	1						
m5	,527	,698	,731	,753	1					
m6	,378	,492	,484	,472	,495	1				
m7	,371	,310	,342	,347	,419	,449	1			
m8	,508	,465	,451	,476	,482	,413	,478	1		
m9	,597	,548	,529	,573	,577	,447	,452	,653	1	
m10	,517	,578	,56	,549	,547	,546	,439	,554	,657	1

Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği Maddeleri Korelasyon Değerleri (n=398; p<,001)

	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11
m1	1										
m2	,848	1									
m3	,575	,586	1								
m4	,681	,658	,745	1							
m5	,598	,618	,710	,760	1						
m6	,618	,655	,676	,793	,796	1					
m7	,568	,581	,451	,584	,516	,569	1				
m8	,595	,594	,512	,637	,582	,627	,794	1			
m9	,486	,465	,555	,601	,555	,571	,422	,555	1		
m10	,464	,458	,365	,496	,415	,464	,623	,58	,435	1	
m11	,410	,438	,413	,45	,444	,45	,548	,58	,417	,661	1

Ek 8. Analizlerde Kullanılan MPlus Syntaxları

Mplus DFA Syntaxları:

Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği

```

TITLE: Guven Olcegi DFA Analizi
DATA: FILE = guven.dat;
VARIABLE: NAMES = case cinsiyet y1-y12; !12 tane sürekli değişken var;
USEVARIABLES y1-y12;
MODEL: DTUG BY y4 y5 y6 y9 y10 y11 y12;
      IADSI BY y1 y2 y3 y7 y8;
      y8 WITH y7;
ANALYSIS: TYPE = GENERAL;
          ESTIMATOR=MLM;
          ITERATIONS = 1000;
          CONVERGENCE = .00005;
PLOT: TYPE = PLOT1 PLOT2 PLOT3;
      OUTLIERS = MAHALANOBIS LOGLIKELIHOOD INFLUENCE
      COOKS;
OUTPUT: SAMPSTAT STDYX MODINDICES(3) RESIDUAL;

```

Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği

```

TITLE: Ogr Oncelik Olcegi DFA Analizi
DATA: FILE = ogrenmeoncelikleri.dat;
VARIABLE: NAMES = case cinsiyet y1-y10; !10 tane sürekli değişken var;
USEVARIABLES y1-y10;
MODEL: DUBI BY y1 y6 y7 y8 y9 y10;
      MDTIU BY y2 y3 y4 y5;
ANALYSIS: TYPE = GENERAL;
          ESTIMATOR=MLM;
          ITERATIONS = 1000;
          CONVERGENCE = .00005;
PLOT: TYPE = PLOT1 PLOT2 PLOT3;
      OUTLIERS = MAHALANOBIS LOGLIKELIHOOD INFLUENCE
      COOKS;
OUTPUT: SAMPSTAT STDYX MODINDICES(3) RESIDUAL;

```

Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği

```

TITLE: Ogr Tercih Olcegi DFA Analizi
DATA: FILE = ogrenmetercih.dat;
VARIABLE: NAMES = case cinsiyet y1-y11; !11 tane sürekli değişken var;
USEVARIABLES y1-y11;
MODEL: CEVRIMICI BY y3 y4 y5 y6 y9;
      YUZYUZE BY y1 y2 y7 y8;
      BIREBIR BY y10 y11;
      y2 WITH y1;
ANALYSIS: TYPE = GENERAL;
          ESTIMATOR=MLM;
          ITERATIONS = 1000;
          CONVERGENCE = .00005;
PLOT: TYPE = PLOT1 PLOT2 PLOT3;
      OUTLIERS = MAHALANOBIS LOGLIKELIHOOD INFLUENCE
      COOKS;
OUTPUT: SAMPSTAT STDYX MODINDICES(3) RESIDUAL;

```

Ek 8. (Devamı)

Mplus Ölçme Değişmezliği Syntaxları:

Bölüm 2: Sınıf İçi Değerlendirmede Güven Ölçeği

<pre> TITLE: Guven Olcegi CGDFA DATA: FILE = guven.dat; VARIABLE: NAMES = case cinsiyet y1-y12; USEVARIABLES = y1-y12; GROUPING = cinsiyet(0 = kadin 1 = erkek); ANALYSIS: ESTIMATOR = MLM; MODEL= CONFIGURAL METRIC SCALAR; MODEL: factor1 BY y4 y5 y6 y9 y10 y11 y12; factor2 BY y1 y2 y3 y7 y8; y8 WITH y7; OUTPUT: SAMPSTAT; </pre>	<pre> TITLE: Guven Olcegi Katı Degismezlik DATA: FILE = guven.dat; VARIABLE: NAMES = case cinsiyet y1-y12; USEVARIABLES cinsiyet y1-y12; GROUPING = cinsiyet(0 = kadin 1 = erkek); ANALYSIS: ESTIMATOR = MLM; MODEL: factor1 BY y4 y5 y6 y9 y10 y11 y12; factor2 BY y1 y2 y3 y7 y8; y8 WITH y7; y4 (1); y5 (2); y6 (3); y9 (4); y10 (5); y11 (6); y12 (7); y1 (8); y2 (9); y3 (10); y7 (11); y8 (12); OUTPUT: SAMPSTAT; </pre>
---	--

Bölüm 3A: Değerlendirmeye İlişkin Mesleki Öğrenme Öncelikleri Ölçeği

<pre> TITLE: Ogrenme Oncelikleri Olcegi CGDFA DATA: FILE = ogrenmeoncelikleri.dat; VARIABLE: NAMES = case cinsiyet y1-y10; USEVARIABLES = y1-y10; GROUPING = cinsiyet(0 = kadin 1 = erkek); ANALYSIS: ESTIMATOR = MLM; MODEL= CONFIGURAL METRIC SCALAR; MODEL: factor1 BY y1 y6 y7 y8 y9 y10; factor2 BY y2 y3 y4 y5; OUTPUT: SAMPSTAT; </pre>	<pre> TITLE: Ogrenme Oncelikleri Olcegi Katı Degismezlik DATA: FILE = ogrenmeoncelikleri.dat; VARIABLE: NAMES = case cinsiyet y1-y10; USEVARIABLES cinsiyet y1-y10; GROUPING = cinsiyet(0 = kadin 1 = erkek); ANALYSIS: ESTIMATOR = MLM; MODEL: factor1 BY y1 y6 y7 y8 y9 y10; factor2 BY y2 y3 y4 y5; y1 (1); y6 (2); y7 (3); y8 (4); y9 (5); y10 (6); y2 (7); y3 (8); y4 (9); y5 (10); OUTPUT: SAMPSTAT; </pre>
--	---

Ek 8. (Devamı)

Bölüm 3B: Değerlendirmeye İlişkin Mesleki Öğrenme Tercihleri Ölçeği

<pre> TITLE: Ogrenme Tercihleri Olcegi CGDFA DATA: FILE = ogrenmetercih.dat; VARIABLE: NAMES = case cinsiyet y1-y11; USEVARIABLES = y1-y11; GROUPING = cinsiyet(0 = kadin 1 = erkek); ANALYSIS: ESTIMATOR = MLM; MODEL= CONFIGURAL METRIC SCALAR; MODEL: factor1 BY y3 y4 y5 y6 y9; factor2 BY y1 y2 y7 y8; factor3 BY y10 y11; y2 WITH y1; OUTPUT: SAMPSTAT; </pre>	<pre> TITLE: Ogrenme Tercihleri Olcegi Katı Degismezlik DATA: FILE = ogrenmetercih.dat; VARIABLE: NAMES = case cinsiyet y1-y11; USEVARIABLES = y1-y11; GROUPING = cinsiyet(0 = kadin 1 = erkek); ANALYSIS: ESTIMATOR = MLM; MODEL: factor1 BY y3 y4 y5 y6 y9; factor2 BY y1 y2 y7 y8; factor3 BY y10 y11; y2 WITH y1; y3 (1); y4 (2); y5 (3); y6 (4); y9 (5); y1 (6); y2 (7); y7 (8); y8 (9); y10 (10); y11 (11); MODEL erkek; OUTPUT: SAMPSTAT; </pre>
--	---

Ek 9. Envanterin Orjinal Formu

APPROACHES TO CLASSROOM ASSESSMENT INVENTORY

PART 1: SCENARIO-BASED QUESTIONS

You will be presented with four scenarios in this section. Each scenario has 12 actions. For each action, please identify how likely you are to do each of the following as part of your overall response to the scenario (1=highly unlikely; 2=unlikely; 3=somewhat unlikely; 4=somewhat likely; 5=likely; 6=highly likely). Please interpret the scenarios in relation to your current teaching context (i.e., grade, school, community); if you teach across contexts, select one context and keep this context in mind when completing the entire survey. There are no right or wrong answers. We are interested in your honest responses to the scenarios. If you do not understand the statement, select 'Do not Know'.

Scenario 1: You give your class a paper-pencil summative unit test with accommodations and modifications for identified learners. Sixteen of the 24 students fail.

As a teacher in this situation, how likely are you to do each of the following actions as part of your overall response to the scenario? If you have experienced this scenario in your teaching context, base your response on the actions you took.

	Highly Unlikely						Highly Likely		Don't Know
	1	2	3	4	5	6			
Action 1. Keep and record the test grade for all students and move on to the next unit.									
Action 2. Disregard the test grade, then reteach and retest the material.									
Action 3. Have students generate a plan to relearn the material.									
Action 4. Analyze test results to determine if failure was the result of poor test design.									
Action 5. Remove test questions that most students struggled with and re-calculate student scores without those questions.									
Action 6. Schedule student conferences to discuss test performance and next steps.									
Action 7. Allow the entire class to retake a similar test.									
Action 8. Allow only students with exceptionalities who struggled with the test to take a new assessment.									
Action 9. Provide a new assessment to every student who struggled with the test.									
Action 10. Identify test questions that most students consistently answered incorrectly, then provide students with new questions to re-test those concepts.									
Action 11. Adjust student grades based on previously collected formative assessment information.									

Ek 9. *Devami*Scenario 1. (*Continue*)

Action 12. Analyze test results to identify test design errors and consider student circumstances for failure then adjust individual student grades accordingly.

If your response to this scenario was not listed above, how would you most likely respond?

Scenario 2: You discover that a student has plagiarized some of their assignment (e.g., an essay, lab report).

	Highly Likely		Don't Know			
	1	2	3	4	5	6
Action 1. Reduce the student's grade.						
Action 2. Have them re-write the plagiarized section in their own words, then re-grade the assignment.						
Action 3. Ask them to reflect on why plagiarism is a problem and what they would do differently next time.						
Action 4. As the teacher, reflect on how the assignment could have been structured differently to deter plagiarism.						
Action 5. Adjust the student grade to reflect the portion of work that was plagiarized.						
Action 6. Discuss with the student the reasons for the plagiarism, severity of plagiarism, and negotiate potential next steps for their learning.						
Action 7. Apply the same consequence you would for other students to ensure all students are treated equally.						
Action 8. Consider if the student has identified learning needs and exceptionalities before determining response to plagiarism.						
Action 9. Discuss why the student plagiarized and agree upon an appropriate alternative assignment.						
Action 10. Apply all aspects of school policy on plagiarism to ensure consistency across all students.						
Action 11. Consider the original aspects of the assignment and the plagiarized text to determine what the student knows and does not appear to know related to content expectations.						
Action 12. Apply the school policy on plagiarism unless there are extenuating circumstances that led to the plagiarism.						

If your response to this scenario was not listed above, how would you most likely respond?

Ek 9. *Devami*

Scenario 3: Out of 28 students in your class, 4 students are classified/ identified with an exceptionality and have an Individual Education Plan (IEP) (i . e., each student requires accommodations but not a modified curriculum) as well as several other unidentified students with differentiated learning needs. You must decide how to accurately assess learning in your class.

	Highly Unlikely						Highly Likely		Don't Know
	1	2	3	4	5	6			
Action 1. Provide all identified students with accommodations on all summative assessments.									
Action 2. Use formative assessments with all of your students based on their individual learning needs to support their learning.									
Action 3. Allow each student to develop a learning plan based on their strengths and goals to support their own learning.									
Action 4. Design a variety of assessment methods to target each learning objective.									
Action 5. Develop and use different rubric criteria to assess identified students.									
Action 6. Explain to students and parents the purpose of accommodations and how they will be implemented to support student learning and assessment.									
Action 7. Grade students using the same assessments and without accommodations.									
Action 8. Assume that accommodated assessments provide an equally valid measure of identified students' learning.									
Action 9. Provide a variety of assessment options and allow students to choose how they will demonstrate their learning.									
Action 10. Use the same scoring rubric for all students.									
Action 11. Consider individual student learning needs when assessing student work.									
Action 12. Use the same scoring rubric for all students but use professional judgment to apply criteria differently based on individual student capability.									

If your response to this scenario was not listed above, how would you most likely respond?

Ek 9. *Devami*

Scenario 4: You are planning a unit for your class.

	Highly Unlikely						Highly Likely Don't Know	
	1	2	3	4	5	6		
Action 1. Start by designing a summative assessment based on curriculum expectations then plan your lessons.								
Action 2. Use formative assessments to guide your subsequent lesson planning.								
Action 3. Co-construct learning goals with students then have each student develop a personal learning plan for the unit.								
Action 4. Design a summative unit assessment based on what was taught to students.								
Action 5. When planning a unit, consider when assessments and feedback will be provided to students.								
Action 6. Discuss assignments and co-construct grading criteria for the unit with your students.								
Action 7. Plan class lessons and assessments that are the same for all students and encompass the curriculum expectations.								
Action 8. Ensure that assessments throughout unit are accommodated for formally identified learners (i.e., learners with exceptionalities).								
Action 9. Give all students a diagnostic assessment at the beginning of the unit and have students use their results to select personalized appropriate learning and assessment activities.								
Action 10. Work with other teachers to develop a reliable summative assessment for the unit.								
Action 11. In designing your assessments, prioritize the alignment between the assessment and the intended learning objectives.								
Action 12. Adapt assessments that have worked well with other students like yours to take into consideration the context and activities of your enacted lessons.								

If your response to this scenario was not listed above, how would you most likely respond?

Ek 9. *Devami*

PART2: CONFIDENCE IN CLASSROOM ASSESSMENT

In Part 2, teachers were asked to self-assess their current confidence level related to each standard on a 5-point scale: 1 (novice), 2 (beginner), 3 (proficient), 4 (competent), and 5 (expert).

	Novice	Beginner	Proficient	Competent	Expert	Don't Know
	1	2	3	4	5	
1. My practices have a clear purpose (e.g., diagnostic, formative, summative) that supports teaching and learning toward curriculum expectations.						
2. My assessment practices align with established curriculum expectations.						
3. I provide adequate student preparation for assessments in terms of resources, time, and learning opportunities.						
4. I communicate purposes and uses of assessment to parents/guardians when appropriate.						
5. I monitor and revise my assessment practice to improve the quality of my instruction.						
6. I am able to use a variety of strategies to analyze test and assessment results at both student and class levels.						
7. I provide timely feedback to students to improve their learning.						
8. I provide useful feedback to students to improve their learning.						
9. My reports are based on a sufficient body of evidence and provide a summary of student learning toward meeting curriculum expectations.						
10. I engage students in monitoring their own learning and using assessment information to develop their learning skills and personalized learning plans.						
11. I have thought deeply about my approach to assessment.						
12. I am able to articulate my personal philosophy of assessment recognizing its alignment and misalignment with assessment policies and theory.						

Ek 9. *Devami*

PART 3 (A): ASSESSMENT PROFESSIONAL LEARNING PRIORITIES

Please prioritize your interest in learning about the following aspects of classroom assessment using the scale provided (1= very low, 2= low, 3= moderate, 4= high, 5= very high).

	Very Low				Very High
	1	2	3	4	5
1. Choosing the appropriate purpose of assessment (e.g., diagnostic, formative, summative) based on instructional goals and assessment					
2. Constructing assessments in alignment with current assessment theory, principles, and practices.					
3. Administering assessments in alignment with current assessment theory, principles, and practices.					
4. Scoring assessments in alignment with current assessment theory, principles, and practices.					
5. Interpreting and using assessment information in alignment with current assessment theory, principles, and practices.					
6. Communicating assessment purposes, processes, and results to students, parents/guardians, and other stakeholders.					
7. Cultivating fair assessment conditions for all learners, with sensitivity to student diversity and exceptional learners.					
8. Disclosing accurate information about assessments. Protecting the rights and privacy of students that are assessed.					
9. Integrating formative assessment (including assessment for and as learning) during instruction to guide next steps in teaching and learning.					
10. Analyzing and using assessment information to guide instructional decisions and support student learning.					

Ek 9. *Devami*

PART 3 (B): ASSESSMENT PROFESSIONAL LEARNING PRIORITIES

Please indicate your preferred mode(s) of educational support for learning about assessment. (Scale: 1 = not preferred, 2 = somewhat preferred, 3 = moderately preferred, 4 = preferred, 5 = highly preferred).

	Not Preferred					Highly Preferred
	1	2	3	4	5	
1. Face-to-face full-length assessment course with a cohort of other practising teachers						
2. Face-to-face short assessment module with a cohort of other practising teachers						
3. Online full-length assessment course working independently						
4. Online full-length assessment course working with cohort of other practising teachers						
5. A series of short online assessment modules offered as independent studies						
6. A series of short online assessment modules working with cohort of practising teachers						
7. Classroom-embedded collaborative learning/inquiry working with colleagues						
8. Classroom-embedded collaborative learning/inquiry working an expert						
9. Webinars						
10. One-on-one mentoring with peer teacher						
11. One-on-one mentoring with a support teacher						

If you would be willing to participate in future research on classroom assessment, please provide your contact information (i.e., name, email) below. Please note this question is optional.

Name:

Email:

We welcome your feedback. If you have any comments about this survey, please provide them here:

.....

.....

.....

Ek. 10 (Devamı)

Eğer, bu senaryoya vereceğiniz cevabınız, yukarıdaki listede yer almıyorsa, bu senaryoya nasıl tepki verirdiniz? Lütfen yazınız.

Senaryo 2: Öğrencilerinizden birinin ödevinde (örn. makale, laboratuvar raporu) intihal yaptığını fark ettiniz. Bu durumla karşılaşan bir öğretmen olarak aşağıdaki davranışları yapma olasılığınız nedir?

	Düşük Olasılıkla					Yüksek Olasılıkla	Fikrim Yok
	1	2	3	4	5	6	
Davranış 1. Öğrencinin notunu düşürürüm.							
Davranış 2. Öğrenciye intihal yaptığı kısmı kendi kelimeleriyle yeniden yazdırır ve ödevde tekrar puan veririm.							
Davranış 3. Öğrenciye intihalin neden bir sorun olduğunu ve bir dahaki sefere neyi farklı yapacağını düşünmesini isterim.							
Davranış 4. Bir öğretmen olarak, intihal yapılmasını önlemek için, ödevi nasıl farklı şekilde yapılandırabileceğimi düşünürüm.							
Davranış 5. Öğrencinin puanını ödevinde intihal yaptığı kısmı yansıtacak şekilde düzenlerim.							
Davranış 6. Öğrenciyle intihalin nedenleri ve ciddiyetiyle ilgili görüşüp öğrenme süreci için sonraki olası adımlar hakkında tartışırım.							
Davranış 7. Bütün öğrencilere eşit davranıldığımı göstermek için öğrenciye intihal yapmanın sonuçlarını uygulardım.							
Davranış 8. Genel intihal ilkelerini uygulayıp uygulamamaya karar vermeden önce, öğrencinin öğrenmeyle ilgili ihtiyaçları ve kendisine özgü durumlarını (öğrenme güçlüğü vb.) göz önünde bulundururum.							
Davranış 9. Öğrenciyle neden intihal yaptığını görüşerek uygun bir alternatif ödevde karar veririm.							
Davranış 10. Bütün öğrencilere tutarlı davranıldığından emin olmak için intihalle ilgili okul politikalarını öğrenciye uygulardım.							
Davranış 11. Teslim edilen ödevin orijinal ve intihal olan kısımlarına ayrı ayrı odaklanarak öğrencinin ödevin kapsamıyla ilgili neler bilip bilmediğini öğrenmeye çalışırım.							
Davranış 12. Öğrencinin intihal yapmasına neden olan hafifletici nedenler yoksa intihalle ilgili disiplin yönetmeliğini uygulardım.							

Eğer, bu senaryoya vereceğiniz cevabınız, yukarıdaki listede yer almıyorsa, bu senaryoya nasıl tepki verirdiniz? Lütfen yazınız.

Ek. 10 (Devamı)

Senaryo 3: Sınıfınızdaki 28 öğrenci içinden 4'ü özel eğitim öğrencisi olarak belirlendi ve bu öğrencilerin her biri için hazırlanmış bireyselleştirilmiş eğitim planı (BEP) mevcut. Bu öğrencilerin dışında sınıfınızda başka eğitsel ihtiyaçları olan fakat henüz özel eğitim öğrencisi olarak tanımlanmamış öğrenciler de var. Sınıfınızda öğrencilerinizin öğrenimlerini nasıl doğru bir şekilde ölçebileceğinize karar vermeniz gerekiyor. Bu durumla karşılaşan, bir öğretmen olarak, aşağıdaki davranışları yapma olasılığınız nedir?

	Düşük					Yüksek Olasılıkla	Fikrim Yok
	1	2	3	4	5		
Davranış 1. Özel eğitim öğrencisi olarak tanımlanan dört öğrenciye tüm sonuç odaklı değerlendirmeler için kolaylık sağlıyorum.							
Davranış 2. Tüm öğrencilerimin bireysel ihtiyaçlarına göre biçimlendirici değerlendirme yapar, öğrenmelerini desteklerim.							
Davranış 3. Her bir öğrencim için onun güçlü yanları, öğrenim ihtiyaçları ve öğrenim hedefleri doğrultusunda kişisel öğrenme planı hazırlarım.							
Davranış 4. Her bir kazanımın ölçülmesini hedefleyen farklı değerlendirme yöntemlerini tasarlarım.							
Davranış 5. Özel eğitime ihtiyaç duyan dört öğrencimi değerlendirmek için farklı puanlama ölçütleri geliştirir ve kullanırım.							
Davranış 6. Özel eğitim öğrencisi olarak tanımlanmasının amaçlarını, bu düzenlemelerin öğrencilerin eğitimlerini desteklemek ve değerlendirmek amacıyla nasıl uygulanacağını öğrencilerime ve velilere açıklarım.							
Davranış 7. Özel eğitim öğrencisi olarak tanımlanmış olmayı dikkate almadan, bütün öğrencilerimi aynı değerlendirme araçları kullanarak notlandırırım.							
Davranış 8. BEP'li öğrenciler için hazırlanan değerlendirme araçlarının, bu öğrencilerin öğrenmelerinin geçerli bir ölçümünü sağladığını varsayarım.							
Davranış 9. Öğrencilere farklı değerlendirme seçenekleri sunar ve onların öğrenmelerini gösterebilecekleri aracı seçmelerine izin veririm.							
Davranış 10. Tüm öğrenciler için aynı puanlama anahtarını kullanırım.							
Davranış 11. Öğrencilerimin çalışmalarını değerlendirirken onların bireysel öğrenme ihtiyaçlarını göz önünde bulundururum.							
Davranış 12. Tüm öğrenciler için aynı yönergeyi kullanırım fakat özel eğitim öğrencileri için puanlama anahtarındaki ölçütleri onların özel durumlarını dikkate alarak kullanırım.							

Eğer, bu senaryoya vereceğiniz cevabınız, yukarıdaki listede yer almıyorsa, bu senaryoya nasıl tepki verirdiniz? Lütfen yazınız.

Ek. 10 (Devamı)

Senaryo 4: Sınıfınız için bir ünitenin planını yapıyorsunuz. Böyle bir durumda, bir öğretmen olarak, aşağıdaki davranışları yapma olasılığınız nedir?

	Düşük Olasılıkla	1	2	3	4	5	6	Yüksek Olasılıkla	Fikrim Yok
Davranış 1. Dönem sonu değerlendirmesini öğretim programı standartlarına göre tasarlamaya başlayarak derslerimi planlarım.									
Davranış 2. Bir sonraki dersimi planlamama rehberlik etmesi için biçimlendirmeye yönelik değerlendirmeleri kullanırım.									
Davranış 3. Öğrenim hedeflerini öğrencilerimle birlikte belirleyerek sonrasında her bir öğrencime ünite için kişisel bir öğrenme planı oluştururum.									
Davranış 4. Öğrencilerime derste öğrettiklerimi içeren ünite sonu not vermeye yönelik bir değerlendirme tasarlarım.									
Davranış 5. Bir üniteyi planlarken öğrencilerime değerlendirme ve geri bildirimlerin ne zaman verilmesi gerektiğini göz önünde bulundururum.									
Davranış 6. Verilen ödevleri ve ünite sonu puanlandırma ölçütünü öğrencilerimle tartışırım.									
Davranış 7. Tüm öğrenciler için aynı olan ve öğretim programının beklentilerini karşılayan sınıf dersleri ve değerlendirmeler tasarlarım.									
Davranış 8. Ünite boyunca kullandığım değerlendirme araçlarının özel eğitim tanısı almış öğrencilere uygunluğundan emin olurum.									
Davranış 9. Ünitenin başlangıcında tüm öğrencilerime bir ön-değerlendirme uygulayım ve öğrencilerimin test sonuçlarını kullanarak kendilerine uygun kişisel öğrenme ve değerlendirme faaliyetlerini seçmelerine olanak sağlarım.									
Davranış 10. Güvenirli ünite sonu değerlendirmeleri geliştirmek için diğer öğretmenlerle birlikte çalışırım.									
Davranış 11. Değerlendirmelerimi tasarlarken öğretilmesi hedeflenen kazanımlar ile değerlendirmelerimin uyumunu ön planda tutarım.									
Davranış 12. Öğrencilerime benzer diğer öğrencilerde iyi sonuç veren değerlendirmeleri, derslerimin bağlamını ve etkinliklerini dikkate alacak şekilde uyarlarım.									

Eğer, bu senaryoya vereceğiniz cevabınız, yukarıdaki listede yer almıyorsa, bu senaryoya nasıl tepki verirdiniz? Lütfen yazınız.

Ek. 10 (Devamı)

BÖLÜM 2. SINIF İÇİ DEĞERLENDİRMEDE GÜVEN

Bu bölüm, öğretmenlerin çağdaş sınıf değerlendirme uygulamalarıyla ilgili güvenini belirlemek için tasarlanmıştır. Lütfen size sunulan ölçek seçeneklerini dikkate alarak, aşağıdaki değerlendirme uygulamalarına ilişkin, beceri düzeyinizi kendiniz değerlendirin.

1=Henüz çok acemiyim
2=Temel düzeyde gerçekleştirebilirim
3=Tecrübeliyim.
4=Artık bu konuda yetkinim.
5=Bu konuda uzmanım.

Bölüm 2: Sınıf İçi Değerlendirmede Güven	Sizi ne kadar tanımlıyor? (1= En düşük; 5=En yüksek)					Fikrim yok
	1	2	3	4	5	
1. Değerlendirme uygulamalarımın öğretme ve öğrenmeyi destekleyen net bir amacı (biçimlendirici, ünite sonu, tanılayıcı vb.) vardır.						
2. Değerlendirme uygulamalarım öğretim programıyla uyumludur.						
3. Ders kaynakları zaman ve öğrenme fırsatları bakımından öğrencilerimin yaptığım değerlendirmelere yeterli hazırlık yapmalarını sağlarım.						
4. Uygun şartlar olduğunda velilerle sınıfta yaptığım değerlendirmelerin kullanımı ve amaçları ile ilgili görüşürüm.						
5. Öğretim yöntemlerimi geliştirmek için sınıfta yaptığım değerlendirmeleri izler ve gözden geçiririm.						
6. Yaptığım sınav ve değerlendirme sonuçlarını hem öğrenci hem de sınıf düzeyinde analiz etmek için çeşitli stratejiler kullanabilirim.						
7. Öğrencilerimin öğrenmelerini geliştirmek için onlara anlık geribildirimler veririm.						
8. Öğrencilerimin öğrenmelerini geliştirmek için onlara faydalı geribildirimler veririm.						
9. Değerlendirme raporlarım yeterli kanıta dayalıdır ve hem öğretim programının beklentilerinin karşılar hem de öğrenci öğrenmelerinin bir özetini sunar.						
10. Öğrencilerimin yaptığım değerlendirme bilgilerini kullanarak kendi öğrenmelerini denetlemesini, kendi yeteneklerini ve kişisel öğrenme planlarını geliştirmelerini sağlarım.						
11. Sınıftaki değerlendirme yaklaşımım hakkında derinlemesine düşünürüm.						
12. Sınıfta yaptığım değerlendirmelerin, değerlendirme ilkeleri ve teorisi ile uyumu ya da uyumsuzluğunu fark ederek kendi değerlendirme felsefemi oluşturabilirim.						

Ek. 10 (Devamı)

BÖLÜM 3. DEĞERLENDİRMEYE DÖNÜK MESLEKİ ÖĞRENME ÖNCELİKLERİ VE DEĞERLENDİRME TERCİHLERİ

Envanterin bu bölümü öğretmenlerin mesleki öğrenme önceliklerini ve değerlendirmedeki tercihlerini belirlemek amacıyla tasarlanmıştır. Bu bölümün birinci kısmı **Değerlendirmeye Dönük Mesleki Öğrenme Öncelikleridir**. Öğretmenlerden, 1'den (çok düşük) 5'e (çok yüksek) kadar Likert tipi bir ölçek kullanarak sınıf değerlendirmesinin tanımlanmış yönleri hakkında öğrenmeye olan ilgilerine öncelik vermeleri istenmektedir.

Bölüm 3 (A): Değerlendirmeye Dönük Mesleki Öğrenme Öncelikleri

	Ne derece yaparsınız?				
	1	2	3	4	5
1. Öğretim hedeflerine ve değerlendirmeye uygun değerlendirme amacı (tanılayıcı, biçimlendirici, dönem sonu vb.) seçerim.					
2. Güncel değerlendirme teorisi, ilkeleri ve uygulamalarıyla uyumlu olacak şekilde değerlendirmeler oluştururum.					
3. Güncel değerlendirme teorisi, ilkeleri ve uygulamalarıyla uyumlu olacak şekilde değerlendirmeler uygularım.					
4. Güncel değerlendirme teorisi, ilkeleri ve uygulamalarıyla uyumlu olacak şekilde değerlendirmeleri puanlarım.					
5. Güncel değerlendirme teorisi, ilkeleri ve uygulamalarıyla uyumlu olacak şekilde değerlendirmelerden elde edilen bilgileri kullanır ve yorumlarım.					
6. Değerlendirmenin amaçları, süreçleri ve sonuçları ile ilgili öğrencilerle, velilerle ve diğer ilgili paydaşlarla görüşürüm.					
7. Öğrencilerin çeşitliliğini ve özel eğitime ihtiyaç öğrencileri de gözetenerek, bütün öğrenciler için uygun değerlendirme koşulları geliştiririm.					
8. Değerlendirmeye tabi tutulan öğrencilerin özel bilgilerini ve haklarını koruyarak, değerlendirmeler ilgili doğru bilgileri veririm.					
9. Öğretim esnasında biçimlendirici değerlendirme bilgilerini öğretimin sonraki aşamalarına rehberlik etmesi için kullanırım.					
10. Öğrencilerin öğrenmesinin desteklenmesi ve öğretimsel kararlara rehberlik etmesi amacıyla değerlendirme verilerini analiz eder ve kullanırım.					

Ek. 10 (Devamı)

İkinci kısım ise Değerlendirmeye Dönük Mesleki Öğrenme Tercihleri'dir. Öğretmenlerden 1'den (tercih edilmeyen) 5'e (çok tercih edilen) kadar Likert tipi bir ölçek kullanarak tercih ettikleri mesleki öğrenme biçimlerini (yüz yüze grup öğrenmesi, çevrimiçi öğrenme, bireysel öğrenme vb.) belirtmeleri istenmiştir.

Bölüm 3 (B): Değerlendirmeye Dönük Mesleki Öğrenme Tercihleri

	Tercih Etme Düzeyiniz				
	1	2	3	4	5
1. Alanda çalışan diğer öğretmenlerle birlikte yüz yüze tam zamanlı değerlendirme kurslarına katılıyorum.					
2. Alanda çalışan diğer öğretmenlerle birlikte yüz yüze kısa zamanlı değerlendirme modüllerine katılıyorum.					
3. Bağımsız bir şekilde çalışabileceğim tam zamanlı çevrimiçi değerlendirme kurslarına katılıyorum.					
4. Alanda çalışan diğer öğretmenlerle birlikte tam zamanlı çevrimiçi değerlendirme kurslarına katılıyorum.					
5. Bağımsız çalışmalar tarafından sunulan bir dizi çevrimiçi kısa değerlendirme modüllerine katılıyorum.					
6. Alanda çalışan diğer öğretmenlerle birlikte bir dizi çevrimiçi kısa modüllere katılıyorum.					
7. Meslektaşlarla birlikte sınıfla bütünlük işbirliğine dayalı öğrenmeyle/ sorgulayarak öğrenirim.					
8. Uzmanla birlikte sınıfla bütünlük işbirliğine dayalı öğrenmeyle/ sorgulamayla öğrenirim.					
9. Webinarlar (internet seminerleri, Zoom vb.) ile öğrenirim.					
10. Akran öğretmenlerle bire bir danışma yoluyla öğrenirim.					
11. Yardımcı öğretmenle bire bir danışma yoluyla öğrenirim.					

Sınıf değerlendirmeleriyle ilgili gelecekte yürütülebilecek araştırmalara katılmak isterseniz lütfen iletişim bilgilerinizi (isim, soy isim) aşağıdaki form aracılığıyla bizimle paylaşın. Bu katılım tamamen isteğe bağlıdır.

İsim :

Soy isim :

Geri bildirimleriniz bizim için çok önemlidir. Bu envanter ile ilgili yorumlarınız varsa lütfen bizimle paylaşın:



SINIF İÇİ DEĞERLENDİRME YAKLAŞIMLARI ENVANTERİ KULLANICI REHBERİ

Lütfullah DİNÇER
İngilizce Öğretmeni

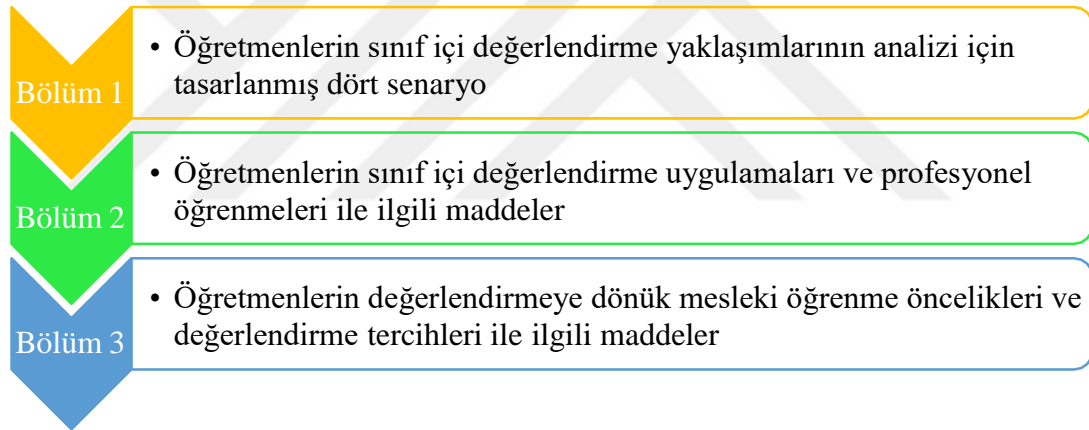
Doç. Dr. Yeşim Özer ÖZKAN
Ölçme ve Değerlendirme A.B.D. Öğretim Üyesi
Gaziantep Üniversitesi

Ek. 11 (Devamı)

SINIF İÇİ DEĞERLEDİRME YAKLAŞIMLARI ENVANTERİ

Sınıf İçi Değerlendirme Yaklaşımları Envanteri öğretmenlerin değerlendirme okuryazarlıklarının ölçülmesi amacıyla Deluca, LaPointe-McEwan ve Luhanga (2016b) tarafından geliştirilmiştir. Envanterin Türkçe'ye uyarlaması ise Dinçer ve Özer Özkan (2023) tarafından yapılmıştır. Envanter www.degerledirmeprofilim.com.tr sitesi üzerinden kullanıcıların çevrimiçi erişimine açıktır. Envanterin cevaplanmasıyla birlikte öğretmenlerin kişisel değerlendirme profilleri oluşturulmaktadır. Oluşturulan profillerde öğretmenlerin sınıf içi değerlendirme yaklaşımları betimlenmektedir. Böylece öğretmenlerin güncel değerlendirme uygulamaları hakkındaki bilgileri analiz edilebilir, öğrenme hedefleri belirlenebilir ve ayrıca öğretmenlerin mesleki gelişimlerine yönelik çalışmalar yapılabilir. Araştırmacılar profil analizleri sonuçlarını kullanarak öğretmenlerin değerlendirme okuryazarlıkları hakkında bilgi edinilebilir ve ölçme değerlendirme eğitimi üzerine çalışmalar gerçekleştirebilir.

Envanterin cevaplanma süresi yaklaşık 20 dakikadır. Envanter maddelerinin doğru ya da yanlış cevabı bulunmamakta, öğretmenlerin kişisel tercihlerini ve önceliklerini yansıtmaktadır. Envanter toplam üç bölümden oluşmaktadır.



Envanterin bölümleri altı farklı bölgede (Birleşik Devletler, Kanada, Birleşik Krallık, Avrupa, Avusturalya ve Yeni Zelanda) tanımlanan ölçme değerlendirme standartları ve Sınıf Değerlendirme Standartları (Klinger ve ark., 2015) incelenerek geliştirilmiştir (Deluca, ve ark., 2016a). İncelenen ölçme değerlendirme standartları yazarlar tarafından *Değerlendirme Amaçları*, *Değerlendirme Süreçleri*, *Değerlendirme Adilliği*, *Değerlendirme Teorisi*, *Değerlendirme Eğitimi ve Öğretmelerin Desteklenmesi* temaları altında yapılandırılmıştır. Envanterin birinci ve ikinci bölümleri *Değerlendirme Amaçları*, *Değerlendirme Süreçleri*, *Değerlendirme Adilliği*, *Değerlendirme Teorisi* temaları; üçüncü bölümü ise *Değerlendirme Eğitimi ve Öğretmelerin Desteklenmesi* teması çerçevesinde hazırlanan ölçme araçlarından oluşturulmuştur. Envanterin birinci ve ikinci bölümlerine ait temalara ilişkin bilgiler Tablo 1'de verilmiştir.

Ek. 11 (Devamı)

Tablo 1. *Değerlendirme Temaları ve Alt Boyutları (Deluca, Rickey ve Coombs, 2021, s.10)*

Değerlendirme Temaları ve Alt Boyutlar	Açıklamalar
Değerlendirme Amaçları	
Öğrenmenin Değerlendirilmesi (AoL)	Öğrenci öğrenmelerinin özetlenmesi ve notlandırılması amacıyla değerlendirilme yapılmasıdır.
Öğrenme İçin Değerlendirme (AfL)	Biçimlendirici değerlendirme yöntemlerinin kullanılarak eğitsel faaliyetlerin sonraki aşamalarına rehberlik edilmesidir. Öğretmen tarafından yönlendirilen öğrenci merkezli bir yaklaşımdır.
Öğrenirken Değerlendirme (AaL)	Öğrencilerin üstbilişsel yeteneklerini ve öğrenme becerilerini (öz Değerlendirme, hedef belirleme, öğrenme planları oluşturma vb.) geliştirmek için onlara geribildirim verilmesini amaçlayan, temelde öğrenci-merkezli fakat öğretmenin de sürece dahil olduğu yaklaşımdır.
Değerlendirme Süreçleri	
Tasarım	Öğrenme hedefleriyle ilişkili olarak öğrencinin öğrenmesini ölçen güvenilir değerlendirmelerin ve ölçme araçlarının geliştirilmesidir.
Kullanım/Puanlandırma (K/P)	Değerlendirme senaryolarının yanıtlanmasında puanlandırma anahtarlarının düzenlenmesi ve kullanımına odaklanan alt boyuttur.
İletişim	Öğrenciler ve velilerle iletişim yoluyla değerlendirme sonuçlarının ve geri bildirim yorumlanmasına odaklanan alt boyuttur.
Değerlendirme Adilliği	
Standart Yaklaşım (SY)	Tüm öğrenciler için eşit değerlendirme süreçlerinin yürütülmesidir.
Eşitlikçi Yaklaşım (EY)	Özel eğitime ihtiyaç duyan öğrenciler için (BEP'li vb.) değerlendirme süreçlerinin yapılandırılmasıdır.
Farklılaştırılmış Yaklaşım (FY)	Her bir öğrencinin kendisine özgü öğrenme ihtiyaçlarının ve hedeflerinin göz önünde bulundurularak öğrenme fırsatlarının ve değerlendirmelerin bireyselleştirilmesidir.
Değerlendirme Teorisi	
Tutarlı	Farklı öğretmenler tarafından ya da farklı zaman aralıklarında yapılan değerlendirmeler arasında tutarlılığın sağlanmaya çalışılmasıdır.
İçeriksel	Ölçme ve değerlendirmenin amacına ve ölçülmesi hedeflenen kazanımlara uygun yapıldığından emin olunması, sonuçlarının analizinde geçerliliğin sağlanmasıdır.
Dengeli	Ölçme ve değerlendirmenin amacına, ölçülmesi hedeflenen kazanımlara uygunluğuna ve ne derecede gerçekleştirilebildiğine odaklanması, ölçme ve değerlendirmede tutarlılığın sağlanmaya çalışılmasıdır.

Bölüm 1'de yer alan senaryo maddeleri temalara göre kategorik olarak gruplandırılmıştır. Senaryo maddelerinin temalara göre dağılımları Tablo 2'de verilmiştir.

Ek. 11 (Devamı)

Tablo 2. Senaryo Maddelerin Temalara Göre Dağılımı

	Madde No	Değerlendirme Amaçları			Değerlendirme Süreçleri			Değerlendirme Adillliği			Değerlendirme Teorisi		
		AoL	AfL	AaL	Tasarım	K/P	İletişim	SY	EY	FY	Tutarlı	İçeriksel	Dengeli
Senaryo 1	1	☒											
	2		☒										
	3			☒									
	4				☒								
	5					☒							
	6						☒						
	7							☒					
	8								☒				
	9									☒			
	10										☒		
	11											☒	
	12												☒
Senaryo 2	1	☒											
	2		☒										
	3			☒									
	4				☒								
	5					☒							
	6						☒						
	7							☒					
	8								☒				
	9									☒			
	10										☒		
	11											☒	
	12												☒
Senaryo 3	1	☒											
	2		☒										
	3			☒									
	4				☒								
	5					☒							
	6						☒						
	7							☒					
	8								☒				
	9									☒			
	10										☒		
	11											☒	
	12												☒
Senaryo 4	1	☒											
	2		☒										
	3			☒									
	4				☒								
	5					☒							
	6						☒						
	7							☒					
	8								☒				
	9									☒			

Ek. 11 (*Devamı*)

$$4 / (4 + 2.5 + 4) = 0,381 \text{ \%}38,1$$

AfL için => AfL(ilk hesaplama) / AoL(ilk hesaplama) + AfL (ilk hesaplama) + AaL(ilk hesaplama)

$$2.5 / (4 + 2.5 + 4) = 0,238 \text{ \%}23,8$$

AaL için => AaL(ilk hesaplama) / AoL(ilk hesaplama) + AfL (ilk hesaplama) + AaL(ilk hesaplama)

$$4 / (4 + 2.5 + 4) = 0,381 \text{ \%}38,1$$

Her bir temada yer alt boyutlar için hesaplanan yüzdesel oranlar kişisel değerlendirme profilinin analizinde ve raporlanmasında kullanılmaktadır. Detaylı analizler ve raporlamalar için www.degerlendirmeprofilim.com.tr sitesinde envanter çevrimiçi doldurulabilir ve profil analizleri incelenebilir.

Ek 12. Etik Kurul Onay Yazısı

Evrak Tarih ve Sayısı: 20.07.2022-211153



T.C.
GAZİANTEP ÜNİVERSİTESİ REKTÖRLÜĞÜ
Sosyal ve Beşeri Bilimler Etik Kurulu

Sayı : E-87841438-604.01.01-211153
Konu : Sosyal ve Beşeri Bilimler Etik
Kurulu'nun 05.07.2022 Tarihli Kararları

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Üniversitemiz Sosyal ve Beşeri Bilimler Etik Kurulu'nun 05.07.2022 tarih ve 08 nolu toplantısında alınan 4, 5, 6, 7, 8, 9, 10, 11 ve 12 sayılı kararlar ekte sunulmuştur. Gereğini arz ve rica ederim.

5950

5950

Prof.Dr. Atilla Ahmet UĞUR
Sosyal ve Beşeri Bilimler Etik Kurulu Başkanı V.

Ek: Sosyal ve Beşeri Bilimler Etik Kurulu'nun 05.07.2022 Tarihli Kararları

GAZİANTEP ÜNİVERSİTESİ
SOSYAL VE BEŞERİ BİLİMLER ETİK KURULU TOPLANTI TUTANAĞI

Toplantı No : 08
Toplantı Tarihi : 05.07.2022
Toplantı Saati : 11:00

Sosyal ve Beşeri Bilimler Etik Kurulu 05.07.2022 tarihinde toplanarak yapılan başvuruları değerlendirdi ve aşağıdaki kararları aldı:

- Eğitim Bilimleri Enstitüsü'nün 16.06.2022 tarih, 196567 sayılı ve "Bilimsel ve Eğitim Amaçlı (Lütfullah DİNÇER)" konulu yazısı incelenmiş olup Üniversitemiz Gaziantep Eğitim Fakültesi Öğretim Üyesi Doç. Dr. Yeşim ÖZER ÖZKAN'ın Eğitim Bilimleri Enstitüsü'nde danışmanlığını yürüttüğü Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı Tezli Yüksek Lisans Programı öğrencisi Lütfullah DİNÇER'in "Sınıf İçi Değerlendirme Yaklaşımları Envanteri'nin Türkçe'ye Uyarlanması ve Ölçme Değişmezliğinin İncelenmesi" başlıklı yüksek lisans tez çalışmasının, Üniversitemiz Sosyal ve Beşeri Bilimler Etik Kurulu'nca değerlendirilmesi istenmektedir. Kurula yapılan başvuru; çalışmanın amacı, yöntemi, veri kaynakları ve veri toplama araçları açısından değerlendirilmiştir. Kurulumuza beyan edilen belgelere dayalı olarak yapılan incelemeler sonucunda, başvuruya ilişkin etik aykırılık tespit edilmemiş olup adı geçen öğrencinin söz konusu yüksek lisans tez çalışmasını yapabilmesinin uygun görülmesine;

ÖZGEÇMİŞ

Lütfullah Dinçer ilköğrenimini Kahramanmaraş'ta, ortaöğrenimini ise İstanbul'da tamamlamıştır. Akdeniz Üniversitesi İngilizce Öğretmenliği programından ise 2019 yılında mezun olmuştur. Lütfullah Dinçer ileri düzeyde İngilizce bilmektedir. 2020 yılından bu yana Kahramanmaraş'ta İngilizce öğretmeni olarak çalışmaktadır.

VITAE

Lütfullah Dinçer completed his primary education in Kahramanmaraş and his secondary education in Istanbul. He graduated from Akdeniz University English Language Teaching program in 2019. Lütfullah Dinçer has an advanced level of English. He has been working as an English teacher at Kahramanmraş since 2020.