

**TÜRKİYE İÇİN PROGRAM DEĞERLENDİRME STANDARTLARI
OLUŞTURMA ÇALIŞMASI**

İsmail YÜKSEL

DOKTORA TEZİ

Eğitim Bilimleri (Eğitim Programları ve Öğretim) Anabilim Dalı

Danışman: Prof. Dr. Mustafa SAĞLAM

Eskişehir

Anadolu Üniversitesi Eğitim Bilimleri Enstitüsü

Mart 2010

DEVELOPMENT OF TURKISH PROGRAM EVALUATION STANDARDS

İsmail YÜKSEL

DISSERTATION

Department of Educational Sciences

Supervisor: Prof. Dr. Mustafa SAĞLAM

Eskişehir

Anadolu University Graduate School of Educational Sciences

March 2010

DOKTORA TEZ ÖZÜ

TURKIYE İÇİN PROGRAM DEĞERLENDİRME STANDARTLARI OLUŞTURMA ÇALIŞMASI

İsmail YÜKSEL

Eğitim Bilimleri Anabilim Dalı Eğitim Programları ve Öğretim Programı
Anadolu Üniversitesi Eğitim Bilimleri Enstitüsü,
Mart, 2010

Danışman: Prof. Dr. Mustafa SAĞLAM

Eğitim programları, yetiştirilecek insan özelliklerinin belirleyicisidir. Bir ülkenin, bir kurumun eğitim programlarına bakılarak bu programların uygulanması sonucunda ortaya çıkacak eğitim çıktıları konusunda yordamlar yapılabilir. Bu nedenle uygulanan programların aksaklık ve eksiklikleri giderilmeli toplumdaki ve bilim alanındaki değişmelere göre yeniden düzenlenmeli diğer bir ifadeyle program sürekli değerlendirilmelidir. Öte yandan eğitim programları gelişigüzel değerlendirilemez. Program değerlendirmenin etkili olarak gerçekleştirilebilmesi için tıpkı program geliştirme sürecine benzer süreçlerden geçirilmesi; plânlanması, uygulanması ve değerlendirilmesi gerekmektedir. Program değerlendirmenin bu sistematik içinde yürütülebilmesi için başta Amerika Birleşik Devletleri olmak üzere birçok gelişmiş ülke program değerlendirme standardı, yönergesi ya da ilkesi geliştirmiştir.

Türkiye’de program değerlendirme standartlarının geliştirilmesine ilişkin herhangi bir çalışma bulunmamaktadır. Ancak son yıllarda eğitim programlarının değerlendirilmesine yönelik alanyazında oldukça fazla sayıda program değerlendirme çalışmasının yapıldığı görülmektedir. Ancak bu çalışmaların genellikle programın bir boyutuna odaklanmış olduğu, çoğunun nicel yöntemler ve genellikle yalnızca anketler yoluyla yürütüldüğü, program değerlendirme sürecini tamamen yansıtmadığı, genellikle akademik kaygılarla derslerin etkililiğine yönelik ve okul, bölüm, fakülte veya il düzeyinde yürütülen çalışmalar niteliğinde olduğu ve bulgularının genellikle yetersiz olduğu görülmüştür. Bu sorunlar Türkiye’de program değerlendirme çalışmalarında izlenilmesi gereken değerlendirme ölçüt ve standartlarının eksikliğini de ortaya

çıkarmaktadır. Bu eksikliğin giderilmesi için ilk olarak ABD’de Ortak Komite tarafından geliştirilmiş olan ve daha sonra İsviçre, Almanya, Avustralya, Kanada ve Kenya (Afrika ülkelerinde kullanılmak üzere) tarafından kendi ülkelerine uyarlanan program değerlendirme standartları temel alınarak Türkiye’deki program değerlendirme çalışmaları için kullanılabilir “Program Değerlendirme Standartları”nın oluşturulmasına gereksinim vardır. Bu araştırma bu gereksinimi karşılamaya dönük olarak hazırlanmıştır.

Bu araştırma, tarama modellerinden kullanılarak gerçekleştirilmiştir. Araştırmanın çalışma evrenini doktora eğitim programları ve öğretim alanında tamamlamış ve 2008–2009 öğretim yılında 94 devlet üniversitesi bünyesinde yer alan 64 eğitim fakültesinde görev yapan görev yapan 158 öğretim üyesi oluşturmaktadır. Araştırmanın çalışma evreni içinde yer alan 158 uzmandan 138’ine (%87) ulaşılabilmektedir. Araştırmanın amacı kapsamında gerekli olan veriler, araştırmacı tarafından alan uzmanlarına uygulanmak üzere hazırlanan “Türkiye İçin Program Değerlendirme Standartları Oluşturma Çalışması Anketi” aracılığıyla web üzerinden toplanmıştır. Toplanan verilerin çözümlenmesinde sayı, yüzde, aritmetik ortalama, standart sapma, tek yönlü varyans çözümlenmesi ve Tukey HSD testlerinden yararlanılmıştır.

Araştırma sonunda, alan uzmanlarının dört temel standart alanında 23 standart ve bu standartlar altında yer alan 110 gösterge ifadesinin Türkiye için uygunluğuna ilişkin verdikleri yanıtlarının sayı, yüzde ve aritmetik ortalama kullanılarak yapılan çözümlenme sonuçları şunlardır:

- “*Yararlılık*” standartları altında yer alan beş standart Türkiye için “uygun” bulunmuş, bu standartlar altında yer alan 24 göstergeden 21’i uzmanlar tarafından “uygun” bulunmuş, diğer üç gösterge ise uzmanlar tarafından uygun bulunmamıştır. Alana ilişkin genel ortalama (4.10) bu standart alanının beş standart ve 21 göstergesi ile birlikte uzmanlar tarafından Türkiye için uygun görüldüğünü ortaya koymaktadır.
- “*Yürütülebilirlik*” standartları altında yer alan üç standart Türkiye için “uygun” bulunmuş, bir standart “kısmen uygun” bulunmuştur. Dört standart altında yer alan 17 göstergeden 13’ü uzmanlar tarafından “uygun” görülmüş,

dört gösterge ise uygun görülmemiştir. Alana ilişkin genel ortalama (3.96) bu standart alanının dört standart ve 13 göstergesi ile birlikte uzmanlar tarafından Türkiye için uygun görüldüğünü ortaya koymaktadır.

- “*Uygunluk*” standartları altında yer alan altı standart Türkiye için “uygun” bulunmuş, bu standartlar altında yer alan 33 göstergeden 30’u uzmanlar tarafından “uygun” bulunmuş, diğer üç gösterge ise uzmanlar tarafından uygun bulunmamıştır. Alana ilişkin genel ortalama (4.21) bu standart alanının altı standart ve 30 göstergesi ile birlikte uzmanlar tarafından Türkiye için uygun görüldüğünü ortaya koymaktadır.
- “*Doğruluk*” standartları altında yer alan sekiz standart Türkiye için “uygun” bulunmuş, bu standartlar altında yer alan 36 göstergeden 30’u uzmanlar tarafından “uygun” bulunmuş, diğer altı gösterge ise uzmanlar tarafından uygun bulunmamıştır. Alana ilişkin genel ortalama (4.09) bu standart alanının sekiz standart ve 30 göstergesi ile birlikte uzmanlar tarafından Türkiye için uygun görüldüğünü ortaya koymaktadır.
- Alan uzmanların görüşlerine göre program değerlendirme standart alanlarının uygunluk düzeyi istatistiksel olarak anlamlı farklılık göstermektedir. Bu fark, “*Yürütülebilirlik*” standartları ile “*Uygunluk*” standartları arasındaki farklılıktan kaynaklanmaktadır.

Sonuç olarak alan uzmanlarının görüşlerine sunulan dört standart alanı altında 23 standart ve bu standartlar altında yer alan 94 gösterge Türkiye için uygun program değerlendirme standartları ve göstergeleri olarak kabul edilmiştir.

ABSTRACT

DEVELOPMENT OF TURKISH PROGRAM EVALUATION STANDARDS

İsmail YÜKSEL

Department of Educational Sciences Curriculum Development and Instruction Program
Anadolu University Graduate School of Educational Sciences,

March, 2010

Supervisor: Prof. Dr. Mustafa SAĞLAM

Educational programs are the determinants of the characteristics of the individuals to be trained. By looking deep into the educational programs of a country or an institution, some connotations regarding educational outcomes may be carried out. Thus, the deficiencies and faults of the implemented programs should be remedied or rearranged in accordance with the changes in society and science; in other words, they should be evaluated systematically. On the other hand, educational programs cannot be evaluated casually. Thus, to conduct an effective program evaluation study, systematic procedures are essential as in program development studies; planning, implementation and evaluation are required. In order to handle a program evaluation in such stages, a few countries, notably the USA, have developed program evaluation standards, principles or guidelines.

In Turkey, there is no study on development of program evaluation standards yet, it can be seen that many studies on evaluation of programs have been recently conducted in literature. However, most of these studies have been focused on only one aspect of the program, and conducted as quantitative studies via questionnaires, obtained inadequate findings. In fact, these studies have been generally conducted to define course efficiency with academic anxiety. These issues reveal the need of standards or criteria to be followed/used in program evaluation studies in Turkey. Thus, to fulfil this need, new evaluation standards for program evaluations studies in Turkey should be developed on the basis of the standards which were developed in the USA by the Joint Committee and

then adapted by Switzerland, Germany, Australia, Canada and Kenya. This study is thus designed to fulfil this need.

This study was conducted with survey research method. The target population of this study consisted of 158 academic teaching members (experts whose PhD is on curriculum and instruction) working in 64 education faculties of 94 state universities in 2008-2009 academic years. Out of 158 academic teaching members, 138 (%87) completed the questionnaire titled as “Development of Turkish Program Evaluation Standards”. While analyzing the collected data, numbers, percentages, one-way ANOVA, and Tukey HSD was applied.

According to the experts’ views on the appropriateness of 23 evaluation standards and 110 indicators in Turkish context, the results are as follows:

- All “*Utility*” standards in the questionnaire (five standards) were found appropriate in Turkish context and of 24 indicators of these standards, 21 indicators were approved for Turkey. On the other hand, three indicators were found inappropriate. The mean of “*Utility*” domain (4.10) showed that the experts accepted five standards with their 21 indicators as “*Utility*” standards for Turkey.
- Of four standards under “*Feasibility*” domain, three were found appropriate and one found partly appropriate. Additionally, the experts approved 13 indicators of 17, but refused four indicators’ appropriateness to Turkish context. The mean of “*Feasibility*” domain (3.96) indicated that the experts accepted four standards with their 13 indicators as “*Feasibility*” standards for Turkey.
- The appropriateness of six “*Propriety*” standards and 30 indicators to Turkish context were *approved* by the participants, but the three indicators were found inappropriate for Turkey. The mean of “*Propriety*” domain (4.21) showed that the experts accepted six standards with their 30 indicators as “*Propriety*” standards for Turkey.

- All “*Accuracy*” standards in the questionnaire (eight standards) and 30 indicators were approved as appropriate for Turkey. However, the experts found six indicators inappropriate. The mean of “*Accuracy*” domain (4.09) indicated that the experts accepted eight standards with their 30 indicators as “*Accuracy*” standards for Turkey.
- According to the experts’ views, the differences between the appropriateness levels of standard domains are statistically significant. This is caused due to the difference between “*Feasibility*” domain and “*Propriety*” domain.

Consequently, it can be concluded that under the four domains, 23 of 23 standards and 94 of 110 indicators were accepted as Turkish program evaluation standards and indicators according to the experts’ views.

JÜRİ VE ENSTİTÜ ONAYI

İsmail YÜKSEL'in "Türkiye İçin Program Değerlendirme Standartları Oluşturma Çalışması" başlıklı tezi tarihinde, aşağıda belirtilen jüri üyeleri tarafından Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca Eğitim Bilimleri Anabilim Dalı Eğitim Programları ve Öğretim Programında, Doktora tezi olarak değerlendirilerek kabul edilmiştir.

	Adı-Soyadı	İmza
Üye (Tez Danışmanı)	Prof. Dr. Mustafa SAĞLAM
Üye	Prof. Dr. Şefik YAŞAR
Üye	Prof. Dr. Ayhan HAKAN
Üye	Doç. Dr. Kıymet SELVİ
Üye	Doç. Dr. Mehmet GÜLTEKİN

Prof. Dr. Esmahan AĞAOĞLU
Eğitim Bilimleri Enstitüsü Müdürü

ÖNSÖZ

Günümüzde hemen tüm gelişmiş ülkelerde eğitimde sürdürülebilir niteliğin sağlanması ve bu niteliğin güvence altına alınması amacıyla yoğun çalışmalar yürütülmektedir. Bu kapsamda, eğitimde niteliğin temel bileşenlerinden biri olan eğitim programlarının geliştirilmesine özel bir önem verilmektedir. Eğitim programlarının geliştirilmesi, alan uzmanı bireyler ya da kuruluşlar tarafından geliştirilen belirli ölçütler, ilkeler ve standartlar çerçevesinde yapılan program değerlendirme çalışmalarının sonuçlarına dayanır. Türkiye’de de program değerlendirme çalışmalarına katkı sağlamak amacıyla önemli bir kesiminin yüksek lisans ve doktora tez araştırmalarının oluşturduğu bilimsel çalışmaların sayısı son yıllarda artmıştır. Ancak Türkiye’de program değerlendirme çalışmalarında temel alınacak ve araştırmacılara yol gösterecek standartlar ya da ölçütler henüz oluşturulmamıştır. Bu nedenle, bu çalışmada alan uzmanlarının görüşleri doğrultusunda Türkiye’de program değerlendirme çalışmalarının planlanması, yürütülmesi ve değerlendirilmesinde kullanılacak standartlar ve göstergeler geliştirilmiştir.

Bu çalışmanın ortaya çıkmasında ilk olarak, doktora eğitimim süresince ve araştırmanın her aşamasında değerli katkılarıyla, bıkmadan, usanmadan bana yol gösteren, desteğini ve yardımlarını esirgemeyen, akademik olarak gelişmemde bana büyük katkı sağlayan hem kişiliği hem de akademik kimliğiyle her zaman örnek aldığım hocam ve tez danışmanım sayın Prof. Dr. Mustafa Sağlam’a, program değerlendirme alanını seçmemde büyük katkısı olan ve bu alanda ilerlemem konusunda beni sürekli güdüleyen, alanla ilgili bilgilerini ve kaynaklarını her zaman benimle paylaşan hocam sayın Prof. Dr. Şefik Yaşar’a, kendisinden aldığım derslerle program geliştirme ve değerlendirme bilgilerimi şekillendirdiğim ve veri toplama aracının oluşturulmasında eşsiz önerilerini aldığım hocam sayın Prof. Dr. Ayhan Gaffar Hakan’a, veri toplama aracının madde havuzunun oluşturulmasında çok değerli görüşlerini aldığım, tez jürilerindeki değerli katkıları ve olumlu eleştirileri ile beni yönlendiren hocam sayın Doç. Dr. Kıymet Selvi’ye ve bu alanda beni sürekli destekleyen ve her zaman yanımda olacağını bildiğim değerli hocam sayın Doç. Dr. Mehmet Gültekin’e içten teşekkürlerimi sunuyorum.

Araştırmamın her aşamasında bana desteklerini esirgemeyen, başım sıkıştığında kendilerine koştuğum dostlarım, arkadaşlarım Yard.Doç.Dr. Abdullah Adıgüzel, Dr. Dilruba Kürüm ve Öğr.Gör. Derya Atik Kara'ya, araştırmanın uygulama sürecinde gerekli web sayfalarının hazırlanmasında büyük emeği geçen dostlarım Araş. Gör. Kerem Kılıçer ve Araş. Gör. Gürkay Birinci'ye ve araştırmanın çeşitli aşamalarında katkıda bulunan ve araştırmanın bilgi toplama aracının uygulanması aşamasında çok büyük yardım ve desteklerini gördüğüm hocalarıma, arkadaşlarıma ve araştırmaya katılan tüm öğretim elemanlarına çok teşekkür ederim.

Ayrıca tüm yaşamım boyunca benim öğrenimim için yaptıkları her türlü fedakârlıkla bugünlere gelmeme vesile olan sevgili annem Altun Yüksel ve sevgili babam Ömer Yüksel ile desteklerini her zaman yanımda hissettiğim kardeşlerim Halil İbrahim, Seyit, Sultan ve Fadime'ye de teşekkür ediyorum.

Son olarak, birlikteliğimizin ilk gününden itibaren daima yanımda olan, akademisyenliği seçmemde etkin rol oynayan ve bu konuda bana her türlü desteği sağlayan, doktora eğitimine başladığım anda benimle beraber sıkıntıları, mutlulukları paylaşan sevgili eşim Araş. Gör. İlknur Yüksel ve ikimizin de doktora eğitimimizin son zamanlarında hayatımıza girmesiyle ailemize neşe getiren, mutluluk kaynağımız, gözümüzün nuru kızımız Bahar Nur Yüksel'e teşekkür ederim.

İsmail YÜKSEL
2010, Eskişehir

İÇİNDEKİLER LİSTESİ

DOKTORA TEZ ÖZÜ	ii
ABSTRACT.....	v
JÜRİ VE ENSTİTÜ ONAYI.....	viii
ÖNSÖZ.....	ix
İÇİNDEKİLER LİSTESİ.....	xiii
TABLolar LİSTESİ.....	xvi
ŞEKİLLER LİSTESİ	xix
1. GİRİŞ	1
1.1. Eğitim Programı	4
1.2. Eğitimde Program Geliştirme	7
1.3. Eğitimde Program Değerlendirme.....	9
1.3.1. Program Değerlendirmenin Amacı.....	18
1.3.2. Program Değerlendirmenin İşlevi	20
1.3.3. Program Değerlendirme Süreci	22
1.3.4. Program Değerlendirme Türleri	31
1.3.5. Program Değerlendirme Yaklaşımları ve Modelleri	36
1.3.5.1. Amaca Dayalı Program Değerlendirme Yaklaşımı ve Modelleri.....	36
1.3.5.2. Yönetime Dayalı Program Değerlendirme Yaklaşımı ve Modelleri	47
1.3.5.3. Uzman Odaklı Program Değerlendirme Yaklaşımı ve Modelleri	58
1.3.5.4. Tüketici Program Odaklı Değerlendirme Yaklaşımı ve Modelleri.....	65
1.3.5.5. Katılımcı Odaklı Program Değerlendirme Yaklaşımı ve Modelleri.....	69
1.4. Eğitimde Program Değerlendirme Standartlarının Tanımı, Kapsamı ve Kullanımı	82
1.4.1. Standardın Kuramsal Çerçevesi.....	83
1.4.2. Standart Geliştirme Aşamaları.....	84
1.4.3. Standart Türleri.....	87
1.5. Program Değerlendirme Standartlarının Kuramsal Çerçevesi	90
1.5.1. Program Değerlendirme Standartlarının Amacı ve İşlevi	94
1.5.2. Program Değerlendirme Standartlarının Gelişimi.....	96
1.6. Sorun.....	116
1.7. Araştırmanın Amacı	117
1.8. Araştırmanın Önemi	118
1.9. Sayıtlar.....	119
1.10. Araştırmanın Sınırlılıkları.....	119
1.11. Tanımlar.....	119
1.12. Kısaltmalar.....	121
2. İLGİLİ ARAŞTIRMALAR	123

2. 1. Yurtdışında Yapılan Araştırmalar	123
2. 2. Yurtiçinde Yapılan Araştırmalar	134
3. YÖNTEM.....	136
3.1. Araştırma Modeli.....	136
3.2. Çalışma Evreni	136
3.3. Verilerin Toplanması	138
3.4. Veri Toplama Aracının Hazırlanması.....	138
3.5. Veri Toplama Aracının Uygulanması.....	142
3.6. Verilerin Çözümlemesi.....	143
4. BULGULAR VE YORUMLARI	146
4.1. Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarının Uygunluğuna İlişkin Görüşleri.....	146
4.1.1. Alan Uzmanlarının <i>Paydaşların Tanımlanması</i> Standartının Göstergelerine İlişkin Görüşleri	146
4.1.2. Alan Uzmanlarının <i>Değerlendirici Yeterliği ve Güvenirliği</i> Standartının Göstergelerine İlişkin Görüşleri.....	149
4.1.3. Alan Uzmanlarının <i>Verilerin Kapsamı ve Seçimi /Yararlığı</i> Standartının Göstergelerine İlişkin Görüşleri.....	152
4.1.4. Alan Uzmanlarının <i>Raporun Kapsamlılığı ve Anlaşılabilirliği</i> Standartının Göstergelerine İlişkin Görüşleri.....	154
4.1.5. Alan Uzmanlarının <i>Sonuçların Yararlılığı/Etkisi</i> Standartının Göstergelerine İlişkin Görüşleri	158
4.2. Alan Uzmanlarının Program Değerlendirmede “Yürütülebilirlik” Standartlarının Uygunluğuna İlişkin Görüşleri.....	160
4.2.1. Alan Uzmanlarının <i>Uygun Yöntemler</i> Standartının Göstergelerine İlişkin Görüşleri.....	161
4.2.2. Alan Uzmanlarının <i>Grup ve Kurum Desteği/Yararlığı</i> Standartının Göstergelerine İlişkin Görüşleri	163
4.2.3. Alan Uzmanlarının <i>Maliyet-Yarar Çözümlemesi</i> Standartının Göstergelerine İlişkin Görüşleri	165
4.2.4. Alan Uzmanlarının <i>Değerlendirme Araştırmasının Yönetimi</i> Standartının Göstergelerine İlişkin Görüşleri.....	168
4.3. Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarının Uygunluğuna İlişkin Görüşleri.....	171
4.3.1. Alan Uzmanlarının <i>Resmi Sözleşme</i> Standartının Göstergelerine İlişkin Görüşleri.....	171
4.3.2. Alan Uzmanlarının <i>Etik Kurallara Uygunluk</i> Standartının Göstergelerine İlişkin Görüşleri	174
4.3.3. Alan Uzmanlarının <i>Adil Değerlendirme</i> Standartının Göstergelerine İlişkin Görüşleri.....	178
4.3.4. Alan Uzmanlarının <i>Şeffaf Raporlaştırma</i> Standartının Göstergelerine İlişkin Görüşleri.....	180

4.3.5. Alan Uzmanlarının <i>Sonuçların Açıklanması</i> Standartının Göstergelerine İlişkin Görüşleri	182
4.3.6. Alan Uzmanlarının <i>Finansal Uygunluk</i> Standartının Göstergelerine İlişkin Görüşleri.....	185
4.4. Alan Uzmanlarının Program Değerlendirmede “ <i>Doğruluk</i> ” Standartlarının Uygunluğuna İlişkin Görüşleri.....	188
4.4.1. Alan Uzmanlarının <i>Değerlendirilecek Programın Tanımlanması</i> Standartının Göstergelerine İlişkin Görüşleri.....	189
4.4.2. Alan Uzmanlarının <i>Durum Çözümlemesi</i> Standartının Göstergelerine İlişkin Görüşleri	191
4.4.3. Alan Uzmanlarının <i>Değerlendirme Araştırmasının Amacının Belirlenmesi</i> Standartının Göstergelerine İlişkin Görüşleri.....	192
4.4.4. Alan Uzmanlarının <i>Değerlendirme Araştırmasının Yönteminin Belirlenmesi</i> Standartının Göstergelerine İlişkin Görüşleri.....	194
4.4.5. Alan Uzmanlarının <i>Veri Kaynaklarının Tanımlanması</i> Standartının Göstergelerine İlişkin Görüşleri	197
4.4.6. Alan Uzmanlarının <i>Verilerin Geçerliliği ve Güvenirliği</i> Standartının Göstergelerine İlişkin Görüşleri	199
4.4.7. Alan Uzmanlarının <i>Verilerin Çözümlemesi</i> Standartının Göstergelerine İlişkin Görüşleri	201
4.4.8. Alan Uzmanlarının <i>Sonuçların Gereçlendirilmesi</i> Standartının Göstergelerine İlişkin Görüşleri	204
4.5. Program Değerlendirme Standartlarının Uygunluk Düzeyinin Standart Alanlarına Göre İncelenmesi	208
4.6. Alan Uzmanlarının Program Değerlendirme Standartları ve Göstergelerine İlişkin Önerileri	210
4.6.1. Alan Uzmanlarının “ <i>Yararlık</i> ” Standartlarına İlişkin Önerileri.....	211
4.6.2. Alan Uzmanlarının “ <i>Yürütülebilirlik</i> ” Standartlarına İlişkin Önerileri.....	212
4.6.3. Alan Uzmanlarının “ <i>Uygunluk</i> ” Standartlarına İlişkin Önerileri	214
4.6.4. Alan Uzmanlarının “ <i>Doğruluk</i> ” Standartlarına İlişkin Önerileri.....	215
5. SONUÇLAR, TARTIŞMA VE ÖNERİLER.....	217
5.1. Sonuç	217
5.2. Tartışma	225
5.3. Öneriler	239
EKLER.....	242
EK 1.Türkiye Program Değerlendirme Standartları	243
EK 2.Alan Uzmanları İçin Veri Toplama Aracı	247
EK 3.Uzmanlara Gönderilen E-davet Mektubu	255
EK 4.Program Değerlendirme Standartları Oluşturma Çalışması Anketinin Web Sayfası Görüntüleri.....	256
KAYNAKÇA.....	258

TABLolar LİSTESİ

Tablo 1.	Veri Türlerine Göre Çözümleme Teknikleri Örnek Tablosu -----	29
Tablo 2.	Biçimlendirici ve Toplam Değerlendirme Arasındaki Farklılıklar -----	35
Tablo 3.	ÇGSÜ Değerlendirme Modelinin Dört Değerlendirme Türü -----	51
Tablo 4.	Dört Değerlendirme Türünün Biçimlendirici ve Toplam Değerlendirme İşlevleri -----	55
Tablo 5.	Uzmanlığa Dayalı Değerlendirme Yaklaşımının Değerlendirme Türleri ---	59
Tablo 6.	Farklı Uzmanlara Göre Standartların Sınıflandırılması-----	88
Tablo 7.	Program Değerlendirme Standartlarının Karşılaştırılması -----	108
Tablo 8.	Araştırmanın Çalışma Evrenini Oluşturan Uzmanların Akademik Unvanlarına Göre Dağılımı-----	137
Tablo 9.	Uzmanlara Gönderilen ve Uzmanlardan Geri Dönen Anket Sayısı -----	137
Tablo 10.	Kapsam Geçerliliğine Dayalı Olarak Çıkarılması ve Eklenmesi Önerilen Maddeler-----	140
Tablo 11.	Veri Toplama Aracının Standart Alanlarına İlişkin Miles ve Huberman Güvenirlik Katsayıları -----	141
Tablo 12.	Seçeneklere İlişkin Sınırlar ve Gruplamalar -----	144
Tablo 13.	Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarından “Paydaşların Tanımlanması” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	147
Tablo 14.	Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarından “Değerlendirici Yeterliliği ve Güvenirliği” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	150
Tablo 15.	Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarından “Verilerin Kapsamı ve Seçimi /Yararlığı” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	153
Tablo 16.	Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarından “Raporun Kapsamlılığı ve Anlaşılabilirliği” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	155
Tablo 17.	Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarından “Sonuçların Yararlılığı/Etkisi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	158
Tablo 18.	Program Değerlendirmede “Yararlık” Standartlarının Aritmetik Ortalamaları-----	160
Tablo 19.	Alan Uzmanlarının Program Değerlendirmede “Yürütülebilirlik” Standartlarından “Uygun Yöntemler” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	161
Tablo 20.	Alan Uzmanlarının Program Değerlendirmede “Yürütülebilirlik” Standartlarından “Grup Ve Kurum Desteği/Yararlığı” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	163

Tablo 21.	Alan Uzmanlarının Program Değerlendirmede “Yürütülebilirlik” Standartlarından “Maliyet-Yarar Çözümlemesi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	166
Tablo 22.	Alan Uzmanlarının Program Değerlendirmede “Yürütülebilirlik” Standartlarından “Değerlendirme Araştırmasının Yönetimi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	168
Tablo 23.	Program Değerlendirmede “Yürütülebilirlik” Standartlarının Aritmetik Ortalamaları -----	171
Tablo 24.	Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Resmi Sözleşme” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	172
Tablo 25.	Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Etik Kurallara Uygunluk” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	175
Tablo 26.	Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Adil Değerlendirme” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	178
Tablo 27.	Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Şeffaf Raporlaştırma” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	181
Tablo 28.	Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Sonuçların Açıklanması” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	183
Tablo 29.	Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Finansal Uygunluk” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	185
Tablo 30.	Program Değerlendirmede “Uygunluk” Standartlarının Aritmetik Ortalamaları -----	188
Tablo 31.	Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Değerlendirilecek Programın Tanımlanması” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	189
Tablo 32.	Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Durum Çözümlemesi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	191
Tablo 33.	Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Değerlendirme Araştırmasının Amacının Belirlenmesi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	193
Tablo 34.	Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Değerlendirme Araştırmasının Yönteminin Belirlenmesi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	195

Tablo 35.	Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Veri Kaynaklarının Tanımlanması” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	197
Tablo 36.	Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Verilerin Geçerliliği ve Güvenirliği” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	199
Tablo 37.	Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Verilerin Çözümlemesi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri-----	202
Tablo 38.	Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Sonuçların Gerekçelendirilmesi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri -----	205
Tablo 39.	Program Değerlendirmede “Doğruluk” Standartlarının Aritmetik Ortalamaları-----	208
Tablo 40.	Program Değerlendirme Standartlarının Standart Alanlarına Göre Uygunluk Düzeyi-----	209
Tablo 41.	Alan Uzmanlarının Program Değerlendirme Standartlarının Uygunluk Düzeyine İlişkin Görüş Puanlarına Uygulanan Tek Yönlü Varyans Çözümlemesi Sonuçları -----	209
Tablo 42.	Alan Uzmanlarının Program Değerlendirme Standartlarının Uygunluk Düzeyine İlişkin Görüş Puanlarına Uygulanan Tukey HSD Testi Sonuçları-----	210
Tablo 43.	Alan Uzmanlarının “Yararlık” Standartlarının Göstergelerine İlişkin Yorumları ve Önerileri -----	211
Tablo 44.	Alan Uzmanlarının “Yürütülebilirlik” Standartlarının Göstergelerine İlişkin Yorumları ve Önerileri-----	213
Tablo 45.	Alan Uzmanlarının “Uygunluk” Standartlarının Göstergelerine İlişkin Yorumları ve Önerileri -----	214
Tablo 46.	Alan Uzmanlarının “Doğruluk” Standartlarının Göstergelerine İlişkin Yorumları ve Önerileri -----	215

ŞEKİLLER LİSTESİ

Şekil 1.	Eğitim Programının Öğeleri	6
Şekil 2.	Değerlendirme Yönetim Planı Örnek Formu	27
Şekil 3.	Program Değerlendirme Sürecinde Biçimlendirici Değerlendirme-Toplam Değerlendirme İlişkisi	35
Şekil 4.	Amaca Dayalı Program Değerlendirme Modelinin Aşamaları	38
Şekil 5.	Metfessel-Michael Değerlendirme Modeli	40
Şekil 6.	Çelişki/Fark Değerlendirme Süreci	42
Şekil 7.	Provus'un Program Değerlendirme Modelinin Aşamaları.....	43
Şekil 8.	ÇGSÜ Değerlendirme Modelinin Temel Kavramları	52
Şekil 9.	Stake'in Program Değerlendirme İçin Veri Toplama Planı	71
Şekil 12.	Yanıtlayıcı Değerlendirmede Belirgin Eylemler.....	76
Şekil 13.	Katılımcı Değerlendirme Modelinin Temel Özellikleri.....	78
Şekil 14.	Katılımcı Program Değerlendirme Sürecinin Aşamaları	79
Şekil 15.	Standart Geliştirme Süreci	86
Şekil 16.	OKPDS, İsviçre ve Almanya Program Değerlendirme Standartlarının Tarihsel Gelişimi	106

1. GİRİŞ

Bilim ve teknolojiye yaşanan hızlı deęişme ve gelişmeler, tüm toplumları bilgi toplumu olma yolunda zorlamaktadır. Bilginin yarattığı bu hızlı deęişimlere ayak uydurabilmenin önkoşulu, toplumsal kalkınmanın ve gelişmenin temel öğelerinden biri olan insan gücünü oluşturan bireylere çağın gerektirdiği yeni bilgi ve becerileri yeterli düzeyde kazandırmaktır. Her ülkede gereksinim duyulan nitelikli insan gücünün yetiştirilmesinde “eğitim” önemli bir görev üstlenmektedir.

Eğitim, belli amaçlara göre bireyin davranışlarında deęişiklik oluşturma sürecidir. Bu süreçte ortaya konan amaçlar, toplumların ve bireylerin istemlerini, gereksinmelerini yansıtır. Bu açıdan bakıldığında, her çağdaş toplum, yeni yetişen kuşaklara bir yandan var olan kültürünü aktarmayı, bir yandan da yeni kuşakların bu kültürü geliştirmesini ister. Bu durum, günümüzde toplumların yenilikçi ve yaratıcı bireyler yetiştirme konusunda eğitim hizmetlerinden yararlanması zorunluluğunu ortaya çıkarmaktadır. Yaşanan deęişme ve gelişmelere ayak uydurabilecek bireylerin yetiştirilmesi ise eğitimin niteliğinin geliştirilmesini gerektirmektedir.

Eğitimin niteliği, eğitim sisteminin çıktısı ya da ürünü olan ve tüm toplumsal sistemlerin işleyişini ve verimliliğini etkileyen insan gücünün niteliğini de belirler. Bu nedenle eğitimde niteliğin geliştirilmesi, ülkedeki yürütme erkince belirlenen eğitim politikasının genel hedefi ve bu politikaya dayalı kararların uygulandığı eğitim kurumlarının da temel amacıdır. Eğitim kurumlarının bu amacı gerçekleştirebilmeleri, uygulanan eğitim programlarının, bu programı uygulamaktan sorumlu öğretmenlerin, programın gerektirdiği donanımların ve bu programa giren öğrencilerin niteliğine bağlıdır.

Eğitim programları eğitim kurumlarında yetiştirilecek insan özelliklerinin belirleyicisidir. Buna göre, çeşitli eğitim kurumları için hazırlanmış eğitim programlarına bakılarak bu programların uygulanması sonucunda ortaya çıkacak eğitim çıktıları ya da ürünleri konusunda yordamalar yapılabilir (Gözütok, 2001). Çünkü eğitim programı, bir eğitim kurumuna devam eden öğrencilerin kazanmaları istenen davranışların, bu davranışları

kazanmalarını sağlayacak derslerin ve konuların, bu konuların öğrenilme biçim ve yollarının ve öğrencilerin öngörülen davranışları kazanma durumlarının belirlenmesinin açıklandığı ayrıntılı ve çok boyutlu bir dokümandır. Bu boyutlar eğitim programının amaç, içerik, öğretme-öğrenme süreçleri ve değerlendirme öğelerini oluşturur. Bu öğelerin toplumun ve bireyin gereksinimlerine, çağın gereklerine, bilimsel ve teknolojik gelişmelere uygun olarak düzenlenmesi ve yürütülmesi eğitimin niteliğinin temel koşuludur. Bu anlamda eğitim programları eğitimin niteliğinin temel belirleyicisi konumunda olduğundan eğitim kurumlarının işlevlerini ve görevlerini etkili biçimde yerine getirebilmeleri, bir başka deyişle toplumun gereksinimi olan daha nitelikli çıktıları üretebilmesi için, eğitim programlarının düzenli olarak geliştirilmesi gereklidir. Çünkü Tyler'in (1969) da vurguladığı gibi, eğitimin niteliği uygulanan programların aksaklık ve eksiklikleri giderildikçe, toplumdaki ve bilim alanındaki değişmelere göre yeniden düzenlendikçe, diğer bir ifadeyle programlar geliştirildikçe artar.

Eğitim programları gelişigüzel geliştirilemez. Çünkü eğitim kurumlarında uygulanan eğitim programları, toplumun gereksinimleri ve bireylerin özelliklerini göz önünde bulunduran belli ölçütlere göre hazırlanır. Bu ölçütler, programları daha etkili ve işlevsel hale getirecek, daha sağlıklı kararların alınabilmesini sağlayacak durumları içerir. Hazırlanan programların bu ölçütlere uygun olup olmadığının belirlenmesi ve programları daha etkili hale getirecek doğru kararların alınabilmesi, bu kararların dayanaklarının bilimsel çalışmalarla araştırılmasına ve uygulamaların değerlendirilmesine bağlıdır. Uygulanan programlar değerlendirildiği sürece, eğitim kurumlarının işlevlerini yerine getirip getirmediği kontrol edilmiş ve eğitimin daha nitelikli duruma getirilmesi için programların geliştirilmesine olanak sağlanmış olur. Buna göre, Gözütok'un da (2001) belirttiği gibi, eğitim programlarının etkililiğinin sorgulanması ve değerlendirilmesi, programların geliştirilmesi için bir başlangıç noktasıdır.

Eğitim programlarının değerlendirilmesi, bu programlarda belirlenmiş olan amaçların, düzenlenen içeriğin ve uygulanan öğretme-öğrenme süreçlerinin ve değerlendirme durumlarının program sonucunda ulaşılması istenilen yeterlikleri hedef kitleye kazandırmadaki etkililiğinin, işlerliğinin, ekonomikliğinin, kullanılabilirliğinin ve işlevselliğinin sağlanmasındaki yeterliliklerini belirlemede ve elde edilen değerlendirme

sonuçlarına göre ilgili programın sürdürülmesine, geliştirilmesine ya da sonlandırılmasına karar vermede önemli olmaktadır. Program değerlendirme, hazırlanan bir eğitim programının güçlü ve zayıf yönlerini belirlemek üzere program uygulanmadan önce yapılabileceği gibi, programın etkililiğini saptamak amacıyla programın uygulanması sırasında ya da sonrasında da yapılabilir. Her iki durumda da program değerlendirme sonucunda elde edilen sonuçlar, program geliştirmecileri ilgili öğeleri gözden geçirmeye yöneltir.

Program geliştirmede olduğu gibi program değerlendirme de belli ölçütlere göre yapılmak zorundadır. Öncelikle program değerlendirme araştırmasını yapacak araştırmacının/araştırmacıların, program değerlendirmenin amacını, modelini, sürecini, yönetimini ve ilkelerini belirlemesi ve değerlendirme sürecini ve sonucunu etkileyebilecek etmenleri, değerlendirme sonucundan etkilenecekleri ve değerlendirme sonucundan yararlanacakları veya bu sonuçlara göre program hakkında karar verecek ilgilileri dikkate alması gereklidir.

Program geliştirme uzmanları, eğitim kurumlarında istenilen düzeye ulaşmak için eğitim programlarının belli standartlara sahip olması gerektiğini savunmaktadırlar (Edelfelt ve Raths, 1998). Bu standartlar aynı zamanda hem program değerlendirme çalışmaları hem de programların/kurumların akreditasyonunda yol gösterici olmaktadır. Standart, belli bir amaca ulaşmada uygun ve yeterli nitelik düzeyi olarak kabul edilmektedir (Richardson, 1994). Bu anlamda, program değerlendirme çalışmalarının da amaca hizmet edebilmesi için program değerlendirme standartlarına uygun olması gerekmektedir. Çünkü program değerlendirme standartları, hem program değerlendirme çalışması yapacak araştırmacılara çalışmanın planlanmasından sonuçlandırılmasına kadar her aşamasında yol göstermede, hem de yapılacak veya yapılan bir program değerlendirme çalışmasını değerlendirmede veya değerlendirmenin sonuçlarını kullanacak ve buna dayalı karar verecek yetkililere araştırmanın geçerliği ve güvenilirliği hakkında sağlam bilgi sunmada kullanılacak ölçütleri kapsar (Sağlam ve Yüksel, 2007).

Amerika Birleşik Devletleri başta olmak üzere pek çok ülkede eğitimin her alanında olduğu gibi, program değerlendirme çalışmalarında da uzman kuruluşlar tarafından geliştirilmiş olan program değerlendirme standartları kullanılmakta ve bu standartlar yeni

gereksinimler ve gelişmeler doğrultusunda sürekli geliştirilmektedir. Türkiye’de de eğitimde nitelik geliştirme çalışmalarının önemli bir boyutu olarak program değerlendirme çalışmaları yoğun biçimde uygulanmaktadır. Ancak yapılan program değerlendirme çalışmalarının önemli bir kesiminin değerlendirmenin temel öğeleri olan planlama, uygulama ve değerlendirme aşamalarına ilişkin bir takım ilke ve ölçütlere uygun olarak gerçekleştirildiğini söylemek güçtür. Çünkü ulaşılabilen alanyazın incelemesi sonuçlarına göre, Türkiye’de henüz geçerliği ve güvenilirliği sağlanmış program değerlendirme standartları oluşturulmamıştır. Bu durum program geliştirme çalışmalarının vazgeçilmez aşamasını oluşturan program değerlendirme alanında önemli bir eksiklik olarak görülmektedir. Bu eksikliğin giderilmesi için pek çok ülkede kullanılan program değerlendirme standartları da temel alınarak Türkiye’deki program değerlendirme çalışmalarında kullanılacak program değerlendirme standartlarının oluşturulmasına, daha sonra bu standartların gelişen ve değişen koşullara ve durumlara göre oluşturulmasına gereksinim vardır. Bu araştırma bu gereksinime dayanmakta ve bu gereksinimin bir bölümünü karşılaması beklenmektedir.

Araştırmanın bu bölümünde araştırma sorununun tanımlanmasında gerekli görülen eğitimde program geliştirme ve program değerlendirme kavramları ve bileşenleri ile program değerlendirme standartlarına ilişkin bilgiler verilmiştir.

1.1. Eğitim Programı

Program eski Yunancadan gelen bir sözcük olup “yapılacak işin bölümlerini ve her bölümün zamanını gösteren maddelerin tümü” anlamına gelmektedir (Gürkan, 2005). Oğuzkan (1993) ise programı, belli bir çalışmanın amacını, bölümlerini, yöntemini ve süresini gösteren plan olarak tanımlamaktadır. Eğitimde program kavramı ise eğitim programı, öğretim programı, ders programı ve ünite ve konu planı olarak sınıflandırılmıştır. Bunlardan eğitim programı en geniş kapsamlı olanıdır.

Yatırımlarını eğitime yapan devletler, bu eğitimi ülke kalkınmasında katkıya dönüştürmek için çeşitli önlemler almak durumundadırlar. Bu önlemlerden biri eğitimin niteliğinin artırılmasıdır. Eğitim programları ise eğitimin niteliği ile doğrudan ilişkilidir (Varış, 1996). Çünkü programlar, eğitilenlerin davranış standartlarından, öğrenme ve öğretim etkinliklerine değin tüm çalışmalara yön vermekte ve ülkenin genel eğitim

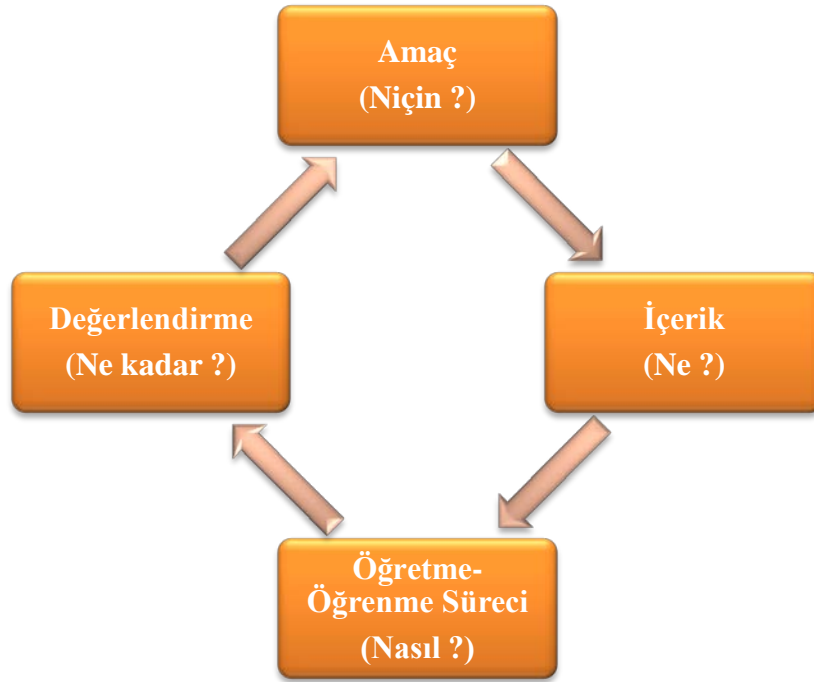
politikalarının uygulamaya dönüştürülmesinde bir aracı rolü oynamaktadır. Kısaca, bir eğitim kurumunun veya sosyal çevrenin, bireylerin yaşantılarını düzenlemek ve zenginleştirmek için ülkenin genel amaçlarıyla uyumlu bir şekilde yürüttüğü tüm eğitim etkinlikleri "eğitim programı" kapsamına girmektedir.

Latince *curriculum* sözcüğüyle tanımlanan eğitim programı kavramı ilk olarak Roma askerlerinin arenalarda at arabalarıyla idman yaptıkları koşu yolu için kullanılmaktaydı (Demirel, 2006). Daha sonra bu kavramın türetildiği "*curricle-atlı araba* ve "*currere-koşmak*" sözcükleri 1600'lü yıllardan itibaren Latince eserler yoluyla İngiliz diline girmiş ve belli başlı üniversitelerde kullanılmaya başlanmıştır. Kavramın eğitimde kullanılmaya başlanması yirminci yüzyılda gerçekleşmiştir (Quinion, 1996). Türkiye'de kullanılması ise 1930'lu yıllara denk gelmektedir. İlk olarak eğitim programı kavramı yerine Arapça "*müfret*" sözcüğünden türetilen *müfredat programı* kavramı kullanılmış ve "tekil olguların arka arkaya sıralanmasından oluşan liste" olarak tanımlanmıştır (Sözer, 1998)

Tyler'a (1969) göre, eğitim programı, geçmişteki ve günümüzdeki profesyonel uygulamalar ve kuramlar kullanılarak planlamayı ve genel ve özel amaçlara ulaşmayı amaçlayan, öğrencilerin bireysel deneyimlerini içeren bir programdır. Saylor ve Alexander (1974), programı "bir öğrenme planı" olarak tanımlarken, Wiles ve Bondi (1993) eğitim programının, öğrencilerin öğrenme durumu içinde ne yapacaklarıyla değil, yaptıkları şeylerin sonunda ne öğrenecekleri ile ilgili olduğunu belirtmektedirler. McNeil (1977) ise eğitim programını bir eğitim kurumunun öğreteceği şeylerin planı olarak tanımlamaktadır

Türkiye'de eğitim programı ile ilgili çalışmaların öncüsü olan Varış'a (1996) göre, eğitim programı, bir eğitim kurumunun çocuklar, gençler ve yetişkinler için sağladığı milli eğitim ve kurumun amaçlarının gerçekleştirilmesine dönük tüm etkinliklerdir. Ertürk (1993) de eğitim programını "yetişek" olarak ele almış, okulun ve öğretmenin yol göstericiliği altında okul içi ve dışındaki tüm öğretme-öğrenme etkinlikleri olarak tanımlamıştır. Gürkan (2005), eğitim programını kısaca öğrencilerin yaşantılarını düzenleme olarak ifade etmiştir.

Eđitim programının kendi iinde belli bir dzeni ve sistematik bir yapısı vardır. Bu yapı eřitli ğelerden oluřmakta ve bu ğeler arasında dinamik bir iliřki yer almaktadır. Dolayısıyla programın ğelerinin herhangi birinde ortaya ıkan bir aksaklık veya deđiřme, sistemde yer alan diđer ğeleri de etkilemektedir. Eđitim programının temel ğeleri ama, ierik, đretme-đrenme sreci ve deđerlendirmedir (řekil 1) .



řekil 1. Eđitim Programının ğeleri

řekil 1’de grldđđ gibi, eđitim programının ama ğesinde, “Bireyleri niin eđitiyoruz?” sorusuna yanıt aranmakta; ierik ğesinde, belirlenen amalar dođrultusunda “ne đretelim” sorusu sorgulanmakta; nc đe olan đretme-đrenme srecinde amalar dođrultusunda dzenlenen ieriđin nasıl đretileceđi irdelenmektedir. Programın son gesi ise deđerlendirmedir. Bu gede nceki ğeler dođrultusunda yapılan eđitimin niteliđinin kontrol yapılmaktadır. Amalara ne dzeyde ulařılıp ulařılmadıđının saptanmaya alıřıldıđı bu gede ortaya ıkan sonular programın geliřtirilmesi iin temel girdi olarak kullanılır. nk eđitim sisteminin bařarısı ve bařarısızlıđında byk payı olan ve bu sistemi ayakta tutmada nemli bir katkısı bulunan eđitim programının hızla deđiřen kořullar altında srekli bir geliřme iinde olması kaınılmazdır. Varıř (1996), programın geliřtirilmesinin ancak đretmenin bu dođrultudaki uyarımı ve isteđi ile olanaklı olduđunu belirtmekte ve bu uyarımın program geliřtirmenin bařlangı kořulunu

oluşturduğunu ve bunu mevcut programın uygulanmasıyla daha iyi ürün ve daha nitelikli öğrenci yetiştirmek için yürütülen etkinliklerin izleyeceğini ifade etmektedir. Öte yandan Varış (1996), program geliştirmenin, düzenlenen programın masa başında değiştirilmesi, bir kısım konuların çıkarılması veya yenilerinin eklenmesi olmadığını; aksine eğitim süreci ile ilgili olan bütün koşulların, bireylerin, ders kitapları ve araçların sürekli biçimde geliştirilmesini içeren uygulamalı bir süreç olduğunu vurgulamaktadır.

1.2. Eğitimde Program Geliştirme

Eğitimde program geliştirme, eğitim programlarının tasarlanması, uygulanması, değerlendirilmesi ve değerlendirme sonucu elde edilen veriler doğrultusunda yeniden düzenlenmesi sürecidir (Erden, 1998; Gürkan, 2005). Demirel (2006), program geliştirmeyi programın öğeleri olan amaç, içerik, öğretme-öğrenme süreci ve değerlendirme boyutları arasındaki dinamik ilişkiler bütünü olarak tanımlamaktadır. Sönmez'e (2001) göre program geliştirme, bilimsel yöntemlerle yapılan program değerlendirme araştırmasının bulguları doğrultusunda programın amaç, içerik, öğretim süreçleri ve değerlendirme boyutlarının geliştirilmesine yönelik koordine çabaların bütünüdür. Varış (1996), program geliştirmenin yazılı bir doküman hazırlamak olmadığını, mevcut programın uygulamada araştırmacı bir yaklaşımla geliştirilerek, öğrencide istendik davranış değişikliğini oluşturmak olduğunu; program geliştirmenin devamlı, kapsamlı ve uygulamalı bir süreç olduğunu belirterek program geliştirmeyi hazırlanmış bir programın araştırmacı bir yaklaşımla uygulamada geliştirilmesi olarak tanımlamaktadır.

Yukarda verilen tanımlara göre, program geliştirme sürecinde iki temel yaklaşımın öne çıktığı görülmektedir. Bunlar ürüne ağırlık veren yaklaşım ve sürece ağırlık veren yaklaşımdır (Fidan, 1997). Ürüne ağırlık veren yaklaşımda temel sayıltı, eğitim amaçlarının önceden belirlenip açıkça ortaya konulabileceğidir. Amaçların belirlenmesinde konu alanı, birey ve toplumun özellikleri önemli rol oynar. Amaçlar belirlendikten sonra amaçlara ulaşılmasına yardım edecek içerik ve öğretme-öğrenme ortamı belirlenir. Değerlendirme etkinliği ile de amaçlara ulaşılma düzeyi belirlenir (Erden, 1998). Sürece ağırlık veren yaklaşımda ise temel sayıltı eğitimin tüm sonuçlarının ve süreçlerinin önceden bilinemez olduğudur. Bu yaklaşımda süreç amaçlara, amaçlar da öğrenci gereksinimlerine göre belirlendiğinden öğretme-öğrenme süreci çeşitli etkenlere

bağlı olarak sürekli değişebilir. Bu yaklaşımda öğrenci gereksinimleri ön planda olduğu için program geliştirme aşamasında uzman görüşünden çok öğrenci ve öğretmenlerin görüş ve tercihleri önem taşımaktadır (Gürkan, 2005).

Program geliştirmede üç temel aşama vardır. Bunlar planlama, uygulama ve değerlendirmedir.

- *Planlama:* Program geliştirmede planlama aşaması, çeşitli çalışma gruplarının oluşturulması, çalışma planının hazırlanması, gereksinim saptama araştırmasının yapılması ve program öğelerinin nasıl düzenleneceğine karar verilmesi işlemlerini/süreçlerini kapsar (Demirel, 2006; Gürkan, 2005). Demirel (2006) çalışma gruplarını üç grup altında toplamıştır: Program Karar ve Koordinasyon Grubu, Program Çalışma Grubu, Program Danışma Grubu. Çalışma planı ne zaman hangi işin yapılacağını gösteren bir plandır. Planlama aşamasında programla ilgili alınacak kararlara temel olacak verilerin de toplanması gerekmektedir. Bu aşamada, toplum, konu alanı ve bireyin beklenti ve gereksinmelerinin ne olduğu sorusunun yanıtı gereksinim saptama araştırmasıyla aranır. Daha sonraki aşamada, gereksinim saptama araştırmasından elde edilen bulgular doğrultusunda eğitim programını oluşturan amaç, içerik, eğitim durumu ve değerlendirme öğeleri düzenlenir (Erden, 1998; Gürkan, 2005).
- *Uygulama:* Bir önceki aşamada hazırlanan programın seçilen belirli okullarda denenmesi ve aksayan yönlerinin giderilmesi çalışmaları program uygulamasının ilk aşamasını oluşturmaktadır. Genel olarak, bir eğitim programı denendikten sonra ülke genelinde uygulamaya konulur. Demirel (2006) programın alanda denenmesi sürecini planlama, uygulama ve sonuç olmak üzere üç aşamada gerçekleştiğini belirtmekte ve bu aşamalarda izlenmesi gereken adımları da şöyle sıralamaktadır (Demirel, 2006):
 - ▀ Uygulamanın planlanması
 - ▀ Deneme yapılacak okulların ve sınıfların seçilmesi
 - ▀ Okul yöneticisi ve öğretmenlerin seçilmesi
 - ▀ Okul yöneticisi ve öğretmenlere denenecek programın tanıtılması

- Hazırlanan program ve öğretim materyallerinin işe koşulması
- Denenen programın ve öğretim materyallerinin değerlendirilmesi
- *Değerlendirme:* Planlanan ve uygulanan programın program değerlendirme modelleri doğrultusunda uygun araştırma yöntem ve teknikleri ile değerlendirilip, değerlendirme sonuçlarının programın geliştirilmesinde kullanılmasıdır.

Günümüzde üretim ve hizmet alanlarında olduğu gibi eğitim alanında da değerlendirmenin çok önemli bir yeri vardır. Dünyada bilim ve teknoloji alanında yaşanan sürekli gelişim ve değişimlere her kurumda olduğu gibi eğitim kurumlarının da ayak uydurması gerekmektedir. Toplumun beklentilerini karşılayan, yeniliklere ve değişime açık bireylerin yetiştirilmesini temel görev olarak benimseyen eğitim kurumlarının üstlendikleri işlevlerini gerçekleştirebilmeleri, ancak toplumun beklentilerini uyumlu, bireylerin gereksinimlerini karşılayan eğitim programlarını uygulamalarıyla olanaklıdır. Bu nedenle nitelikli bir bireyin yetişmesi de nitelikli bir programın hazırlanmasını ve uygulanmasını gerektirir. Bireyden beklenen davranışlar, bireye öğretilecek bilgi ve beceriler, bireyin sahip olması gereken tutumlar ancak nitelikli program yoluyla bireye kazandırılabilir. Bu yönüyle eğitim programları yetiştirilecek bireylerin özelliklerinin belirleyicisidir (Gözütok, 2001). Eğitim sürecinin ürünü olan bireye bakılarak sistem veya program hakkında bilgiye ulaşılabilir, sistemin ve programın işleyip işlemediğine ya da bir aksaklığı olup olmadığına karar verilir. Bu nedenle bir eğitim sistemi uyguladığı programların başarıya ulaşip ulaşmadığını, toplumun beklentilerini karşılayıp karşılamadığını kontrol ederek programı sürekli değerlendirmeli ve değerlendirme sonuçlarını program geliştirme çalışmalarına yansıtmalıdır (Gözütok, 2001).

1.3. Eğitimde Program Değerlendirme

Eğitim, en genel anlamıyla bireyde istenilen yönde davranış değişikliği oluşturma sürecidir. Eğitim programı ise bu davranış değişikliğinin oluşturulacağı hedef kitlenin özelliklerine göre davranış değişikliğinin oluşmasını sağlayacak öğrenme yaşantılarını ve bunların etkilerini belirleme süreçlerini kapsar. Eğitimde değerlendirme de bu tür davranış değişmelerinin, program tasarısında saptanan ölçütlere göre beklenen sonuca ulaşip ulaşmadığını ortaya çıkarmak için başvurulan bir süreç olarak ortaya çıkmaktadır.

Bu süreçteki inceleme sonunda eğitim durumları, öğrenme yaşantıları ve davranış değişimleri arasındaki ilişkiler denencel olarak test edilerek programın amaçlarına hizmet etme dereceleri saptanmaktadır. Bu etkinlikler program öğelerinin değişmeye ve sürekli geliştirilmeye açık tutulması olanağını sağlamaktadır. Bu açıdan program hakkında bir karara ulaşabilmek ve bu karar uyarınca programı geliştirmek ancak değerlendirme sayesinde olanaklıdır (Bilen, 1999).

Alanyazında değerlendirmenin çok fazla tanımı vardır. Sözlüklere bakıldığında değerlendirme “değerini belirleme, değer kazandırma, kıymetlendirme, araştırma ve karar verme” olarak tanımlanmaktadır (TDK, 2009; Birsözlük, 2009; Ötüken sözlük, 2009; Sözlüknet, 2009). Fitzpatrick, Sanders ve Worthen (2004) ise eğitimde değerlendirme alanının öncülerinden biri olan Michael Scriven’in değerlendirme ile ilgili 60 farklı kavram kullandığını ve bunlardan en yaygın olanlarının da; “karar verme”, “değer biçme”, “çözümleme”, “eleştirme”, “inceleme”, “derecelendirme”, “gözden geçirme”, “sınama” ve “denetleme” olduğunu belirttiktedirler.

Fink (1995) değerlendirmeyi bir programın özelliklerinin ve boyutlarının ayrıntılı araştırması olarak ifade etmektedir. Değerlendirmeyi program için vazgeçilmez olarak gören ve değerlendirmenin bir eğitim programında nelerin yolunda gittiğinin ve nelerde aksaklık yaşandığının saptanması olduğunu vurgulayan McCain’e (2005) göre, değerlendirme uygulanan bir programın değerini belirleme ve geliştirme amacıyla verilen eğitim hakkında bir karar verme sürecidir. Morrison (1993) değerlendirmeyi okulların ve öğretmenlerin uyguladıkları ile uygulama sonucunda bekledikleri arasında ne düzeyde bir tutarlığın olduğunu belirleme süreci olarak ifade etmektedir. Demirel (2006), değerlendirmeyi kısaca programın etkililiği hakkında karar verme süreci olarak tanımlarken, Erden (1998) değerlendirmenin bir ölçme işlemi sonucunda elde edilen ölçme sonuçlarının bir ölçütle karşılaştırılması ve buna bağlı olarak bir değer yargısına ulaşma işi olduğunu vurgulamaktadır.

Program değerlendirme kavramı tanımlanırken eğitsel değerlendirme ve öğretimsel değerlendirmenin bilinmesinde yarar vardır. İki kavramdan en geneli olan ve program değerlendirmeyi ve öğretimsel değerlendirmeyi de kapsayan eğitsel değerlendirme, okul ile ilgili tüm değişkenleri içine alırken; öğretimsel değerlendirme öğrencilerin

değerlendirilmesini, öğreticinin performansını ve öğretim sürecindeki belli bir uygulamanın, yaklaşımın ya da yöntemin etkililiğinin değerlendirilmesini içeren değerlendirmedir (Oliva, 1988). Program değerlendirme ise program geliştirme sürecinin son boyutunu oluşturmaktadır. Amaçlanan davranışların gerçekleşme düzeylerini saptama ve uygulama sonucunda programın aksayan yönlerini belirleme ve gerekli düzeltmeleri yapma program değerlendirmenin içeriğini oluşturur (Yaşar, 1998).

Eğitimde değerlendirme çalışmalarının oldukça eski bir geçmişi vardır. İlk bulgular değerlendirmenin, M.Ö 2000’li yıllarda Çinli yöneticilerin saray görevleri için yaptıkları sınavlarla adayların yeterliklerini belirleme amacıyla kullanıldığıdır (Shadish, Cook ve Leviton, 1995). Yine Eski Yunan felsefe bilginlerinden Sokrat’ın öğrenme sürecinin bir parçası olarak değerlendirmeyi sözel olarak yaptığı bilgisi değerlendirme ile ilgili ilk bulgular arasında yer almaktadır (Fitzpatrick, Sanders ve Worthen, 2004). Ancak sistematik program değerlendirme etkinliklerinin örgün eğitimin yaygınlaşması ve test tekniklerinin gelişmesi ile başladığı söylenebilir (Stufflebeam, Madaus ve Kellaghan, 2002).

1792 yılında William Farish sınavları değerlendirmek için ilk kez *nicel* puanlar kullanmıştır. O güne kadar öğrenci değerlendirmeleri için genelde “doğru” sözcüğü ile kullanılan nitel değerlendirmelerin yerini alan nicel değerlendirmeler öğrencilerin başarıya göre sıralanmasına, puanların ortalamalarının alınmasına olanak vermiştir. Bu, psikometri alanındaki ilk önemli gelişmedir ve bu gelişme geçmişten günümüze program değerlendirmeye çok önemli katkılar sağlamıştır (Stufflebeam, Madaus ve Kellaghan, 2002). 1800’lü yıllarda Büyük Britanya’da eğitim ve sosyal programlardaki bazı sorunlar, hükümet tarafından lordlardan oluşan bir komisyonca ilgili kurumların değerlendirilmesiyle giderilmeye çalışılmıştır. Yapılan değerlendirme sonucu yeni programlar uygulamaya konulmuş ve her yıl programları uygulayan okullar müfettişlerce değerlendirilmiş ve sonuçlar raporlaştırılmıştır.

Amerika Birleşik Devletleri’nde okul başarılarının değerlendirilmesine yönelik ilk uygulama 1845’te Boston’da gerçekleştirilmiştir. Bu uygulama bir okulun veya bir programın etkililiğinin değerlendirilmesinde temel veri kaynağı olarak öğrenci test sonuçlarının kullanılmasının da başlangıcı olmuştur (Stufflebeam, Madaus ve Kellaghan,

2002). Alan uzmanlarınca ilk formal eğitim programı değerlendirmesi olarak kabul edilen çalışma ise 1887 ve 1898 yılları arasında ABD’de Joseph Rice tarafından yürütülmüştür. Rice, farklı bölgelerdeki birkaç okulda yürüttüğü bu çalışmasında “telâffuz öğretimi”nde alıştırmaların etkililiğinin belirlenmesi üzerine karşılaştırmalı bir araştırma yapmıştır. Rice, öğretimin etkililiğinin belirlenmesi için test puanlarını ölçüt olarak kullanmış ve böylece değerlendiricilerin deneysel, nicel ve karşılaştırmalı araştırma yapmaları gerektiğini de ortaya koymuştur (Stufflebeam, Madaus ve Kellaghan, 2002).

1900’lü yıllar personel değerlendirmesine ilişkin Fredrick Tylor’un çalışmalarıyla başlar. Tylor’un bu çalışmaları daha sonraki yıllarda yönetim ve endüstri tabanlı değerlendirme yaklaşımlarının en büyük dayanağı olmuştur (Stufflebeam, Madaus ve Kellaghan, 2002). Yine 1910 yılında Flexner, Kanada ve Amerika Birleşik Devletleri’nde 155 tıp okulunu kendi çabalarıyla değerlendirip ilk akreditasyon uygulamasını yürütmüştür. Bu yıllarda Cronbach ve arkadaşları da okulların yönetimini ve ortamlarını ve yerel belediyelerdeki sorunları araştırmışlardır (Fitzpatrick, Sanders ve Worthen, 2004).

1930’lu ve 1940’lı yıllarda Amerika Birleşik Devletleri’nin yarısında eyalet düzeyinde sınavlar, ölçüt dayanaklı testler, standartlar, başarı testleri, birey ve ilgi testleri oldukça yaygınlaşmıştı. Öte yandan “ölçme” ve “değerlendirme” kavramları aynı anlamda kullanılmakta, özellikle “değerlendirme” öğrencilerin test başarılarının ve ödev notlarının özetlenmesi olarak düşünülmekteydi. Bu yıllarda Tyler değerlendirmeyi farklı bir boyuta taşımış, değerlendirmede yeni standartlar geliştirerek ölçüte dayalı değerlendirmeyi geliştirmiştir. Tyler öncelikle eğitim programı ve öğretim programına ilişkin bilgileri inceleyerek bu sözcükleri kavramsallaştırmıştır (Stufflebeam, Madaus ve Kellaghan, 2002). Tyler’ın 1940’lı yıllarda geleneksel ve gelişimci iki farklı okul türünü karşılaştırdığı çalışması program değerlendirme alanında ilk ve kapsamlı değerlendirme çalışması olarak kabul edilmiştir. Tyler bu çalışmasıyla değerlendirmeyi, amaçlanan çıktılar ile gerçek çıktıların karşılaştırması olarak kavramsallaştırmıştır (Stufflebeam, Madaus ve Kellaghan, 2002).

İkinci Dünya Savaşı’ndan sonra değerlendirme alanında bir duraklama olmuşsa da daha sonraki yıllarda hükümetlerin ve federal yönetimlerin de katkılarıyla hız kazanmıştır. Amerika Birleşik Devletleri, 1950’li yılların sonunda Rusya’nın Sputnik uzay aracını

yörüngeye oturtmasıyla harekete geçmiş ve eğitim sistemini gözden geçirerek yeni programlar uygulamaya koymuştur. Yönetim, bu programların sürekli olarak geliştirilmesi için bir bütçe de ayırmıştır. 1960 ve 70'li yıllarda değerlendirmeye ilişkin birçok görüş ortaya atılmış ve bu görüşler çerçevesinde birçok farklı değerlendirme modeli geliştirilmiştir. Değerlendirme üzerine yapılan tartışmalar ve tanımlamalar 1970'li yıllardan sonra program değerlendirmenin bir meslek olarak kabul edilmesini sağlamış ve programların belli ilkelere göre değerlendirilebileceği standartlar oluşturulmuştur (Fitzpatrick, Sanders ve Worthen, 2004).

Türkiye'de program değerlendirme çalışmalarına bakıldığında ilk çalışmaların 1944 yılında ilkokul programlarının değerlendirilmesine yönelik olarak yapıldığı görülmektedir. Yürütülen değerlendirme çalışmalarının amacı, o dönemde Türkiye'de beş yıllık şehir ilkokulu ve üç yıllık köy ilkokullarında uygulanan iki farklı eğitim programının uygulamadaki eksikliklerini, aksaklıklarını ve sorunlarını belirlemek ve ortaya çıkan sonuçlar doğrultusunda yeni ve tek bir eğitim programını geliştirmektir. Bu amaçla 1944 yılında Milli Eğitim Bakanlığı (MEB) Talim ve Terbiye Dairesi tarafından her iki tür programın değerlendirilmesi için öğretmenlere bir anket uygulanmış; müdürler ve müfettişlerle de görüşmeler yapılmıştır (Arslan, 2000). Öğretmenlere uygulanan ankette öğretmenlerden aşağıdaki konularda görüş ve önerilerini belirtmeleri istenmiştir:

- Her iki programın genel olarak eksik ve fazla yönleri
- Programların uygulanmasında karşılaşılan güçlükler
- Programların köy ve şehir ilkokulları için yeni ve tek bir program hâline getirilmesinin uygunluğu
- Yeni programa eklenecek dersler ve konular ile programdan çıkarılacak konular
- Köy ve şehirlerdeki ilkokullar için ayrı ayrı programlar yapılmasından vazgeçildiği takdirde çevrenin özelliklerine göre köyde, kasabada ve şehirde üzerinde durulması gerekli ders konuları ile her üç çevreye göre öğrencilerin gereksinimleri gözetilerek eklenecek yeni konular

Anket sonucunda köy ve şehir ilkokulları eğitim programlarının birleştirilmesi kabul edilerek 1948 İlkokul Programı geliştirilmiş ve program 1948-1949 öğretim yılı başından

itibaren uygulanmaya başlanmıştır. Yeni programın geliştirilmesinde öğretmenlere verilen anket sonuçlarından yararlanılması, ilköğretim müfettişleri ile müdürlerden görüş toplanması ve bu kişilerden oluşan komisyon tarafından incelemeler yapılması, program geliştirme ve değerlendirme çalışmalarının bilimsel tabana oturtulmaya çalışıldığının göstergeleridir (Gözütok, 2003). Ancak yürütülen bu çalışmalar ayrıntılı incelendiğinde, Türkiye’de program geliştirme ve program değerlendirmenin 1950’li yıllara kadar daha çok derslerin ve konuların programa eklenmesi ve çıkarılması olarak algılandığı da görülmektedir (Demirel,1992).

Program geliştirme ve değerlendirme çalışmalarının ilk kez sistematik bir biçime getirilmesi 1952 yılında Türkiye’ye gelerek köy okullarında incelemeler yapan K.V.Wofford’un hazırladığı raporla gerçekleşmiştir. Bu raporlarla ilgili olarak 1952 yılında 25 öğretmen, uygulamalar hakkında bilgi edinmek için Amerika’ya gönderilmiştir. 1953 yılında V. Millî Eğitim Şûrası’nda, uygulanmakta olan 1948 programının, günün değişen gereksinimlerine yanıt verebilecek şekilde yeniden gözden geçirilmesi üzerinde durulmuştur. Şurada programın yeniden gözden geçirilmesi konusunda şu görüşler dile getirilmiştir (Çelenk, Tertemiz ve Kalaycı, 2000):

- Programın amacı ve ilkeleri ile içeriği arasında uyumsuzluk vardır.
- Öğretmene konuların öğretiminde yeterince esneklik tanınmamaktadır.
- Dersler çok çeşitlidir.
- Derslerdeki konular öğrencilerin gelişim seviyelerine uygun değildir.
- Dersler ve konular arasında ilişki kurulamamaktadır.
- Haftalık ders saatleri ders çeşitliliğinden dolayı yeterli değildir.
- Değerlendirme sistemi yalnızca bilgiye dayanmaktadır.

Bu görüşler doğrultusunda Türkiye için bir çerçeve programın hazırlanması ve bu programın uygulanmaya konulmadan önce çeşitli bölgelerden seçilen deneme okullarında uygulanması, değerlendirilmesi ve geliştirilmesi kararları alınmıştır. Bu amaçla, 1948 İlkokul Programı temel alınarak hazırlanan taslak program, 1953-54 öğretim yılında Bolu ve İstanbul’da deneme okullarında uygulanmıştır. Bu çalışmalara paralel olarak 1954-55 öğretim yılında İstanbul Atatürk Kız Lisesi’nde deneme okulu program komisyonu

tarafından geliştirilen deneme okulu programı da uygulanmaya başlanmıştır. Bu çalışmalar, araştırma ve değerlendirme kavramlarının program geliştirme ve uygulamalarında daha sık yer almasından dolayı oldukça önemlidir (Demirel, 1992; Gözütok, 2003).

1961-1962 öğretim yılına kadar “Deneme Okulları Programları”nın geliştirilmesine, denenmesine devam edilmiş ve deneme çalışmaları değerlendirilmiştir (Çelenk, Tertemiz ve Kalaycı, 2000). Deneme okullarında yapılan bu çalışmalar, 1957 yılında çok az bir değişiklikle yeni baskısı yapılan 1948 İlkokul Programının yeniden geliştirilmesi gereğini ortaya koymuştur. Bu amaçla, Talim Terbiye Dairesi, İlköğretim Genel Müdürlüğüyle işbirliği yaparak, ilkokul, ortaokul, öğretmen okulu öğretmenleri ile uzmanlardan oluşan 16 kişilik bir komisyon oluşturmuştur. Bu komisyon 1948 İlkokul Programında yapılacak değişikliklerle ilgili bir rapor hazırlayarak Bakanlığa sunmuş ve Bakanlık raporu, tüm illere, ilçelere, öğretmen derneklerine, basına, ilgili resmi ve özel kurumlara göndererek değerlendirmelerini istemiştir. Bakanlık aynı zamanda bir genelge ile Milli Eğitim Teşkilâtı, öğretmen dernekleri, okul aile birlikleri ve öğretmenlerden de programı değerlendirmelerini istemiştir. Program paydaşlarının tümünün katılımıyla gerçekleşen bu program değerlendirme çalışmaları sonucu 1962 yılında “1962 İlkokul Program Taslağı” geliştirilmiş ve beş yıl süre ile deneme okullarında uygulanmıştır. Deneme uygulamaları yapılan taslak program Merkez Değerlendirme Komitesi’nce görevlendirilen 120 kişilik bir komisyon tarafından sürekli olarak değerlendirilmiş ve 1968-1969 Öğretim Yılı’nda uygulamaya konulmuştur (Çelenk, Tertemiz ve Kalaycı, 2000). Bu sistematikte geliştirilen bu program yaklaşık 1990’lı yıllara kadar uygulanmıştır (Fer, 2005).

1970’li yıllarda gelindiğinde 1973 yılında çıkarılan 1739 sayılı Millî Eğitim Temel Kanunu ile Türk Millî Eğitim Sistemi yeni baştan düzenlenmiştir. Bu kanunla birlikte ilköğretim, beş yıllık ilk ve üç yıllık ortaokullar birleştirilerek “8 yıllık Temel Eğitim” adını almıştır. Bu yeni kurumlar için program geliştirme çalışmaları gündeme gelmiş ve bu çalışmalara hız verilmiştir. Dokuz üyeden oluşan çalışma grubu sekiz yıllık ilköğretim okullarının amaçlarını ve eğitim ilkelerini belirlemiş, ancak bu çalışmalar denemeden öteye gidememiştir (Demirel, 1992). Bu dönemde daha önce geliştirilmiş olan ilköğretim ve ortaöğretim programlarının uygulamasına devam edilmiştir

1980'li yıllarda program geliştirme çalışmalarında yeni bir arayış başlamıştır. Bu arayış öncelikle program geliştirmede model oluşturmaya yöneliktir. Bu amaçla MEB, üniversitelerdeki bilim adamlarıyla işbirliği içerisinde yeni bir program modeli oluşturmuştur. Bu modelde programların hazırlanması ve geliştirilmesi konusunda görev alacak kişiler ile program geliştirme grubunun çalışma esasları belirlenmiş ve her programda genel, ünite ve konu amaçlarının belirlenmesinin, her ünitenin ayrı ayrı davranışlarının tespit edilmesinin gerekli olduğunun altı çizilmiş, programların bir yıllık uygulanmasından sonra değerlendirilmesinin yapılarak, değerlendirme sonuçlarına göre programların geliştirilmesi karara bağlanmıştır (Gözütok, 2003). Ancak bu model konusunda bağlayıcı bir karar bulunmaması nedeniyle programlar dersler düzeyinde farklı modellerle hazırlanarak geliştirilmiştir. Böylelikle ders programlarının geliştirilmesinde standartlaşma yerine çeşitliliğe doğru gidilmiştir.

Türkiye'de program geliştirme çalışmaları, 1990 yılında başlatılan Dünya Bankası desteğindeki Millî Eğitimi Geliştirme Projesi ile önemli gelişmeler göstermiştir. Bu projesi çerçevesinde Eğitimi Araştırma ve Geliştirme Dairesi (EARGED) tarafından 1993 yılında yeni bir program geliştirme modeli geliştirilmiştir. Programın alanda denenmesi ve sonrasında değerlendirilmesini de temel alan bu modelin amacına ulaşması için Millî Eğitim Bakanlığı bir yıl sonra Müfredat Laboratuvar Okullarını (MLO) kurulmuştur.

MLO'nun kurulmasının temel amacı, taslak olarak hazırlanan öğretim programlarının alanda denenerek süreçte karşılaşılan sorunlarının ve eksikliklerinin giderilmesi ve tüm ülke düzeyinden uygulanmaya başlanmasından önce son kez programın değerlendirilmesidir. Bu yönüyle MLO okullarının iki temel işlevinden söz edilebilir:

- Taslak öğretim programının niteliğini arttırmaya dönük önlemler almak (biçimlendirici program değerlendirme işlevi)
- Taslak öğretim programının ülke genelinde uygulanmasından önce programın niteliğini/sağlamlığını onaylamak (toplam program değerlendirme işlevi)

Bu model ile birlikte program geliştirme sürecinin daha ayrıntılı olarak ele alındığı söylenebilir. Örneğin, 1998 Hayat Bilgisi Öğretim Programı ve 1998 Matematik Öğretim Programı bu modele göre geliştirilmiştir.

Görüldüğü gibi, 1980 yılından sonra eğitim programlarının toplu geliştirme uygulaması yerini tek tek dersler bazında program geliştirme uygulamasına bırakmıştır. Bu uygulama beraberinde değerlendirme çalışmalarının da tek tek ders bazında yapılmasını getirmiştir (Çelenk, Tertemiz ve Kalaycı, 2000). Ancak yapılan değerlendirme çalışmalarının merkezden değil, genellikle üniversitelerde görev yapan öğretim elemanlarınca bireysel olarak ve amaçların değerlendirilmesine dayalı olarak yürütüldüğü görülmektedir (Yaşar ve diğ., 2005).

1968 İlkokul Programlarının toplu biçimde geliştirilmesinden sonraki ilk toplu program geliştirme çalışmaları 2004-2005 yılında MEB tarafından yürütülen Yeni İlköğretim Programlarının geliştirilmesidir. Yeni İlköğretim Programlarının geliştirilmesinde, önceki programın uygulayıcılarından ve uygulama sonuçlarından bilimsel bir süreç içinde dönütler alınmamasına karşın bu programların bir yıl süreyle pilot uygulamaları yapılmış ve bu uygulamalar değerlendirilmiştir. 2004-2005 yılında taslak olarak geliştirilen ve aynı öğretim yılında dokuz ilde bulunan 120 İlköğretim okulunda pilot uygulaması yapılan programların değerlendirme çalışmalarını EARGED yürütmüştür (MEB, 2005).

Programların pilot uygulamalarının değerlendirilmesine ilişkin raporlar incelendiğinde, her ne kadar program geliştirme süreci yönünden program geliştirme ilkelerine uygun olarak geliştirilmemiş olsa da yeni programların değerlendirme çalışmalarının bilimsel temeller çerçevesinde kapsamlı olarak ele alındığı söylenebilir. Programların değerlendirilmesi için gerekli veri toplama araçlarının (anketler ve bilgi formları) danışman akademisyenler tarafından hazırlanması, uygulamada ilgili müfettişlerin, okul müdürlerinin, öğretmenlerin ve velilerin görüşlerinin alınması, bulguların raporlaştırılarak paydaşlarla paylaşılması vb. yönlerden program değerlendirme ilkelerini taşıdığı söylenebilir (MEB, 2005).

Genel olarak Türkiye’de program değerlendirme çalışmalarının bir değerlendirmesi yapıldığında, değerlendirme çalışmalarının, özellikle belirli dönemlerde program geliştirme çalışmalarıyla bir bütünlük içinde yürütüldüğü söylenebilir. Örneğin, 1948 İlkokul Programının geliştirilmesinde önceki programların değerlendirme sonuçlarından yararlanılması, 1962 Taslak İlköğretim Programlarının geliştirilmesinde 1950’li yıllardan beri denenen bazı uygulamaların değerlendirmelerinin dikkate alınması, yine bu taslak

programın da altı yıl süreyle alanda denenmesi ve değerlendirilmesi sonucu uygulamaya konulması, 2006 yılında ülke düzeyinde uygulamaya konulan Yeni İlköğretim Programlarının uygulama öncesi pilot uygulamalarının değerlendirilmesi önemli örneklerdendir. 1980’li yıllardan 2000’li yıllara kadar olan dönemde ise programların tek tek geliştirildiği bu nedenle toplu bir program değerlendirme çalışmasının yapılmadığı görülmektedir. Bu sürede yapılan değerlendirme çalışmalarının öğretim elemanları tarafından bireysel olarak yürütüldüğü, ancak bu çalışmalarının sonuçlarının yeni program geliştirme çalışmaları için öneriden öteye geçmediği görülmektedir. Ayrıca bu çalışmaların sonuçlarının programların geliştirilmesine dönük karar vericiler tarafından da dikkate alındığını söylemek güçtür.

1.3.1. Program Değerlendirmenin Amacı

Eğitim kurumlarının en önemli görevlerinden biri toplumun, gelişen dünyanın, bilim ve teknolojinin istemlerini karşılayabilen özelliklere sahip insanları yetiştirmektir. Örgün ve yaygın eğitim kurumları bu işlevini, nitelikli programlar hazırlayarak, bu programları uygun koşullarda uygulamaya koyarak ve programlarını sürekli olarak geliştirerek yerine getirebilir. Eğitim sistemleri, gelişmelere uyum sağlayabilecek ve gelişmeleri yönlendirebilecek bireyleri, ancak ülkenin kalkınma amaçlarıyla tutarlı eğitim programları uygulayabildikleri zaman yetiştirebilirler. Çünkü programlar; eğitilecek bireylere kazandırılacak davranış standartlarını, öğretilecek bilgi, beceri ve değerleri belirler, öğretme-öğrenme etkinliklerine ve değerlendirme çalışmalarına rehberlik eder. Demirel’in (2006) de belirttiği gibi, Milli Eğitim Politikasının uygulamaya geçişi ancak eğitim programlarının aracılığı ile sağlanabilir.

Eğitimin niteliği eğitim programlarının niteliği ile özdeştir. Çünkü nitelikli bir eğitim ancak nitelikli eğitim programları yoluyla sağlanabilir. Bir ülkenin, bir kurumun eğitim programlarına bakılarak bu programların uygulanması sonucunda ortaya çıkacak eğitim çıktıları konusunda yordamlar yapılabilir. Kuşkusuz bir programın değeri uygulamadaki başarısıyla doğru orantılıdır. Bireyin ve toplumun gereksinimleri ile bilim ve teknolojide meydana gelen değişme ve gelişmelere koşut olarak programlarını geliştiremeyen toplumlar ya da ülkeler yurttaşlarını gelecekte için yetiştirmiyorlar demektir. Bu açıdan bir eğitim sistemi, sistemin başarısından emin olmak ve uyguladığı programların toplumun istemlerini karşılayıp karşılamadığını kontrol etmek üzere programlarını sürekli

olarak deęerlendirmeli ve deęerlendirme sonularını program geliřtirme alıřmalarına yansıtmalıdır (Gözütok, 2001).

Program deęerlendirmeye iliřkin tek bir tanım yapılamayacağı gibi, program deęerlendirmenin amacına iliřkin de tek bir amatan söz etmek olanaksızdır. Deęerlendirme kuramcıları farklı ortamlarda farklı deęerlendirme amalarından söz etmişlerdir. Genellikle program deęerlendirmenin yapılma nedeni programın geliştirilmesi ve programın devam ettirilmesi veya sonlandırılmasına iliřkin karar verilmesidir. Ancak deęerlendirme olmaksızın programın etkili olup olmadığı, öğrencinin başarılı olup olmadığı veya eğitime harcanan paraların iyi kullanılıp kullanılmadığı bilinmez (Sanders ve Nafziger, 1976). Bu nedenle, program deęerlendirmenin eğitimin tüm türlerinde ve alanlarında yapılması gerektiğini ifade eden Sanders ve Nafziger'e (1976) göre deęerlendirme řu amalarla yapılır:

- Programın güçlü ve zayıf yanlarını belirlemek ve geliřtirmek
- Programda karşılaşılan sorunları, özümü güçleřmeden önce belirlemek
- Eğitim gereksinimlerini belirlemek
- Eğitimde kullanılacak kaynakları belirlemek
- İstenen eğitsel çıktıları belirlemek
- Planlama ve karar vermede kullanılacak bilgileri toplamak
- Eğitim giderlerini azaltmada kullanılabilir finansal bilgileri toplamak

Worthen, Borg ve White, (1993) program deęerlendirmedeki temel amacın, programın devam etmesine, sonlandırılmasına veya gözden geçirilmesine karar vermek olduğunu ifade etmekte ve program deęerlendirme yoluyla programın deęerinin ortaya konulduğunu belirtmektedir. Program deęerlendirmenin birçok amaca hizmet ettiğini, ancak temel hedefinin programın deęerini ve yeterliğini ortaya koymak olduğunu belirten Scriven (1967), 1994 yılında deęer ve yeterlik kavramlarına nitelik kavramını da eklemiřtir. Talmage (1982), program deęerlendirmenin programın aksayan yönünü ortaya koymak, bu doęrultuda politikalar oluřturarak karar vericiye yardımcı olmak amacıyla yürütüldüğünü belirterek program deęerlendirmenin amacının, programla ilgili işlevsel politikalar oluřturmak olduğunu vurgulamaktadır.

Chelimsky ve Shadish (1997) program deęerlendirmenin amacını, programa bakış açısı kazandırmak, programın içeriğini geliřtirmek řeklinde açıklamakta ve ona gre yeni yzyılda dnyadaki deęiřimleri yakalamak; vrenin korunması, yeni teknolojiler, geliřim, insan hakları, terrizm karřıtı olma, nfus yapısı, gler gibi nedenler program deęerlendirmenin amacını ortaya koymaktadır.

Erden'e (1998) gre deęerlendirme iki amaca dnk olarak yapılır. Bu amalar řunlardır:

- ęrencinin bařarısını deęerlendirerek bir dersin hangi ęrenciler tarafından tekrar edilmesi gerektięine karar vermek
- Eęitim programlarının etkililięi hakkında yargıda bulunmak ve programdaki aksaklıkların hangi ęe ya da ęelerden kaynaklandığını belirleyerek gerekli dzeltmelerin yapılmasına olanak saęlamak

Yıldırım'a (1999) gre, deęerlendirmenin temel amacı, eęitim alıřmalarının etkililik derecesini belirlemek ve bu etkililięi arttırıcı nlemlere zemin hazırlamaktır. Bu amalar dıřında Yıldırım (1999), deęerlendirmenin ařaęıda verilen beř farklı amacından sz etmektedir:

- Araların geerlięini ve uygunluęunu saęlamak
- Varsayımları doęrulamak
- ęrenme zorluklarını tanımak
- İlgiyi canlandırmak
- Rehberlik etmek

1.3.2. Program Deęerlendirmenin İřlevi

Program deęerlendirmenin yukarıda sz edilen amalarının yanı sıra farklı iřlev ve kullanımları vardır. Eisner deęerlendirmenin beř temel iřlevi ve buna baęlı kullanımları olduęunu vurgulamaktadır (Morrison, 1993):

- *Tanı Koyma İřlevi:* Program deęerlendirme, bir eęitim programının ğretme-ęrenme srecinde yařanan eksikliklerini ortaya ıkarmada kullanılır.

- *Gözden Geçirme İşlevi:* Program değerlendirme, bir eğitim programının gözden geçirilerek programın geneliyle ilgili dönütler sağlar.
- *Karşılaştırma İşlevi:* Program değerlendirme, bir eğitim programının eş değer bir programla maliyet, etkililik vb. yönlerden karşılaştırmasında kullanılır.
- *Eğitimsel Gereksinimleri Belirleme İşlevi:* Program değerlendirme, hedef kitlenin eğitsel gereksinimlerinin saptanmasında kullanılır.
- *Amaçlara Ulaşıp Ulaşılmadığını Saptama İşlevi:* Program değerlendirme, bir eğitim programının gereksinimler doğrultusunda belirlenmiş amaçlarına ulaşıp ulaşılmadığını belirlemede kullanılır.

McCain (2005) ise değerlendirme için toplanan bilginin analiz edildikten sonra ortaya çıkan sonuçların şu alanlarda kullanılması gerektiğini ifade etmektedir:

- Öğrenme yaşantılarını tasarlama
- Amaçların gerçekleşme düzeylerini saptama
- İçeriğin uygunluğunu belirleme
- Öğretim stratejilerin etkililiğini ve uygunluğunu belirleme
- Öğrenmeyi arttırma
- Öğretmene dönüt sağlama
- Öğretimde uygun hızın ve etkinliklerin sıralamasını belirleme
- Katılımcılara dönüt sağlama
- Gelecekteki programlara kimlerin katılabileceğini belirleme

Kimi konu alanı uzmanları da program değerlendirmenin amaçlarını doğrudan tanımlamak yerine, program değerlendirmenin sonuçlarının kullanıldığı alanları ya da program değerlendirmenin sonuçlarını kullanacak kişileri gruplayarak program değerlendirmenin amacını ve işlevini dolaylı olarak açıklamaktadırlar. Örneğin, Fitzpatrick, Sanders ve Worthen (2004) değerlendirmenin şu alanlarda kullanılabileceğini ifade etmektedir:

- Okullardaki ders programlarının niteliği hakkında yapılacak çalışmalarda
- Düşük düzeyde de olsa okulların akredite edilmesinde

- Okulu mali yönden destekleyen kurumların eğitim programının etkililiğine ilişkin bilgilendirmede
- Ailelerin ve öğrencilerin daha etkili bir okul seçimi yapabilmelerinde
- Öğretmenlerin mesleki gelişimleri ile ilgili eksikliklerini görmelerinde ve bu eksiklikleri gidermelerinde

Değerlendirmenin kullanılacağı bu alanların yanında Fink (1995), değerlendirme sonuçlarını kimlerin kullanabileceğine ilişkin beş farklı gruptan söz etmektedir. Fink, bu grupları şöyle tanımlamaktadır:

- Devlet ya da yerel hükümetler (örn. şehirlerin, eyalet devletlerinin eğitim birimleri)
- Şehirlerdeki okullarda ve sosyal hizmet birimlerinde program geliştirme çalışmalarında görev yapan program geliştiriciler
- Devletin ya da hükümetin eğitimle ilgili alt birimlerinde çalışan karar vericiler
- Programları mali yönünden destekleyen kurumlar
- Hükümetlerde, üniversitelerde ve diğer kurumlardaki araştırmacılar

1.3.3. Program Değerlendirme Süreci

Program değerlendirme, yalnızca program uygulaması sonunda yapılan gelişmiş güzel bir işlem olmayıp, programı geliştirmek amacıyla verilerin bir araya getirildiği ve yargıya varıldığı bir süreçtir. Bu yönüyle değerlendirme bir süreçtir ve bilimsel araştırma süreçlerinde izlenen işlemler program değerlendirme sürecinin de basamaklarını oluşturur. Ancak bilimsel araştırma ile program değerlendirme birebir örtüşmez. Her ne kadar bazı değerlendirme çalışmaları sosyal bilimler araştırma geleneğinden etkilenmişse de değerlendirme ile bilimsel araştırma arasında belirgin farklar vardır.

Program değerlendirme ile araştırma arasındaki ilk farklılık amaçlarından kaynaklanmaktadır (Fitzpatrick, Sanders ve Worthen, 2004). Araştırma, alana bir yenilik getirmek ve yeni bir kuram ortaya koymak amacı taşırken, değerlendirmede programın gelişime katkıda bulunmak, program hakkında bir karara varmak ve programın tasarımını ya da uygulama sürecinin değerlendirilerek var olan durumu ortaya koymak

amaçlanmıştır. İkinci temel farklılık ise araştırmanın sonuçlar üzerinde odaklanması, değerlendirmenin ise daha çok yargılara odaklanmasıdır. Araştırmanın sonuçlar üzerinde odaklanması; sonuçların genelleştirilebilmesini sağlar, ancak bu durum değerlendirme için geçerli değildir. Program değerlendirme sonuçları yalnızca başka çalışmaların sonuçları ile karşılaştırmada kullanılabilir. Öte yandan değerlendirmede varılan sonuçlara dayalı olarak verilen kararlar başka programlara ilişkin olarak genellenemezler. Araştırmaların doğruluğu için genelde iç geçerlik ve dış geçerlik çalışmaları kullanılır. Ancak değerlendirme çalışmalarının doğruluğunun belirlenmesi için bu ölçütler yeterli değildir. Program değerlendirme çalışmaları doğruluk, uygunluk, yararlık ve yürütülebilirlik yönünden değerlendirilirler. Bunun içinde program değerlendirme için geliştirilen ölçütler ve standartlar kullanılır (Fitzpatrick, Sanders ve Worthen, 2004).

Araştırma ile değerlendirme arasındaki diğer bir fark, araştırmacının yetiştirilmesi ile değerlendiricinin yetiştirilmesindedir. Araştırmacı yalnızca bir alanda derinlemesine bir eğitim sürecinden geçerek yetiştirilmekte, değerlendirici ise paydaşların ve kullanıcıların farklı alanlarda farklı bilgi gereksinimlerini karşılamakla sorumlu olduğu için disiplinler arası bir anlayışla yetiştirilmektedir. Program değerlendirme bir yargıda bulunma sürecidir. Araştırma ise; bu yargıda bulunabilmek için veri toplama sürecidir (Oliva,1988). Değerlendirme türleri içerik, girdi, ürün ve süreç değerlendirme olarak sınıflandırılırken; araştırma türleri eylem, betimsel, tarihsel ve deneysel olarak sınıflandırılmaktadır (Oliva, 1988).

Program değerlendirme çalışmalarının bilimsel araştırmada olduğu gibi sistematik bir sürece dayandırılması değerlendirmenin daha kapsamlı bir hale getirilmesine de yardımcı olur. Birbiri ile yakından ilişkili işlemlerin olduğu program değerlendirme süreci planlama, uygulama ve değerlendirme olmak üzere üç aşamadan oluşur. Program değerlendirme sürecine başlamadan önce bu aşamaları yürütecek olan değerlendiricinin seçimi de oldukça önemlidir. Program değerlendirme sadece program tasarımcısı, öğretmen ya da değerlendiricinin sorumluluğunda değildir. Liderliği kurumlar üstlenmiş olsa da, müdürler, öğretmenler hatta öğrenciler de program değerlendirmeden sorumlu olabilirler (McCain, 2005). Bununla birlikte, genel olarak program değerlendirme sürecini yürütebilecek iki grup değerlendiriciden söz edilmektedir. Bunlar dış değerlendiriciler ve iç değerlendiricilerdir (Fitzpatrick, Sanders ve Worthen, 2004).

- *Dış Değerlendiriciler:* Bağımsız, danışman, sözleşmeli değerlendiriciler olarak da tanımlanan dış değerlendiriciler program sonunda başarımları ölçen toplam değerlendirmelere daha çok katılırlar. Bu tür değerlendiriciler programlara daha uzak kaldıkları için yansız bir bakış açısına sahiptirler. Ancak bu durum bazen olumsuzluk yaratabilmektedir. Öte yandan program katılımcıları dış değerlendiricilere daha duyarlı yanıtlar verebildiğinden iç değerlendiricilere göre daha fazla tercih edilirler. Bu tür değerlendiriciler daha çok üniversitelerde, kolejlerde, danışmanlık şirketlerinde bulunurlar ve iç değerlendiricilere göre daha fazla kaynağa ulaşma olanağına sahiptirler (Fitzpatrick, Sanders ve Worthen, 2004; W.K. Kellogg Foundation, 1998). Örneğin, doktora tez jürilerinde tezin bitiminde tez savunmasına çağrılan uzman dış değerlendiriciyi temsil etmektedir.
- *İç Değerlendiriciler:* İç değerlendiriciler programın içinden ya da projeye yakından ilgili bireylerdir. Bu tür değerlendiriciler program ve süreç hakkında daha fazla bilgiye sahiptirler. Biçimlendirici değerlendirmelerde dış değerlendiricilere göre daha fazla tercih edilirler (Fitzpatrick, Sanders ve Worthen, 2004). Programla ilgili hızlı dönütlerin sağlanması, değerlendiricinin programın katılımcıları ile birebir bir ilişki içinde olması olumlu yönlerindedir, ancak çoğu zaman dış değerlendiricinin sahip olduğu teknik olanaklar ve bakış açısına gereksinim duyarlar (W.K. Kellogg Foundation, 1998). Örneğin, doktora tez izleme jürilerinde yer alan uzmanlar iç değerlendiriciyi temsil etmektedir.

Bazı değerlendirme çalışmalarında her iki değerlendirici de görev alabilmektedir. Bu tür durumlarda iç değerlendirici dış değerlendiricinin danışmanlığını ve teknik desteğini, dış değerlendirici de iç değerlendiricinin program bilgisini kullanır. Programı değerlendirecek uzmanın seçiminden sonra programın değerlendirilmesi için şu süreçler izlenir:

- *Planlama Süreci:* Değerlendirme sürecinin başlangıcı ve başarısını etkileyen en önemli aşamasıdır. Planlama aşamasında desenin (deneysel-betimsel), evren-örneklem belirlenmesi, veri toplama yöntemlerinin tanımlanması, verilerin seçimi, hangi ölçme araçlarının kaç kez, ne zaman, nasıl, hangi amaçla

kullanılacağıının belirlenmesi, bu araçların geçerlik ve güvenilirlik çalışmalarının yapılması gerekir. Değerlendirici her değerlendirme sorusunu en iyi yanıtlayacak bilgilerin hangileri olduğuna karar vermeli; uygun bilgi kaynaklarını belirlemelidir. Diğer taraftan personel gereksinimi saptanmalı ve değerlendirme bütçesi geliştirme ve değerlendirme için işbirliği anlaşmaları oluşturulmalıdır. Son olarak bir yönetim planı oluşturularak işlemlerin yerine getirilip getirilmediği kontrol edilmeli ve eksiklikler giderilmelidir (Fitzpatrick, Sanders ve Worthen, 2004). Bu işlemler şunlardır:

- *Değerlendirmenin Amacının Belirlenmesi:* Değerlendirmenin amacının önceden saptanması ve anlaşılır bir biçimde ortaya konması ve değerlendirmenin hangi kullanıma hizmet edeceği değerlendirmeye başlamadan önce yapılacak en önemli etkinliklerdendir. Amaçlar, paydaşların, program destekleyicilerinin ve değerlendirme sonuçlarını kullanacak bireylerin gereksinimleri doğrultusunda oluşturulur. Sürece başlamadan önce programın değerlendirilebilirliğinin ortaya konması gereklidir. Bu belirlendikten sonra değerlendirmenin neden yapılacağı, değerlendirme bulgularının ne amaçla kullanılacağı, kim tarafından kullanılacağı, programın hangi zaman aralığında ve hangi kapsamda değerlendirileceği, programın daha önce değerlendirilip değerlendirilmediği ve değerlendirme bütçesi gibi konular açıklığa kavuşturulur. Daha sonra değerlendirmeyi yürütecek değerlendiriciye karar verilir (iç-dış değerlendirici) (Fitzpatrick, Sanders ve Worthen, 2004).
- *Değerlendirme Sınırlıklarının ve Çerçevesinin Belirlenmesi:* Değerlendirme amaçları belirlendikten sonra değerlendirmenin sınırlıklarının belirlenmesi gereklidir. Bu, değerlendirme çerçevesini ortaya koymada önemlidir. Sınırlıkların belirlenmesi için öncelikle tüm katılımcıların gereksinimleri ve görüşleri alınarak değerlendirmenin çerçevesi ortaya konur. Daha sonra değerlendirilecek programın ayrıntılı bir tanımlaması yapılmalıdır. Bu tanımlama programın amaçları ve etkinliklerini içerebilir. Programın tanımının yapılmasında temel kaynaklar daha önceki program değerlendirmeleri, paydaş görüşleri ve raporlardır (Fitzpatrick, Sanders ve Worthen, 2004).

- *Değerlendirme Sorularının ve Ölçütlerinin Belirlenmesi:* Program değerlendirme, programın ilgili kesimler tarafından kabul edilmesine, programın devam ettirilmesine veya programın geliştirilmesi ile ilgili soruların yanıtlanması amacıyla yürütülür. Dolayısıyla sorular değerlendirmenin temelini oluşturur ve süreç içinde değerlendirmeye yön verir. Değerlendirme soruları belirlenmeden yapılan bir değerlendirmede neyin, nasıl ve ne amaçla değerlendirileceği belirsiz kalır. Kaya (1997) değerlendirmelerin başarısız olmalarının temel nedenlerinden birinin değerlendirme sorularının doğru belirlenememesi olduğunu vurgulamaktadır. Değerlendirme sorularının belirlenmesinde genellikle paydaşların soruları ve görüşlerinden, belirlenen değerlendirme modelinin özelliklerinden, değerlendirme standartları ve ölçütlerinden, farklı uzman deneyimlerinden ve son olarak değerlendirici görüşlerinden yararlanılır (Fitzpatrick, Sanders ve Worthen, 2004).
- *Araştırma Deseni, Evreni ve Örneklemine Belirlenmesi:* Araştırma desenleri araştırmanın yorumlanmasını nesnelleştirecek ve bulguların genellenebilirliği hakkındaki şüpheleri azaltacak deneysel ve istatistiksel yöntemler bileşimidir (Turgut, 1983). Diğer bir ifadeyle desenler, veri toplamaya yönelik yapı ve düzenlemelerdir (Sanders ve Nafziger, 1976). Bu aşamada değerlendirmenin planlama aşamasında belirlenen sorulara dayalı olarak hangi desenlerin uygun olduğuna karar verilir. Değerlendirici bu aşamada betimsel ya da deneysel desenlerden birini seçer. Desen belirlendikten sonra araştırmanın evreni ve örneklemini açıklığa kavuşturulur (Fitzpatrick, Sanders ve Worthen, 2004).
- *Değerlendirme Uygulamasının Planlanması:* Planlamanın bu aşamasında değerlendirici diğer tüm aşamalarda etkinliklere ek olarak uygulamaya geçmeden önce değerlendirmenin herkes tarafından daha anlaşılabilir olmasını sağlar. Bunun için de, değerlendirmenin biçimlendirici mi yoksa toplam değerlendirme mi olduğu, gereksinim çözümlemesine mi yoksa bir ürüne mi yönelik olduğunu belirler. Dolayısıyla değerlendirici, öncelikle değerlendirme deseninin betimsel mi ya da deneysel mi olacağına karar verir. Bu seçim doğrultusunda hangi yöntemlerin kullanılacağını, hangi kaynaklardan veri toplanacağını, evreni örneklemini, verilerin seçiminde ne tür

ölçütlere bakacağını, hangi çözümleme tekniklerine başvuracağını ve taraflarla yapılacak sözleşmenin içeriğini kesinleştirir (Fitzpatrick, Sanders ve Worthen, 2004). Bu sürecin sonunda değerlendirici işlemlerin kolaylıkla izlenebilmesi için bir yönetim planı hazırlar (Şekil 2).

Değerlendirme Soruları ve Amaçları	• Araştırmanın amacı ve araştırmada yanıtı aranacak sorular
Gereksim Duyulan Bilgiler	• Cinsiyet Okul puanları • Notlar Kurum türü vb
Veri Kaynağı	• Öğretmenler Öğrenciler • Veliler Müdürler vb.
Veri Toplama Yöntemleri	• Tarama Deneysel • Odak grup görüşmeleri vb.
Veri Toplama Süreci	• Kişi Araştırmacı, Kurum • Ortam Verilerin toplanacağı yer • Zaman Verilerin toplanacağı zaman
Çözümleme Süreci	• İlgili istatistiksel tekniklerin seçimi, uygulaması
Raporlaştırma	• Kime Raporun hedef kitlesi • Nasıl Raporun biçimi • Ne Zaman Raporun zamanlaması

Şekil 2. Değerlendirme Yönetim Planı Örnek Formu

Kaynak: Fitzpatrick, Sanders ve Worthen, 2004'ten uyarlanmıştır.

Şekil 2'de görüldüğü gibi değerlendiricinin hazırladığı bu yönetim planı, değerlendiricinin değerlendirmenin uygulama aşamasında sürekli olarak başvuracağı özet bir çalışmadır. Değerlendirici bu yönetim planı aracılığıyla paydaşlarla, katılımcılarla ve ilgililerle değerlendirimin amacı ve soruları, değerlendirme için gerekli bilgilerin neler olduğu, değerlendirmede başvurulacak veri kaynaklarının neler ve kimler olduğu, bu verilerin hangi yöntemlerle, kimler tarafından nerde ve ne zaman toplanacağı, toplanan verilerin çözümünde hangi istatistiksel tekniklerin kullanılacağı ve sonuçların kimlere ne şekilde ve ne zaman sunulacağı hakkında bilgileri paylaşır.

- **Uygulama Süreci:** Program değerlendiriminin uygulama aşamasında, planlama aşamasında alınan kararlar doğrultusunda desen, örneklem ve maliyete dayalı olarak verilerin toplanması, analiz edilmesi ve sonuçların yorumlanması

gerçekleştirilir. Değerlendirme bilgilerinin kullanılması ve raporlaştırılması ile değerlendirme sürecinin uygulama aşaması tamamlanır. Rapor da yer verilmesi gereken temel başlıklar özet, giriş, değerlendirme amaçları ve yanıtlanacak sorular, değerlendirme planı ve yöntemlerin belirtilmesi, bulgular, sonuç ve önerilerdir. Bu yolla süreç sistematik biçimde rapora yansıtılır (Fitzpatrick, Sanders ve Worthen, 2004). Uygulama aşamasında izlenmesi gereken işlemler şunlardır:

- *Verilerin Toplanması, Çözümü ve Yorumlanması:* Araştırmanın deseni ve evren-örnekleme belirlendikten sonra, belirlenen değerlendirme soruları doğrultusunda hangi verilerin, hangi yöntemlerle toplanacağı ortaya konur. Değerlendirici bu aşamada belirlenen verileri toplamak için nicel, nitel ya da her ikisinin bir arada uygulandığı karma yöntemleri işe koşar (Mertens, 2004). Program değerlendirme çalışmalarında genellikle doküman incelemeleri, kayıtlar, gözlemler, alan ziyaretleri, anketler, yüz yüze görüşmeler, telefon görüşmeleri, odak grup görüşmesi, yazılı yoklama sınavları gibi veri toplama teknikleri kullanılmaktadır. Veriler toplandıktan sonra verilerin özelliklerine göre çözümleme teknikleri uygulanır. Genellikle nicel veriler için betimsel ve diğer istatistiksel teknikler kullanılırken, nitel veriler için örüntü çözümlemesi, temalar, içerik çözümlemeleri yapılmaktadır. Tablo 1’de nicel ve nitel yöntemlerle toplanan verilerin çözümlenmesi için hangi istatistik tekniklerinin kullanılabileceğine ilişkin örnekler verilmiştir. Gerekli çözümlemeler yapıldıktan sonra bulgulara ilişkin yorumlamalar yapılmalıdır. Bu süreçte hedef kitlenin ve program paydaşlarının etkinleştirilmesi gereklidir (Fitzpatrick, Sanders ve Worthen, 2004).

Tablo 1.

Veri Türlerine Göre Çözümleme Teknikleri Örnek Tablosu

Veri türü	Yöntem	Çözümleme Tekniği
Görüşmeler	Nicel ve nitel	Yapılandırılmış sorular için betimsel istatistik, açık uçlu sorular için örüntü çözümlemesi
Gözlemler	Nicel ve nitel	Gözlem süresi vb. ile ilgili değişkenlerle ilgili betimsel istatistikler, etkinliklerle ilgili istenilen kısa gözlem raporlarının çözümlemeleri için içerik ve örüntü çözümlemesi
Likert Tipi Araçları	Nicel	Betimsel ve ilişkisel istatistik teknikleri
Anketler	Nicel ve nitel	Yapılandırılmış maddeler için betimsel ve ilişkisel istatistik teknikleri, açık uçlu sorular için içerik çözümlemesi
Sınav Sonuçları	Nicel	Betimsel ve ilişkisel istatistik teknikleri

Kaynak: Fitzpatrick, Sanders ve Worthen, (2004)'den özetlenmiştir.

- **Raporlaştırma:** Yukarıda sözü edilen tüm süreçlerin ve bu sürecin sonuçlarının raporlaştırılması programın geliştirilmesi, programın devamı ya da sonlandırılması kararlarının alınması, ilgili kesime programda yaşanan bir sorunla ilgili bilgi verilmesi, programa destek sağlanması, program ilişkin tutumların değiştirilmesi için değerlendirme sürecinin vazgeçilmezidir. Değerlendirme raporları, zamanlama, uzunluk, içerik ve dil yönünden hedef kitlenin beklentileri ve gereksinimleriyle örtüşmelidir. Bu nedenle sonuçlar resmi raporlar, kısa özetler, bildiri sunumları vb. farklı biçimlerde sunulabilir. Burada dikkat edilmesi gereken konu, sonuçların anlamlı bir şekilde sunumudur. Aksi halde uygun olarak toplanmış veriler değerlerini yitirebilirler (Fitzpatrick, Sanders ve Worthen, 2004). Kapsamlı bir değerlendirme raporu şu içeriğe göre oluşturulur:

- I. Özet
- II. Giriş
 - Değerlendirmenin amacı
 - Değerlendirme hedef kitlesi
 - Değerlendirme sınırlıkları
 - Rapor içeriğinin özeti
- III. Değerlendirme Odağı
 - Değerlendirilen programın betimlenmesi
 - Değerlendirme amacı ve soruları
 - Değerlendirme için gerekli bilgiler
- IV. Değerlendirme Planı ve Sürecine İlişkin Özet
- V. Değerlendirme Sonuçlarının Sunumu

- Bulguların özeti
- Bulguların yorumu
- VI. Sonuç ve Öneriler
 - Programla ilgili karar alınmasında kullanılan standartlar ve ölçütler
 - Programla ilgili kararlar
 - Öneriler
- VII. Azınlık Raporu (Kararları Benimsemeyenlerin Görüşleri)
- VIII. Ekler

- *Değerlendirme Süreci*: Program değerlendirmede son aşama olan değerlendirmenin değerlendirilmesi aşamasında, değerlendirme sürecinde yaşanan eksiklik ve hataların ortaya çıkarılması amaçlanır. Bu sürece aynı zamanda *meta-değerlendirme* de denilmektedir (Sağlam ve Yüksel, 2007). Planlanan ve planlama doğrultusunda uygulamaya konan ve raporlaştırılan program değerlendirme çalışmalarının değerlendirilmesi, değerlendirme çalışmalarının niteliğinin güvence altına alınması amacıyla yürütülmektedir (Sağlam ve Yüksel, 2007). Ayrıca meta-değerlendirme yoluyla planlama ve uygulama sürecinin güçlü ve zayıf yönlerinin belirlenmesi ve değerlendirme bulgularının güvenilirliği ile ilgili önemli bilgilerin sağlanması da olanaklıdır. Bununla birlikte, bu tür çalışmalar değerlendiricilere değerlendirme sürecinde öz eksikliklerini, zayıf kaldıkları yönlerini ortaya çıkarmada yardımcı olur. Brinkerhoff ve diğ. (1983) meta-değerlendirmenin şu amaçlarla gerçekleştirildiğini belirtmektedirler:

- Değerlendirme amaç ve sorularının değerini ve sağlamlığını değerlendirmek
- Değerlendirme deseninin sağlam olup olmadığını, uygun uygulanıp uygulanmadığını belirlemek
- Bilgi toplama yöntemleri ve bu yöntemlerle toplanan bilgilerin konuyla olan ilgi düzeyini ve niteliğini değerlendirmek
- Analizlerin ve analiz yorumlarının doğruluğu ve yeterliğini değerlendirmek
- Değerlendirme raporunu düzen, yeterlik ve yararlık yönlerinden değerlendirmek
- Yapılan harcamaların kabul edilebilirliğini belirlemek için değerlendirmenin ne kadar iyi yürütüldüğü ve ne kadar ekonomik olduğunu değerlendirmek

Bu amaçlarla gerçekleştirildiğinde meta-değerlendirmenin *biçimlendirici meta-değerlendirme* (formative meta-evaluation) ve *toplam meta-değerlendirme*

(summative meta-evaluation) olarak iki işlevinden söz edilmektedir (Stufflebeam, 1974). Biçimlendirici meta-değerlendirme, değerlendirmenin planlanması, yürütülmesi, yorumlanması ve raporlaştırılmasına ilişkin kararların alındığı değerlendirmedir. Ayrıca değerlendirmenin niteliğinin artırılması için değerlendirme sürecinin gözlemlendiği ve varsa eksiklik ve aksaklıklara ilişkin dönütler sunulduğu bir meta-değerlendirmedir. Toplam meta-değerlendirme ise amaçların, desenlerin, uygulamaların ve sonuçların bir araya getirilerek tek bir çalışma olarak ortaya çıkarıldığı bir meta-değerlendirmedir (Sağlam ve Yüksel, 2007).

1.3.4. Program Değerlendirme Türleri

Farklı amaçlar için farklı değerlendirme türleri bulunmaktadır. Değerlendirme türlerinden önce program değerlendirmeler iki yolla yürütülür: Formal ve informal değerlendirme.

- *Formal Değerlendirme:* Sitemli, yapılandırılmış, amaçları belirlenmiş nerede, ne zaman, hangi uzmanlar tarafından yürütüleceği belirlenmiş, veri toplama araçlarının güvenilirliği ve geçerliği yüksek olan değerlendirme türüdür (Fitzpatrick, Sanders ve Worthen, 2004).
- *İnformal Değerlendirme:* Formal değerlendirmenin aksine bir sistematığı olmayan ve düzensiz veri toplama teknikleri sonucu daha çok bireysel kararlardan öteye geçemeyen değerlendirme türüdür. Bu değerlendirme türü daha çok günlük yaşamdaki olaylarda ortaya çıkmaktadır. Bir okul müdürünün okulda çalışan öğretmeni gözlemlemesi ve etkililiği hakkında yargılarda bulunması, bir öğretmenin yeni ortaya atılan bir yöntemin sınıf içinde kullanılabilirliği hakkında yorumlarda bulunup karar vermesi gibi örnekler bir sistematik içerisinde yapılan değerlendirme değildir. Deneyim, içgüdü, genelleştirme ve mantık informal değerlendirme sonuçları üzerinde etkilidir. Bununla birlikte, formal değerlendirmelerin gerçekleşmesinin olanaklı olmadığı yerlerde, informal değerlendirmeler bilgili, deneyimli ve dürüst bireylerce gerçekleştirilir (Fitzpatrick, Sanders ve Worthen, 2004).

Bu değerlendirme türleri dışında program değerlendirmeler iki amaca yönelik olarak gerçekleşir: Programın geliştirilmesi ya da niteliğinin artırılması ve programın yürütülmesine ilişkin bir karar verilmesi. Bu amaçlara dönük olarak iki tür değerlendirmeden söz edilir. Bunlar, biçimlendirici değerlendirme ve toplam değerlendirmedir.

Biçimlendirici ve toplam değerlendirme kavramları ilk olarak Scriven (1967) tarafından eğitim programlarının değerlendirilmesine ilişkin bir makaleyle alana kazandırılmıştır. Scriven'in biçimlendirici ve toplam değerlendirme arasındaki farklılığı betimlediği bu çalışması, ABD'de program değerlendiriciler tarafından alana özgü ilk ve en etkili çalışma olarak kabul edilmiştir (Popham, 2006). Scriven bu çalışmasında değerlendirmenin iki temel işlevi olduğunu belirtmiştir: Niteliğini artırma/biçimlendirme, geriye dönük bilgilendirme/geleceğe dönük karar alma. Scriven'a (1967) göre, biçimlendirici değerlendirmede programla ilgili uygulamada karşılaşılan eksikliklerin, aksaklıkların ve sorunların belirlenmesine dönük bilgilerin sağlanması amaçlanmakta; toplam değerlendirmede ise uygulamanın bitiminde ya da bitimine yakın bir zamanda, biçimlendirici değerlendirmenin de sonuçları dikkate alınarak programın niteliğine dönük /geleceğine ilişkin bir kararın verilmesi amaçlanmaktadır. Programın olduğu gibi devam ettirilmesi, geliştirilmesi ya da sonlandırılması kararları toplam değerlendirme sonuçlarına göre verilir.

Bu iki kavram daha sonraki yıllarda değerlendiriciler tarafından kabul görmüş ve farklı alanlarda kullanılmıştır. Bloom (1968), bir yıl sonra biçimlendirici değerlendirme ve toplam değerlendirme kavramlarını kendi geliştirdiği tam öğrenme modelinde kullanarak bu kavramlara katkıda bulunmuştur. Ancak Bloom, biçimlendirici ve toplam değerlendirme kavramlarını "evaluation" sözcüğünün yerine "assessment" sözcüğü ile tanımlamış ve bu kavramları programların değerlendirilmesinde değil; öğrenci/öğretimin değerlendirilmesinde kullanmıştır (Bloom, Hastings ve Madaus, 1971; CERİ, 2005; Pearson, 2008). Bu nedenle Türkçe'ye çevirileri "değerlendirme" olarak yapılan bu iki kavramın işlevleri farklıdır. Bu farkın belirginleştirilmesi için "evaluation" kavramının "değerlendirme" ile "assessment" kavramının ise "öğretimin değerlendirilmesi" ile tanımlanması yerinde olacaktır. Dolayısıyla "formative evaluation-biçimlendirici değerlendirme" ile "formative assessment-biçimlendirici öğretim değerlendirmesi" ve

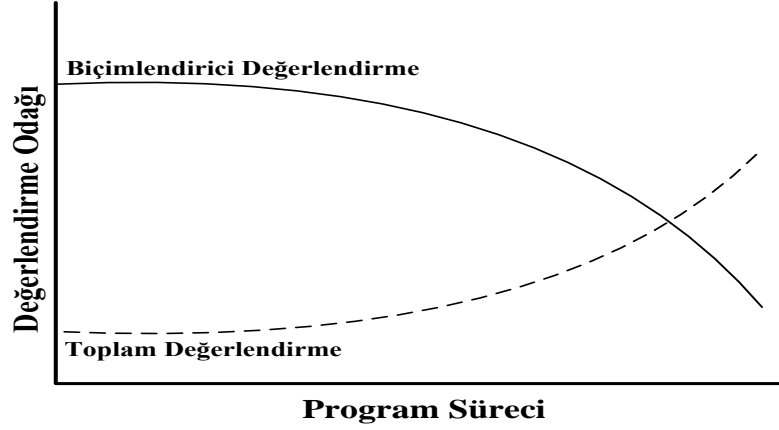
“summative evaluation-toplam değerlendirme” ile “summative assessment-toplam öğretim değerlendirmesi” kavramları arasında da aynı fark vardır (Popham, 2006).

Biçimlendirici değerlendirme ve toplam değerlendirmenin yanlış kullanımlarından biri de biçimlendirici değerlendirmenin “süreç değerlendirmesi”, toplam değerlendirmenin “ürün değerlendirmesi” olarak tanımlanmasıdır. Bu kavramları alana kazandıran Scriven (1996), biçimlendirici değerlendirmenin süreç değerlendirmesi olarak tanımlanmasının değerlendirme kapsamını daraltacağını, bu değerlendirme amacın süreci değil süreçte programın niteliğini değerlendirme olduğunu belirtmektedir. Scriven, biçimlendirici değerlendirmenin aksine toplam değerlendirmenin geniş kapsamlı bir süreç değerlendirmesi olarak tanımlanabileceğini de ifade etmektedir (Fitzpatrick, Sanders ve Worthen, 2004). Çeşitli araştırmacıların görüşlerine dayanılarak biçimlendirici değerlendirme ve toplam değerlendirmeye ilişkin daha ayrıntılı bilgiler ve bu iki değerlendirme türü arasındaki ilişki aşağıda verilmiştir.

- *Biçimlendirici Değerlendirme:* Biçimlendirici değerlendirmenin amacı, programın gelişimi ve niteliğinin artırılması için gerekli verilerin sağlanmasıdır (Stufflebeam, Madaus ve Kellaghan, 2002). Biçimlendirici değerlendirme, programın uygulanması sürecinde program yöneticilerine, uygulayıcılarına henüz program tamamlanmadan gerekli önlemlerin alınması fırsatı verir. Amaçların oluşturulmasında, önceliklerin belirlenmesinde, taslak planın uygulanmasının ve süreç içinde yapılan iç değerlendirme sonuçlarının değerlendirilmesinde biçimlendirici değerlendirmeye gereksinim duyulur. Yapılan tüm bu değerlendirmeler geliştirme sürecindeki tüm etkilere yansıtılır. Dolayısıyla değerlendirici sürekli olarak program yöneticileriyle ve uygulayıcılarıyla iletişimde kalmalı ve karar vermelerinde yol göstermelidir. Değerlendirme planı esnek ve gereksinimleri karşılayıcı olmalıdır (Stufflebeam ve Shinkfield, 2007).
- *Toplam Değerlendirme:* Program değerlendirmenin diğer türü toplam değerlendirme değildir. Programın uygulanmasının bitirilip, programın tamamlanmasından sonra geçmişe dönük değerlendirmelerin yapıldığı toplam değerlendirmede elde edilen verilerle programla ilgili olarak ileriye dönük kararların alınması amaçlanır (Stufflebeam, Madaus ve Kellaghan, 2002).

Programın devam ettirilmesi veya sonlandırılması toplam değerlendirme sonuçlarına göre yapılır. Toplam değerlendirmeler değerlendirme sonuçlarını kullanacak olan kesime başarı ve başarısızlıkla ilgili nedenler ve ürünlerin niteliği hakkında bilgilerin sağlanmasında yararlıdır (Stufflebeam ve Shinkfield, 2007).

- *Biçimlendirici ve Toplam Değerlendirme Arasındaki İlişki:* Programın geliştirilmesi veya niteliğinin artırılmasına dönük kararlar alınmasında ve programın değerinin ve geleceğinin belirlenmesinde hem biçimlendirici hem de toplam değerlendirmeye gereksinim vardır. Sağlam bir program değerlendirmesi için her ikisinin de yapılması zorunludur. Ancak genellikle program değerlendirme çalışmalarında ya toplam değerlendirme veya biçimlendirici değerlendirmeye ağırlık verilmektedir. Bu durum programın değeri, geleceği ya da geliştirilmesi yönünden bazı olumsuzluklara neden olabilmektedir. Örneğin, bir program değerlendirme çalışmasında sadece toplam değerlendirmeye odaklanılırsa, bu durum program geliştirme sürecini kısıtladığından, yetersiz ve yanlış sonuçların ortaya çıkmasına neden olmakta ve gerekli önlemlerin zamanında alınmasını ve eksikliklerin giderilmesini engellemektedir. Bu durum program değerlendirme açısından sakıncalı görülmektedir. Aynı şekilde bazen de toplam değerlendirme yerine biçimlendirici değerlendirmeye ağırlık verilmektedir. Bu durum da programın sadece uygulama sürecindeki niteliğini belirlediğinden, buna karşılık programın geleceğine ilişkin kararları ortaya koymadığından sakıncalı görülmektedir. Oysa programın değerinin bilinmesi, hangi sonuçları ortaya çıkardığının belirlenmesi, programın devam edip etmeyeceğinin tespit edilmesi için toplam değerlendirmeye gereksinim vardır. Öte yandan biçimlendirici değerlendirme ile elde edilen veriler aynı zamanda toplam değerlendirme için gerekli verilerin bir bölümünü oluşturur (Mertens, 2004). Buna göre, hangi tür değerlendirmeye daha fazla odaklanılacağı genellikle program süreci boyunca değişebilir (Şekil 3).



Şekil 3. Program Değerlendirme Sürecinde Biçimlendirici Değerlendirme-Toplam Değerlendirme İlişkisi

Kaynak: Fitzpatrick, Sanders ve Worthen, 2004.

Şekil 3'te de görüldüğü gibi genellikle biçimlendirici değerlendirme, programın geliştirilmesinin ilk aşamalarında, toplam değerlendirme ise son aşamalarında yapılmaktadır. Ancak kimi araştırmacılar tarafından her iki değerlendirmenin de yalnızca bu zaman aralıklarında uygulanmasının yanlış olduğu görüşü dile getirilmektedir (Shadish, Cook ve Leviton, 1995). Bu nedenle, bir program değerlendirme çalışmasında iki değerlendirme türünün de kullanılması önerilmektedir (Fitzpatrick, Sanders ve Worthen, 2004; Mathison, 2005; Scriven, 1991b). Biçimlendirici değerlendirme ile toplam değerlendirme arasındaki farklar Tablo 2'de gösterilmiştir.

Tablo 2.

Biçimlendirici ve Toplam Değerlendirme Arasındaki Farklılıklar

	Biçimlendirici Değerlendirme	Toplam Değerlendirme
Amaç	Niteliği belirleme	Niteliği belirleme
Kullanım	Programı geliştirme	Programın geleceği veya onaylanması ile ilgili karar verme
Hedef Kitle	Program yöneticileri ve personel	Karar vericiler, politika oluşturanlar, yöneticiler
Değerlendiriciler	İç değerlendiriciler (bazı durumlarda dış değerlendiriciler)	Dış değerlendiriciler (bazen iç değerlendirici yardımı alınabilir)
Katkısı	Programın geliştirilmesi için dönüt sağlar	Karar alıcılara programın devam etmesi veya sonlandırılması ile ilgili bilgi sağlar
Veri toplama amacı ve veri toplama sıklığı	Tanımlama/sık	Yargıda bulunma/Sık değil
Örnekleme büyüklüğü	Küçük örnekleme	Genellikle büyük örnekleme

Kaynak: Stufflebeam ve Shinkfield, 2007 ve Fitzpatrick, Sanders ve Worthen, 2004'den sadeleştirilmiştir.

Tablo 2’de görüldüğü gibi, biçimlendirici değerlendirmeler genellikle program uygulayıcıları, yöneticileri tarafından yapılırken, toplam değerlendirmeler genellikle dış değerlendiriciler tarafından yapılmaktadır. Ancak yine de biçimlendirici değerlendirme ile toplam değerlendirme arasındaki temel farklar kâğıt üzerinde olduğu gibi açıkça anlaşılamayabilir. Scriven (1991b) her iki değerlendirmenin iç içe olduğunu vurgulamakta ve toplam değerlendirme sonuçlarının hem karar verme hem de daha sonra biçimlendirici amaçla kullanılabilceğini ifade etmektedir.

Bu değerlendirme türleri dışında Türkçe alanyazında sıkça değinilen ve Bloom’un Scriven’in değerlendirme türlerinden esinlenerek ortaya attığı Tanılayıcı Değerlendirme, Biçimlendirici Değerlendirme ve Toplam Değerlendirme de vardır. Bu değerlendirme türleri daha çok öğretme ve öğrenmenin değerlendirilmesinde ya da öğrenci başarısının değerlendirilmesinde kullanılan değerlendirme türleridir. Ancak bu değerlendirme türlerinden elde edilen veriler de program değerlendirme çalışmaları için vazgeçilmez verilerdir.

1.3.5. Program Değerlendirme Yaklaşımları ve Modelleri

Bir eğitim programını değerlendirmek amacıyla veri toplama ve toplanan verileri yorumlamada değerlendiriciler farklı yaklaşımlar izlemektedir. Program değerlendiricilerin, değerlendirme konusundaki bilgi ve beceri düzeyleri, benimsedikleri değerlendirme kuramları ve felsefî değerleri, onların program değerlendirme yaklaşımlarını yapılandırmaktadır. Bu yaklaşımlar temel aldıkları öğeye göre amaca dayalı, yönetime dayalı, uzman odaklı, tüketici odaklı ve katılımcı odaklı olmak üzere beş ana başlık altında toplanmıştır. Ayrıca her yaklaşımın altında farklı araştırmacılar/alan uzmanları tarafından geliştirilen program değerlendirme modelleri yer almaktadır.

1.3.5.1. Amaca Dayalı Program Değerlendirme Yaklaşımı ve Modelleri

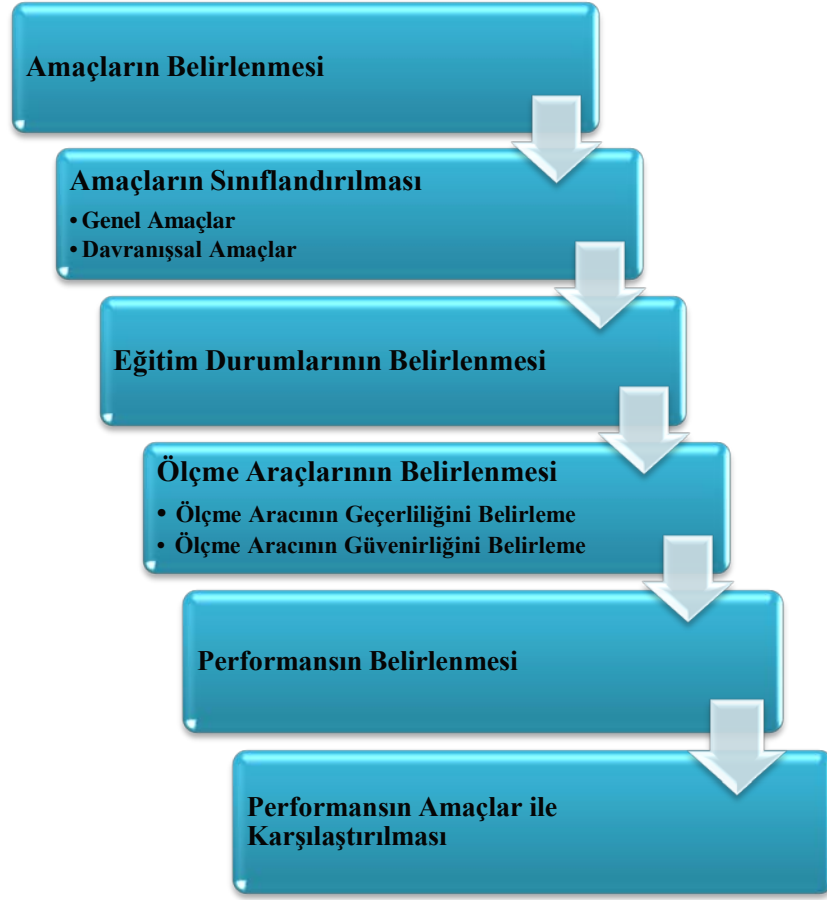
Amaca dayalı program değerlendirme yaklaşımı ve bu yaklaşımın altında yer alan program değerlendirme modelleri, program amaçlarının belirlenmesi ve bu amaçlar doğrultusunda gerçekleşen çıktılarının değerlendirilmesini temel alır. Yapılan değerlendirmeler programın amaçlarının yeniden gözden geçirilmesinde ve bu doğrultuda

kazandırılması amaçlanan davranışların şekillenmesinde, etkinliklerin ve değerlendirme süreçlerinin belirlenmesinde kullanılmaktadır (Mathison, 2005). Amaca dayalı program değerlendirme yaklaşımı kapsamında yer alan program değerlendirme modelleri alanyazında genellikle bu modellerin ilk uygulayıcıları olan Tyler, Metfessel ve Michael, Provus gibi araştırmacıların adıyla anılmaktadır.

Tyler'in Değerlendirme Modeli

1902 Chicago doğumlu olan Tyler, öğretim ve değerlendirme konusunda ilk önemli birikimini 1927 yılında göreve başladığı Ohio Devlet Üniversitesinde kazanmıştır. Bu dönemde Charles Judd ve W.W. Charters gibi dönemin önemli iki eğitimcinin düşüncelerinden etkilenen Tyler, bu birikimlerini 1930'lu yıllarda yaptığı ve ilk uzun süreli ve kapsamlı değerlendirme çalışması olarak da bilinen dört yıl ilköğretim, dört yıl ortaöğretim olmak üzere sekiz yıllık bir sürede yürüttüğü *Sekiz Yıl Çalışması*'nda kullanmıştır (Fitzpatrick, Sanders ve Worthen, 2004; Riles, 2002). 30 ilköğretim ve ortaöğretim okulu, 300 kolej ve üniversitede gerçekleştirdiği bu çalışma, değerlendirmenin eğitimde ilk kez bu kadar kapsamlı ele alınması, değerlendirmenin ilk kez belirli bir sistematik içinde yürütülmesi vb. yönlerden yirminci yüzyılın en önemli değerlendirme çalışmalarından biri olarak kabul edilmiştir (Mathison, 2005; Riles, 2002). Bu çalışma sonrası "Tylerci Değerlendirme" anlayışı oluşmuş ve Tyler kendi değerlendirme anlayışına atfedilen bu kavramı sonraki yıllarda bir görüşmede onaylamıştır (Tyler, 1981).

Tyler, program değerlendirme yaklaşımını uygulanan programın belirlenen amaçlarının uygulama sonunda ne düzeyde gerçekleştiğinin saptanması üzerine inşa etmiştir (Tyler, 1981). *Sekiz Yıl Çalışması*'nda denediği ve geliştirdiği bu yaklaşımının ayrıntılarını 1942 yılında *Değerlendirme Üzerine Genel İfadeler* adı altında yayınlamıştır. Bu eserinde Tyler, amaca dayalı değerlendirmenin Şekil 4'de biçimsel olarak gösterilen uygulama süreçlerini ayrıntılı olarak ele almıştır (Fitzpatrick, Sanders ve Worthen, 2004; Mathison, 2005).



Şekil 4. Amaca Dayalı Program Değerlendirme Modelinin Aşamaları

Tyler böylece geliştirdiği amaca dayalı program değerlendirme modelinde, beklenen amaçlar ile gerçekleşen öğrenci performansı arasındaki farkın belirlenmesiyle, programdaki eksikliklerin giderilmesini ve bilgilerin program geliştirme sürecinde kullanılmasını amaçlamıştır. Program geliştirme modelini temel alarak geliştirdiği modelinde, programın amaç, öğrenme yaşantıları ve değerlendirme olmak üzere üç temel ögesinden söz etmekte ve amaçların, öğrencilere kazandırılmak istenen davranışları; öğrenme yaşantısının, öğrencilerin istedik davranışları kazanmaları için geçirmeleri gereken yaşantı ve etkinlikleri; değerlendirmenin ise amaçlara ulaşma derecesini belirlemek üzere yapılan etkinlikleri kapsadığını belirtmektedir (Erden, 1998).

Tyler'in "öğrenci performansı" olarak tanımladığı öğretme-öğrenme sürecinin sonunda oluşan ürünün yeterliliğini sorgulayan modeli, önceden belirlenmiş amaçlar doğrultusunda öğrenci davranışlarında oluşan değişimleri ve gelişmeleri temel alır. Dolayısıyla modelin başarıyla uygulanmasında kullanılacak ölçme araçlarının geçerli ve

güvenilir olması son derece önemlidir. Bu yönüyle öğreticinin süreci iyi bilmesi, yürütmesi ve değerlendirme konusundaki sorumluluklarını yerine getirmesi önemlidir. Amaçlara geçerliliği ve güvenilirliği sınanmış ölçme araçlarıyla ulaşıp ulaşılmadığı, eğer ulaşılmamışsa nedeni araştırılır. Sorun belirlenmiş amaçlardan kaynaklı ise amaçlar gözden geçirilir, gerekli görüldüğü durumda bazı amaçlar programdan çıkarılır veya yeniden tanımlanır. Eğer sorun amaçlarla ilgili değilse o zaman öğreticinin yol göstericiliğinde öğrenciler tarafından gerçekleştirilen öğrenme yaşantıları üzerinde durulur. Sorun öğrenme yaşantıları kaynaklı ise gerekli değişiklik ve düzenlemeler yapılır (Erden, 1998; Gözütok, 2001).

Program uygulaması sürecinde öğrencilerin programa girmeden önce, programın amaçladığı kimi davranışlara sahip olma olasılığı daima vardır. Bu nedenle hazırlanan ölçme araçlarının programın başında ve sonunda olmak üzere en az iki kez uygulanması önerilmektedir. Böylelikle kazanılan davranışın program yoluyla kazandırılıp kazandırılmadığı belirlenmiş olacaktır. Ayrıca davranışların kalıcılığını kontrol etmek için, programın bitiminden belli bir süre sonra da davranışların izleme çalışması ile değerlendirilmesi önemlidir (Erden, 1998).

Daha çok deneysel araştırma desenlerinin kullanıldığı bu modelde, bilişsel davranışların ölçülmesinde başarı testleri, devimsel davranışların ölçülmesinde gözlem formları ve performans testleri ve son olarak duyuşsal davranışların ölçülmesinde ise tutum ölçeklerinden yararlanılabileceği belirtilmektedir (Erden, 1998; Gözütok, 2001).

Metfessel-Michael Değerlendirme Modeli

Metfessel ve Michael 1967 yılında Tyler'in geliştirdiği program değerlendirme modelinden etkilenerek yeni bir program değerlendirme modeli geliştirmişlerdir. Mathison (2005) bu yeni değerlendirme modelini Tyler'in modelinden ayıran temel özelliğin, modelin alternatif ölçme araçlarının seçimine daha fazla ağırlık vermesi ve performansın amaçlarla karşılaştırılması yerine belirlenen standartlarla karşılaştırılması olduğunu belirtmektedir. Fitzpatrick, Sanders ve Worthen (2004) de Metfessel ve Michael'in bu değerlendirme anlayışı ile değerlendiricilerin o dönemde ölçme araçlarına ilişkin bakış açılarını önemli ölçüde etkilediğini ifade etmişlerdir.

Fitzpatrick, Sanders ve Worthen (2004) Metfessel ve Michael tarafından geliştirilen bu değerlendirme modelinde aşağıdaki aşamalardan söz etmektedir (Şekil 5):

- Program paydaşlarının (programdan etkilenen bireyler) süreçte yer almasının sağlanması
- Genel ve özel amaçlar arasında uyumun sağlanması
- Özel amaçların öğrencilere kazandırılabilir davranışlara dönüştürülmesi
- Ölçme araçlarının seçilmesi veya geliştirilmesi
- Programın uygulanması sırasında içerik testleri, ölçekler ve diğer ölçme araçlarını kullanarak sürekli gözlemlerin yapılması
- Uygun yöntemler kullanarak verilerin çözümlenmesi
- Verilerin standartlar kullanılarak yorumlanması
- Bulgular doğrultusunda amaçların gözden geçirilmesine, geliştirilmesine ve gelecekteki uygulamalara dönük önerilerin geliştirilmesi



Şekil 5. Metfessel-Michael Değerlendirme Modeli

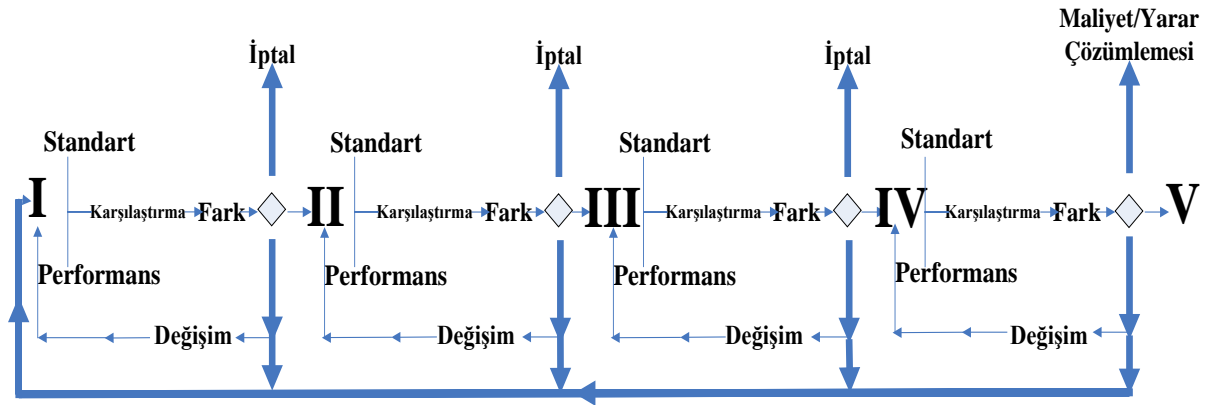
Şekil 4 ve Şekil 5 birlikte incelendiğinde, Tyler'in program değerlendirme modeli ile Metfessel-Michael değerlendirme modelinin uygulama aşamalarının birbirine benzerlik gösterdiği görülmektedir. Ancak Mathison'un (2005) da belirttiği gibi, iki model arasındaki belirgin farklar ölçme araçlarının çeşitliliği ve değerlendirmede standartların kullanılmasıdır.

Provus'un Çelişki / Fark Değerlendirme Modeli

Tyler geleneğinden gelen diğer bir değerlendirme modeli de Malcolm M. Provus'un 1960'lı yılların sonunda yaptığı değerlendirme çalışmaları sonucu geliştirdiği modeldir. Provus bu çalışmalarında Pittsburgh devlet okullarındaki okuma programlarının etkililiğini değerlendirmiştir. Her ne kadar çalışması yönetim tabanlı bir değerlendirme modeli olsa da, Provus'un çalışmasının temel dayanakları Tyler'in modeline dayanmaktadır. Provus'un (1969) geliştirdiği değerlendirme modelinin üç temel özelliği vardır. Bu özellikleri Provus şöyle belirtmektedir:

- Değerlendirme standartlara dayalı olarak yürütülmelidir. Değerlendirmede temel alınacak standartlar paydaşların ortak kararıyla kesinleştirilmelidir.
- Program değerlendirmede aşamalar sonunda elde edilen performans ile önceden belirlenen standartlar arasında bir çelişki veya fark olup olmadığı belirlenmelidir.
- Değerlendirmenin amacı programın durumuna dönük veriler sağlamak olduğundan elde edilen verilerin değerlendirilmesi sonucunda program devam ettirilmeli, geliştirilmeli veya sonlandırılmalıdır.

Provus'un "*Çelişki/Fark Değerlendirme Modeli*" adını verdiği bu model, diğer değerlendirme modellerinden farklı olarak programın geliştirilmesi sürecinde daha çok program tasarımının değerlendirilmesini amaçlamaktadır. Değerlendirmenin beklenen ile gerçekleşen arasındaki çelişkilerin, uyumsuzlukların ve farkların betimlenmesi olarak tanımlandığı bu modelde temel işlem, performans ve standartlar arasında yapılan karşılaştırmadır (Şekil 6).



Şekil 6. Çelişki/Fark Değerlendirme Süreci

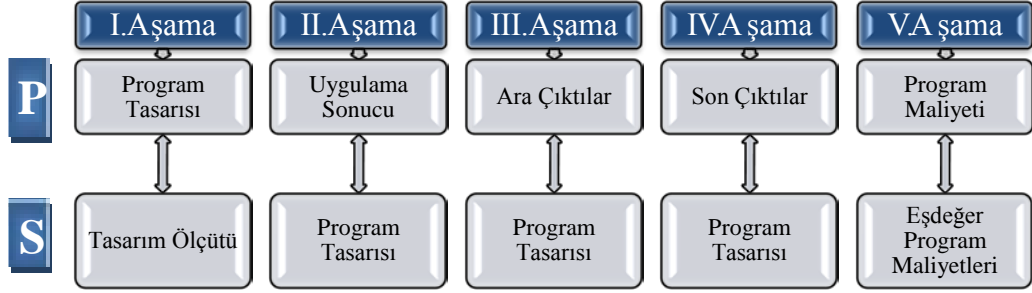
Kaynak: Provus, 1969

Şekil 6'da da görüldüğü gibi, modelde değerlendirme süreci ardışık bir yapı içinde yürütülmektedir. Bu sürecin her aşamasında *performans* ile *standartlar* karşılaştırılmakta ve bu ikili arasındaki *karşılaştırmadan* elde edilen *fark/çelişkiler* doğrultusunda aşağıdaki kararlar verilmektedir (Provus, 1969):

- Performanslar ile standartların karşılaştırılması sonucu eğer bir farklılık yoksa bir sonraki aşamaya geçilir. (I→II→III→IV→V)
- Performanslar ile standartların karşılaştırılması sonucu eğer bir farklılık varsa şu üç karardan biri alınır:
 - ▀ Standartlar ya da performanslar gözden geçirilir. Standartları karşılamada performansın yeterli olmadığı görülüyorsa programın performansı gözden geçirilmeli; standartların uygun ve gerçekçi belirlenmediği düşünülüyorsa da standartlar değiştirilmelidir.
 - ▀ Programın uygulanmasına yeniden başlanabilir ve performans ya da standartlar yeniden düzenlenebilir.
 - ▀ Farklılıklar düzeltilemeyecek kadar fazla ise program sonlandırılır.

Provus'un değerlendirme modelinin temel amacı programın süreç içerisinde değerlendirilerek geliştirilmesidir. Bu nedenle Provus (1969), programa ilişkin yukarıda

belirtilen kararların alınabilmesi için her aşamada birbiriyle karşılaştırılabilecek performans (P) ve standartların (S) neler olduğunu Şekil 7’de göstermiştir.



Şekil 7. Provus'un Program Değerlendirme Modelinin Aşamaları

Kaynak: Provus (1969)'dan özetlenmiştir.

Şekil 7’de görüldüğü gibi, her bir aşamada süreç sonunda bir performans ve sürece ilişkin sağlam bir kararın verilebilmesi için de standartlara gereksinim duyulmaktadır. Örneğin, değerlendirmenin ilk aşamasında program tasarısı (tüm boyutları ile birlikte) bir performans iken, tasarım ölçütleri de standardı temsil etmektedir. Program tasarısı ile tasarım ölçütleri arasındaki fark/çelişki doğrudan değerlendirme yöneticilerine iletilir ve aşamayla ilgili bir karara varılır. Değerlendirme sürecinin bu beş aşamasına ilişkin bilgiler aşağıda yer almaktadır.

- **I. Aşama:** Tasarlama ya da tanımlama olarak da adlandırılan bu aşamada performans, program tasarısıdır. Diğer bir ifadeyle değerlendirilecek olan, hakkında karar verilecek olan program tasarısının kendisidir (Provus, 1969). Program tasarısı, program paydaşlarının ve uygulayıcıların görüşleri doğrultusunda program tasarım ölçütleri dikkate alınarak ortaya çıkarılır. Bu ölçütler, genel olarak amaçlar, bu amaçları davranışa dönüştürecek etkinlikler, ortaya çıkması beklenen davranışlar, amaçları yerine getirmede kullanılacak kaynakları vb. kapsar. Program tasarısı ortaya çıktıktan sonra tasarı, program yöneticileri ve değerlendiricinin katıldığı bir panelde kapsamı, anlaşırılığı, iç tutarlılığı vb. yönlerden değerlendirilir. Tasarım ölçütlerini oluşturan bu değişkenler bu aşamanın standartlarını oluşturmaktadır. Kısaca, bu aşamada ortaya çıkan performans olan program

tasarısının kapsamının tutarlılık, anlaşılabilirlik vb. standartlarla olan uygunluğu değerlendirilmektedir. (Fitzpatrick, Sanders ve Worthen, 2004).

- *II. Aşama:* Deneme ya da kurulum olarak adlandırılan bu aşamada amaç, programın tasarlandığı biçimiyle uygulanmaya hazır hale gelip gelmediğini kesinleştirmektir. Bunun belirlenmesi için de programın denenmesi ve bu sürecin değerlendirilmesi gereklidir. Burada değerlendirmede dikkate alınacak standart program tasarısıdır. Çünkü deneme, kesinleşmiş ve bir önceki aşamada standartlara uygunluğu belirlenmiş tasarıya göre yürütülmelidir. Dolayısıyla deneme süreci performansı; program tasarısı da standardı temsil etmektedir. Değerlendirici deneme (performans) ile program tasarısı (standart) arasında karşılaştırma yaparken herhangi bir çelişkinin/farkın ve uyumsuzluğun olup olmadığını belirlemek için tasarımı madde madde inceleyerek bir dizi uygunluk-örtüşme testi yapar (Provus, 1969). Bu testler gözlemler yoluyla olabileceği gibi doğrudan ölçmelerle yoluyla da gerçekleştirilebilir. Test sonuçları paydaşlar ve uygulayıcıların bir araya geldiği bir oturumda program tasarısı da dikkate alınarak tartışılır. Yapılan inceleme sonucunda bir çelişkiye rastlanması durumunda program yöneticisinin alacağı iki karar vardır: Program tasarısı değiştirilir ya da deneme çalışması yeniden düzenlenir.
- *III. Aşama:* Süreç aşaması olarak adlandırılan ve ilk kez neden-sonuç ilişkisinin araştırıldığı bu aşamada programla değiştirilmek istenen davranışlarla, bu değişimi etkileyen süreç arasındaki ilişki incelenmektedir. Diğer bir ifadeyle bu aşamada performans, girdiyi çıktıya dönüştürebilmiş midir, süreçteki eylemlerin çıktı üzerindeki etkileri nelerdir vb. soruların yanıtları aranmaktadır. Biçimlendirici değerlendirme özellikleri taşıyan bu aşamada değerlendirme sonucu ortaya çıkan ara çıktılar performansı, program tasarısı ise standart olarak nitelendirilmektedir (Provus, 1969).
- *IV. Aşama:* Ürün aşaması olarak adlandırılan bu aşamada sürecin sonunda ulaşılan tüm çıktılarının değerlendirilmesi amaçlanmaktadır. Bu durum programın etkililiğiyle de yakında ilgilidir. Provus (1969) değerlendiricinin süreç sonunda programın belirlenen amaçları gerçekleştirmede etkili olup olmadığının

sorgulamasının programın etkililiğine ilişkin bilgi sağlayacağını, bunu gerçekleştirmesi için de özellikle deneysel araştırma desenleri kullanması gerektiğini ifade etmektedir. Dolayısıyla, toplam değerlendirme özellikleri taşıyan bu aşamada elde edilen son çıktılar performansı; program tasarısı (özellikle çıktı boyutu) ise standardı oluşturmaktadır.

- *V. Aşama:* Programın geliştirilmesine ilişkin tüm aşamaların değerlendirilmesinden sonra programın maliyet-etkililik analizinin yapılması ve bu sonuçların eşdeğer program değerlendirme çalışmaları ile karşılaştırılması önemlidir. Programın etkililiğinden çok, ekonomik olup olmadığının incelendiği bu aşamada maliyet-etkililik analizinin sonuçları programın performansını; farklı programlardan elde edilen veriler ise performansın karşılaştırılacağı standardı oluşturmaktadır. Provus (1969) bu karşılaştırmanın en temel amaçlarından birinin kaynakların etkili paylaşımının yapılması olduğunu vurgulamaktadır.

Sonuç olarak programın değerlendirilmesinin program geliştirilirken yapılmasını temel alan ve bu nedenle daha çok program tasarılarının ölçüt olarak dikkate alındığı bu model, girdilerin beklenen çıktı olarak gerçekleşip gerçekleşmediğini araştırdığı için amaca dayalı; programa ilişkin kararlar yöneticiler tarafından alındığı için de yönetim odaklı program değerlendirme modeli olarak görülebilmektedir.

Amaçtan Bağımsız Program Değerlendirme Modeli

Scriven (1972) tarafından geliştirilen bu program değerlendirme modelinde vurgu programın amaçları üzerinde değildir. Genel olarak amaçtan bağımsız değerlendirmede bir alan çalışması yürütülerek çıktılar ile ilgili bilgi toplama ve gözlenen çıktıların program katılımcılarının gerçek beklentileri ile karşılaştırma amaçlanmaktadır (Patton, 2002). Değerlendirici, program amaçları ile ilgili hiçbir bilgi ile ilgilenmez, program yöneticileri ile amaçlar üzerine tartışmaz, öneri geliştirmez yalnızca gözlenen çıktılar ile katılımcıların beklentilerini karşılaştırır. Patton (2002) amaçtan bağımsız bir değerlendirme yürütmeyi özellikle dar kapsamlı amaçlarla çalışmayı sevmeyen, program sonucunda beklenmedik çıktıları da dikkate alan, amaçlara, yönetime bağlı kalmadan özgürce değerlendirme yapmak isteyen değerlendiricilerin tercih ettiğini bildirmektedir. Scriven'in amaçları dışarıda bırakan bu modeli çeşitli değerlendiricilerden bu

nedenlerden dolayı tepki almıştır. Örneğin, Rossi (1982), açık bir biçimde amaçları tanımlanmamış bir sosyal yardımlaşma programının değerlendirilemeyeceğini ifade ederken, Weiss (1972), değerlendiricinin değerlendirmede neyi arayacağını bilmesi için amaçların açık olması gerektiğini vurgulamıştır.

Amaç odaklı program değerlendirme yaklaşımına alternatif olarak geliştirilen amaçtan bağımsız değerlendirme şu temel özellikleri taşımaktadır (Fitzpatrick, Sanders ve Worthen, 2004):

- Değerlendirici bilinçli olarak programın amaçlarından haberdar olmaktan kaçınır.
- Önceden belirlenen amaçların değerlendirmenin odak noktasını daraltmasına izin verilmez.
- Amaçtan bağımsız değerlendirme program sonucunda beklenen çıktılardan çok, gerçek çıktılar üzerinde yoğunlaşır.
- Değerlendirici program yöneticisi ve personeli ile iletişimini en düşük düzeyde tutar.
- Beklenmeyen çıktılarının belirtilmesi eğilimini artırır. Amaçtan bağımsız değerlendirme beklenmedik amaçların etkililiği olasılığını artırır.

Amaca Dayalı Program Değerlendirme Yaklaşımının Üstünlükleri ve Sınırlıkları

Stufflebeam, Madaus ve Kellaghan (2002) amaca dayalı program değerlendirme yaklaşımının programların amaçları açık ve anlaşılır olduğu durumlarda kullanılmasının daha uygun olduğunu ifade etmektedirler. Anlaşılması ve uygulanması oldukça kolay olan bu yaklaşımın güçlü olduğu diğer yönü modellerin uygulama aşamaları belirgin olduğundan karar vericiler için tüm aşamalarda programla ilgili olarak bol miktarda bilgi sağlanmaktadır. Ayrıca programın amaçlarının açıkça ifade edilmesi, değerlendiricilere program hakkında yargıya varmada hangi temel değerlerin dikkate alınacağını görmelerinde yardımcı olur. Öte yandan bu yaklaşımı eleştirenler bu yaklaşım ile programın değeri ve kazancı konusunda yeterince bilgi sağlanamayacağını; yalnızca amaçların ölçülmesinin ve değerlendirilmesinin yapılabileceğini belirtmektedirler. Yine amaçlarda belirtilmemiş olan önemli program ürünlerinin değerlendirilmesinin göz ardı

edilmesi yaklaşımın önemli eksikliklerindedir (Fitzpatrick, Sanders ve Worthen, 2004; Stufflebeam, Madaus ve Kellaghan, 2002) .

1.3.5.2. Yönetime Dayalı Program Değerlendirme Yaklaşımı ve Modelleri

Yönetim dayalı program değerlendirme yaklaşımı, program değerlendirme sürecinden elde edilen bilgilerin özellikle karar vericiler, kurullar, müdürler vb. yönetim basamağında yer alan bireylerce daha etkili kullanılabileceğini savunan bir yaklaşımdır. 1960'lı yılların sonlarına doğru eleştirilmeye başlanan program değerlendirme modellerine alternatif olarak ortaya atılan bu yaklaşımla birlikte değerlendirme sürecinde program amaçları yerine program yöneticilerinin kararlarına odaklanılmıştır (Fitzpatrick, Sanders ve Worthen, 2004). Bu nedenle, bu değerlendirme yaklaşımına dayalı geliştirilen modellerde de değerlendirme amacı ve soruları yöneticilerin bilgi gereksinimleri doğrultusunda oluşturulur.

Yönetime dayalı program değerlendirme yaklaşımında bir programın değerlendirilmesi programın niteliğiyle ilgili olarak program hakkında karar verme yetkisine sahip bireylerin ya da kuruluşların bilgilendirilmesi amacıyla yapılır. Bu yaklaşıma dayalı program değerlendirme sürecinde üç anahtar kavramdan söz edilmektedir: Karar vericilerin ilgileri, değerlendirmede kullanılacak bilgiler ve değerlendirme sürecinin etkililiğinin sağlanması için belirlenen standartlar. Diğer bir ifadeyle; programı değerlendiren kişi karar vericiler ile birlikte çalışır, verilmesi gerekli kararların neler olduğunu betimler ve belirlenmiş standartlara dayalı olarak her bir kararın olumlu ve olumsuz getirileri hakkında veri toplar.

Yönetime dayalı program değerlendirme yaklaşımının dayandığı temel ilkeleri şunlardır (Fitzpatrick, Sanders ve Worthen, 2004):

- Program değerlendirme planlanmasında ve yürütülmesinde programın amaçlarından çok program yöneticilerinin kararları dikkate alınır.
- Program değerlendirmede, programın girdi, süreç ve çıktı boyutlarının her birine ilişkin kararların alındığı bir sistem yaklaşımı benimsenir.

- Programın değerlendirilmesinde farklı karar vericiler ve buna bağlı olarak alınması gereken farklı kararlar vardır. Bu nedenle değerlendirme sonuçlarını kimlerin kullanacağı, ne amaçla ve nasıl kullanılacağı açıklığa kavuşturulur.
- Yönetime dayalı bir program değerlendirmenin başarısı, değerlendiriciler ile karar vericiler arasındaki iletişim, etkileşim ve takım çalışmasının niteliğiyle doğrudan ilişkilidir.

Yönetime dayalı program değerlendirme yaklaşımının gelişimine katkıda bulunan birçok uzman bulunmakla birlikte bu yaklaşımın öncüsü Stufflebeam olarak kabul edilir. Stufflebeam (2003) amaca dayalı program değerlendirme yaklaşımlarına alternatif olarak geliştirdiği ve bir eğitim programının değerlendirilmesinde, programın amaçlarından çok program yöneticilerinin kararlarını merkeze koyduğu program değerlendirme modelini Çevre-Girdi-Süreç-Ürün (ÇGSÜ) (Context-Input-Process-Product,CIPP) modeli olarak adlandırmıştır. Stufflebeam, bu modelin geliştirilmesinde Henry Bernard, Horace Mann, William Torey Harris ve Carleton Washburne'ün yönetim kuramına ilişkin yaptıkları çalışmalardan etkilenmiştir.

ÇGSÜ modelinin dışında yönetime dayalı program değerlendirme yaklaşımını temel alan modellerden biri de Alkin'in bu modelden esinlenerek UCLA Üniversitesi'nde geliştirdiği UCLA program değerlendirme modelidir. UCLA modelinde, ÇGSÜ modelinden farklı olarak "programın uygulanması" aşaması yer almaktadır (Alkin, 2004). Bu iki modelin yanı sıra Wholey tarafından 1983 ve Patton tarafından 1986 yılında iki model daha geliştirilmiştir.

Çevre-Girdi-Süreç-Ürün Modeli (ÇGSÜ)

Bu değerlendirme modeli, programların, projelerin, personelin, ürünlerin, kurumların ve sistemlerin biçimlendirici ve toplam değerlendirmelerinde kullanılan kapsamlı bir değerlendirme modelidir. Model, Daniel Stufflebeam tarafından ilk olarak 1966 yılında hükümet tarafından kırsal bölgelerdeki okullarda uygulanan programların değerlendirilmesi ve denetlenmesi için geliştirilmiştir. Zamanla Amerika içinde ve dışında kısa ve uzun süreli araştırmalarda sıkça kullanılan model, askeri, işletme ve sağlık alanlarına da uyarlanmıştır.

Stufflebeam'a (2003) göre, program değerlendirme programının amaçları, tasarımı, uygulaması ve çıktılarının değeri ve kazancı hakkında betimsel ve yargısal bilgilerin tanımlanması, toplanması ve uygulanması sürecidir. Değerlendirmenin en önemli amacı ise programın niteliğinin kanıtlanmasından çok, programın sürekli olarak geliştirilmesidir. ÇGSÜ modeli bu yönüyle Provus'un amaca dayalı program değerlendirme modeliyle benzerlikler taşımaktadır. Her iki modelde de programın süreçte geliştirilmesi, program yöneticilerine dönük olarak alınan bazı kararlar ve bu kararların alınmasında başvuru standartları yer almaktadır.

Stufflebeam (2003) ÇGSÜ modelinde yöneticilerin programa ilişkin sağlıklı değerlendirmeler yapabilmelerine yardımcı olacak dört farklı karardan söz etmekte ve bu kararların alınabilmesi için de dört değerlendirme türünün kullanılması gerektiğini ifade etmektedir. "Çevre", "Girdi", "Süreç" ve "Ürün" değerlendirmesi olarak adlandırılan ve değerlendirmede sistem yaklaşımını da benimseyen bu değerlendirme türlerinde sırasıyla "Planlama", "Yapılandırma", "Uygulama" ve "Gözden Geçirme" kararları alınmaktadır. Bu değerlendirme türlerinin ayrıntılı açıklamaları aşağıda verilmiştir.

- *Çevre Değerlendirmesi*: Karar vericilerin programın amaçlarını ve önceliklerini belirlemelerine; kullanıcıların da bu amaçlar, öncelikler ve çıktılar hakkında bir yargıda bulunmalarına yardımcı olabilecek gereksinimlerin, sorunların ve fırsatların değerlendirilmesini temel alan değerlendirme türüdür. Çevrenin değerlendirilmesinin temel amacı, programın amaçlarının ne olması gerektiğine karar vermektir (Fitzpatrick, Sanders ve Worthen, 2004). Stufflebeam, Madaus, ve Kellaghan (2002) ve Kellaghan, Stufflebeam ve Wingate (2003), çevrenin değerlendirilmesinin; gereksinimlerin amaçlar ve kaynaklarla uyumluluğu ve tutarlılığı ile amaçların sorunların çözümünde yeterli olup olmadığını belirlemede önemli işlevi olduğunu ifade etmektedir. Demirel (2006) ise çevrenin değerlendirilmesinin, gereksinim çözümlemesi sürecinde karşılanamayan gereksinimlerin, kaçırılmış olanakların niçin karşılanamadığının ve kaçırıldığının belirlenmesi nedeniyle önemli olduğunu vurgulamaktadır.
- *Girdi Değerlendirmesi*: Stufflebeam'a (2003) göre, en çok ihmal edilen ancak en önemli değerlendirme türü olan girdi değerlendirmesi, programın amaçlarının

gerçekleştirilmesi ve gereksinimlerin karşılanabilmesi için gerekli etkinlik planlarının, personel planlarının ve uygulamada kullanılacak yaklaşımların maliyet-etkililik ve uygulanabilirlik yönünden değerlendirilmesini kapsar (Stufflebeam, Madaus ve Kellaghan, 2002). Girdilerin değerlendirilmesi ile karar vericiler, planların işlevselliği, çözüm stratejilerinin belirlenmesi, destek kaynakların seçimi ve etkinliklerin nasıl yapılandırılması gerektiği konusunda kararlar alırlar. Diğer bir ifadeyle değerlendirmenin yapılandırılmasına ilişkin kararlar bu değerlendirme sonucu alınır.

- *Süreç Değerlendirmesi:* Programın nasıl uygulandığının, sürecin nasıl ilerlediğinin, programın yasal ve etik ilkeler doğrultusunda yürütülüp yürütülmediğinin, süreçte karşılaşılan sorunların neler olduğunun kısaca uygulama süreçlerinin değerlendirilmesidir. Fitzpatrick, Sanders ve Worthen (2004), sürecin değerlendirilmesinde değerlendiricinin aşağıdaki sorulara yanıt aradığını belirtmektedirler:

- Uygulanan etkinlik planı ne kadar iyidir?
- Planın başarısını etkileyen etmenler nelerdir?
- Planın gözden geçirilmesi gerekli noktaları var mıdır?

Değerlendirmenin uygulanmasına ilişkin kararların alındığı süreç değerlendirme sonunda program tasarımının düzeltilmesi ve geliştirilmesi ile ilgili yargılara da varılır. Buradaki yargılar ürünlerin değerlendirilmesi sürecine de temel oluşturmaktadır.

- *Ürün Değerlendirme:* Bu değerlendirmede programın genel ve özel çıktılarının incelenmesi, beklenen çıktıların ölçülmesi, beklenmeyen çıktıların tahmin edilmesi, programın etkisinin değerlendirilmesi, etkili bir maliyet yararlık çözümlemesinin yürütülmesi amaçlanmıştır. Kısaca, kısa ve uzun dönemli, beklendik ve beklenmedik çıktıların neler olduğu sorgulanmakta ve programın yeniden düzenlenmesine ilişkin alınacak kararlar belirlenmektedir (Stufflebeam, 2003; Kellaghan, Stufflebeam ve Wingate, 2003). Fitzpatrick, Sanders, ve Worthen (2004), programın kazanımları ile ilgili bir değerlendirmenin yapılabilmesi için değerlendiricinin şu sorulara yanıt vermesi gerektiğini belirtmektedirler:

- Değerlendirmede hangi sonuçlara ulaşıldı?
- Gereksinimler karşılandı mı?
- Değerlendirmeden sonra program ile ilgili hangi kararlar alınmalıdır.

Stufflebeam ve Shinkfield (2007), ÇGSÜ modelindeki değerlendirme türlerinin amacı, yöntemi ve değerlendirme sürecinde verilmesi gerekli kararların özeti Tablo 3'te verilmiştir.

Tablo 3.

ÇGSÜ Değerlendirme Modelinin Dört Değerlendirme Türü

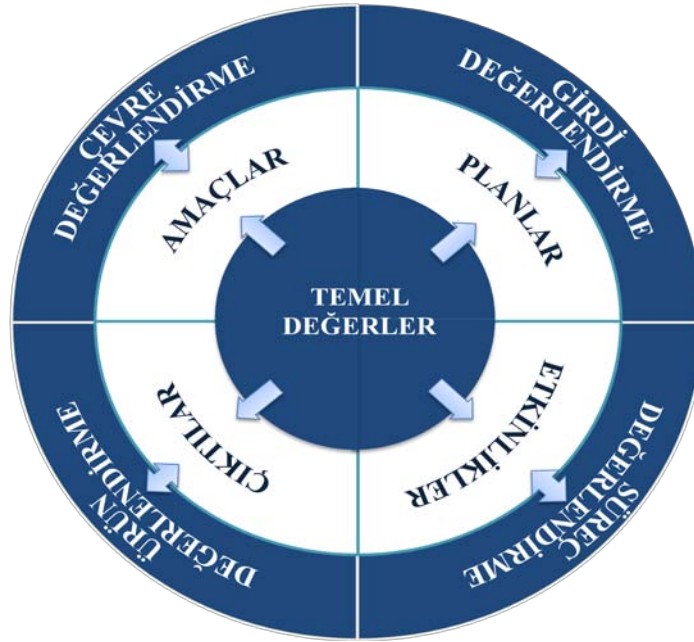
	Amaç	Yöntem	Verilecek Karar
Çevre Değerlendirme	Kurumsal çevreyi tanımlamak, hedefi grubu ve gereksinimlerini belirlemek, sınırlılıkları teşhis etmek, öncelikleri ve kaynakları tanımlamak, amaçların gereksinimleri karşılayıp karşılamayacağına karar vermek	Sistem analizi, tanılayıcı testler, kontrol listeleri, doküman analizi, görüşmeler, delfi yöntemi, uzman panelleri, odak görüşmeler vb.	Uygulama çevresinin, amaçların, önceliklerin vb. karşılaştırılması
Girdi Değerlendirme	Uygulama için seçilen araştırma desenlerini, belirlenen bütçeyi, sistemin olanaklarını belirlemek ve değerlendirmek	Alanyazın taraması, özlem ve görüşmeler, ziyaretler, uzman görüşleri, paneller, pilot denemeler vb.	Yapılandırma etkinlikleri için yöntem, bütçeye, zamanlamaya ilişkin gerekli seçim kararları
Süreç Değerlendirme	Uygulamada karşılaşılan sorunları belirlemek, gerçekleşen etkinlikleri kaydetmek ve değerlendirmek	Katılımcı gözlemleri, bağımsız gözlemler, görüşmeler, doküman analizi vb.	Program planının uygulanmasına ilişkin kararlar
Ürün Değerlendirme	Çıktıları tanımlamak ve çıktılarının amaçlarla uygunluğunu değerlendirmek, onların değerini ve etkisini yorumlamak	Paydaş değerlendirmeleri, nitel ve nicel analizler	Programın sürdürülmesi, sonlandırılması ya da yenilenmesine dönük kararlar

Kaynak: Stufflebeam, Madaus ve Kellaghan (2002)'den özetlenmiştir.

Tablo 3'teki özet bilgiler incelendiğinde, çevrenin değerlendirilmesinin daha çok değerlendirme süreci başlamadan önce süreçte dikkate alınması gerekli değişkenlerin değerlendirilmesi ile ilgili olduğu; girdi değerlendirmesinde, bu değişkenlere bağlı olarak çalışma planının, uygulamada işe koşulacak yöntemlerin, tekniklerin ve değerlendirme bütçesinin seçimine dönük kararların alındığı, süreç değerlendirmesinin tümüyle

uygulama etkinliklerinin değerlendirilmesini içerdiği ve ürün değerlendirmesinde de programın geleceğine dönük kararların alındığı görülmektedir.

Stufflebeam değerlendirme türleri arasındaki ilişkiyi de farklı bir şekilde aşağıdaki gibi görselleştirmiştir (Şekil 8).



Şekil 8. ÇGSÜ Değerlendirme Modelinin Temel Kavramları

Kaynak: Stufflebeam, Madaus ve Kellaghan (2002)'den alınmıştır.

Şekil 8’de gösterildiği gibi, en içteki daire programla ilgili temel değerleri ifade etmektedir. Çünkü program değerlendirmede kök kavram “değer”dir. “Değer” bir toplum, grup ya da bireyin sahip olduğu görüşlerdir. Örneğin, insan hakları, fırsat eşitliği, kaynakların etkili kullanımı, ürünlerin ve sürecin güvenliği vb. birer değeri ifade etmektedir. Temelde değerlendirci bir programı ya da kurumun hizmetlerini ilgili sosyal, kurumsal, mesleki ve teknik değerlere göre değerlendirir. Bu anlamda değerler değerlendirmede kullanılacak ölçütlerin de temellerini oluşturur. Örneğin, programla ilgili ölçütler öğrencinin zihinsel, psikolojik, estetik, sosyal, fiziksel, ahlaki ve mesleki gelişiminin göstergeleri olabilir. Bu ölçütler program gereksinimlerinin belirginleştirilmesinde ve daha sonra ölçme araçlarının oluşturulmasında oldukça yararlı olabilir. Sonuçta değerlendirci program değerlendirmenin temeline değerleri almak

zorundadır. Stufflebeam (2003) modelinin bu yönüyle “değer tabanlı program değerlendirme modeli” olarak da tanımlanabileceğini ifade etmektedir.

Şekil 8’de “temel değerler” dairesini çevreleyen ikinci daire bir programı etkileyen ve değerlendirilmesi gereken dört ögesini temsil etmektedir. Bunlar; amaçlar, planlar, etkinlikler ve çıktılardır. En dıştaki daire ise bu dört ögeyle ilişkili dört değerlendirme türünü ifade etmektedir: Çevre, girdi, süreç ve ürün değerlendirmesi. Çift yönlü oklar programın ögesi ile değerlendirme türü arasındaki karşılıklı ilişkiyi temsil etmektedir. Örneğin, amaçlar çevre değerlendirmede sorulabilecek soruların temelini oluşturur. Bu işlemin karşılığı olarak yapılan çevre değerlendirmesi sonucunda ise bu amaçların geçerliği kanıtlanır ya da amaçlar yeniden geliştirilir. Planlama etkinlikleri girdi değerlendirmesi için gerekli verileri sağlarken, girdi değerlendirmesi de planların değerlendirilmesine ve gerekiyorsa güçlendirilmesine ilişkin geri dönütler sağlar. Program etkinlikleri süreç değerlendirmesinde dikkate alınacak temel girdilerdir. Diğer taraftan süreç değerlendirmesi de bu etkinliklerin niteliğine dönük bilgiler sağlar. Son olarak programdaki başarılar, başarısızlıklar ve yan etkiler ürün değerlendirmesinin temelini oluşturur. Ürün değerlendirmesi ise bu ürünler hakkında gerekli kararları ve nasıl daha iyi çıktılar alınabileceğine ilişkin geri bildirimler sağlar.

Stufflebeam çevre, girdi, süreç ve ürün değerlendirmelerinin tasarlanmasına ilişkin değerlendiricilere yol gösterecek bir yönerge hazırlamıştır. Bilimsel araştırma sürecinin aşamalarına benzeyen bu yönergenin temel aşamaları, başlangıç, bilginin toplanması ve düzenlenmesi, bilginin analizi, raporlaştırılması ve değerlendirme yönetimidir. Her bir aşamada yapılması gerekenler ise şu şekilde açıklanmıştır (Fitzpatrick, Sanders ve Worthen, 2004):

- *Başlangıç:* Değerlendirmenin başlangıcında değerlendiricinin alması gerekli kararları ifade etmektedir. Bu kararlar şunlardır:
 - ▀ Değerlendirmede verilecek kararların düzeyinin belirlenmesi (yerel, eyalet düzeyi ya da ulusal)
 - ▀ Belirlenen her düzey için kararların yer, özellikler, zamanı vb. yönlerden açıklanması
 - ▀ Her bir karar durumu için standartların belirlenmesi

- Değerlendiricinin hangi yönergeleri izleyeceğinin belirlenmesi
- *Bilginin Toplanması ve Düzenlenmesi:* Değerlendiricinin değerlendirme için gerekli bilgilerin toplanması ve düzenlenmesine dönük alması gerekli kararları ifade etmektedir. Bu kararlar şunlardır:
 - Bilgi kaynaklarının tanımlanması
 - Gerekli bilgileri toplamada kullanılacak yöntem ve araçların belirlenmesi
 - Toplanacak bilgi için bir ortak bir yapının oluşturulması
 - Evren/örneklemin belirlenmesi
 - Bilgi toplamaya dönük bir çalışma takviminin oluşturulması
- *Bilginin Analizi:* Değerlendiricinin toplanacak bilgilerin çözümlenmesine dönük alması gerekli kararları ifade etmektedir. Bu kararlar şunlardır:
 - Çözümleme tekniklerinin belirlenmesi
 - Çözümleme aracının belirlenmesi (paket programlar ya da birey vb.)
- *Raporlaştırma:* Değerlendiricinin değerlendirme sonucu ulaşılabilecek bulguların raporlaştırılmasına dönük alması gerekli kararları ifade etmektedir. Bu kararlar şunlardır:
 - Değerlendirme raporunun paydaşlarının/ilgililerinin tanımlanması
 - Paydaş/ilgililere bilginin nasıl sunulacağına ilişkin araçların belirlenmesi
 - Değerlendirme raporu için uygun bir yapının belirlenmesi (özet rapor, kapsamlı rapor vb.)
 - Raporlaştırma takviminin belirlenmesi
- *Değerlendirmenin Yönetilmesi:* Değerlendiricinin değerlendirme sürecinde değerlendirmenin yönetimine dönük alması gerekli kararları ifade etmektedir. Bu kararlar şunlardır:
 - Değerlendirme takviminin özetlenmesi
 - Kadro ve kaynak gereksinimlerinin belirlenmesi ve bunların karşılanması için plan yapılması
 - Geçerli, güvenilir, ölçülebilir bilgi sağlayacak potansiyel değerlendirme planının değerlendirilmesi
 - Değerlendirme planının düzenli olarak güncellenmesi için gerekli araçların

belirlenmesi ve uygulama izlencesinin çıkarılması

- Programın değerlendirilmesi için gerekli bütçenin sağlanması

ÇGSÜ değerlendirme modelinde temel amaç, karar verme ve programla ilgili geriye dönük bilgi sağlamadır. Bu iki amaç, modeldeki değerlendirme türlerinin programın geliştirilmesi, niteliğinin arttırılmasına dönük kararların alındığı biçimlendirici değerlendirme ve programın niteliğinin ortaya konulduğu ve programla ilgili geriye dönük bilgilerin sağlandığı toplam değerlendirmenin işlevlerinin olduğunu göstermektedir. Tablo 4’de ÇGSÜ modelindeki dört değerlendirme türünün biçimlendirici değerlendirme ve toplam değerlendirme işlevi verilmiştir.

Tablo 4.

Dört Değerlendirme Türünün Biçimlendirici ve Toplam Değerlendirme İşlevleri

	Çevre	Girdi	Süreç	Ürün
Biçimlendirici Değerlendirme (<i>karar vermede</i>)	Amaçların seçilmesinde ve sıralanmasında yardımcı olması	Araştırma deseninin değerlendirmesine ve kaynak dağıtım planlarının seçiminde yardımcı olması	Planı uygulamada yol göstermesi	Programı devam ettirme, sonlandırma ya da yenilemede yol göstermesi
Toplam Değerlendirme (<i>programın değeri, etkisi, önemine ilişkin bilgi sağlamada</i>)	Amaçların ve önceliklerin gereksinimler, sorunlar ve fırsatlar ile karşılaştırılması	Program tasarımının ve bütçesinin alternatif tasarımlarla ve kullanıcı gereksinimleriyle karşılaştırılması	Tasarlanan ve gerçekleşen sürecin ve maliyetin karşılaştırılması	Çıktıların ve yan çıktıların(beklenmedik) amaçlar ve gereksinimlerle ve olanaklı ise farklı program çıktıları ile karşılaştırılması

Kaynak: Stufflebeam, Madaus ve Kellaghan (2002)’den özetlenmiştir.

Tablo 4’de de görüldüğü gibi, program değerlendirme çalışmasını yürüten araştırmacıların programların ve projelerin geliştirilmesinde veya iyileştirilmesinde (biçimlendirici rol) ya da denetim için gerekli bilginin sağlanmasında (toplam değerlendirme rolü) çevre, girdi, süreç ve ürün değerlendirmelerini kullanabileceklerini göstermektedir. Buna göre, değerlendiriciler öğretmenlerin, yöneticilerin veya diğer hizmet sağlayıcılarının bir program, proje veya hizmeti tasarlamaları ve yürütmelerine yardımcı olacak bir değerlendirme tasarlayabilir ve yürütebilir.

Alkin’in UCLA Program Değerlendirme Modeli

1969 yılında Mary Alkin tarafından UCLA Üniversitesinde (University of California, Los Angeles) geliştirilen ve geliştirildiği üniversitenin adını alan bu program değerlendirme

modeli Stufflebeam'in geliřtirdiđi GSÜ modeliyle bazı boyutlarda benzer özellikler göstermektedir. Örneđin, her iki model de program deđerlendirmede sistem yaklařımını temel aldıđı ve iki modelde yer alan deđerlendirme türlerinde yapılan etkinliklerin aynı olduđu görölmektedir. Modelin SGÜ modelinden farkı son iki deđerlendirmenin biçimlendirici ve toplam deđerlendirme özelliklerini taşımasıdır (Alkin, 2004). Bu modeldeki deđerlendirme türleri řöyledir (Fitzpatrick, Sanders ve Worthen, 2004):

- *Sistemlerin Deđerlendirilmesi*: Programın uygulandıđı sistemin mevcut durumu ile ilgili bilgilerin sađlanması amaçlandıđı bu deđerlendirme türü GSÜ modelinde çevrenin deđerlendirilmesi ile benzer özellikler taşır.
- *Programın Tasarlanması*: Eđitsel gereksinimleri karřılayacak uygun programların seilmesinin amaçlandıđı bu deđerlendirme türü de GSÜ modelinde girdilerin deđerlendirilmesi ile benzer özellikler taşır.
- *Programın Uygulanması*: Modelin GSÜ modelinden farklı tek deđerlendirme türüdür. Ancak burada gerekleřen eylemler GSÜ modelindeki süreç deđerlendirmesinde de yer almaktadır. Bu deđerlendirme türünde programın amaçlanan řekilde hedef kitleye sunulup sunulmadıđına iliřkin bilgilerin toplanması amaçlanmaktadır.
- *Programın Geliřimi*: GSÜ modelinde sürecin deđerlendirilmesine benzer olan bu deđerlendirme türünde programın uygulanması sırasında amaçların gerekleřip gerekleřtirilmediđinin, beklenmeyen ıktıların ortaya ıkıp ıkmadıđının deđerlendirilmesi amaçlanmaktadır. Diđer bir ifadeyle programla ilgili biçimlendirici deđerlendirmenin yapıldıđı deđerlendirmediir.
- *Programın Onaylanması*: Programla ilgili toplam deđerlendirmenin yapıldıđı bu deđerlendirme GSÜ modelindeki ürün deđerlendirmesi ile aynıdır. Bu deđerlendirme ile programın deđerleri ile ilgili bilgiler sađlanmaktadır.

Alkin (2004), geliřtirdiđi bu deđerlendirme modeliyle deđerlendirme alanına deđerlendirme ile ilgili dđrt varsayım kazandırdıđını belirtmektedir. Alkin'in belirttiđi bu varsayımlar řunlardır (Fitzpatrick, Sanders ve Worthen, 2004):

- Deđerlendirme ilgili verilerin toplanması sđrecidir.
- Deđerlendirmede toplanan veriler programın geleceđine iliřkin karar verme sđrecinde kullanılır.
- Deđerlendirme verileri karar vericilerin onları etkili bir řekilde kullanabilecekleri ve karıřıklıđa ve yanlış kullanıma neden olmayacak řekilde sunulmalıdır.
- Farklı kararlar farklı deđerlendirme uygulamalarını gerektirir.

Yönetime Dayalı Program Deđerlendirme Yaklařımının Üstünlükleri ve Sınırlılıkları

Her program deđerlendirme yaklařımında olduđu gibi, yönetime dayalı program deđerlendirme yaklařımı kapsamında yer alan deđerlendirme modellerinin de üstünlükleri ve sınırlılıkları bulunmaktadır. Bu yaklařıma dayalı modellerin üstünlükleri řunlardır:

- Bu yaklařım, deđerlendiricilere ve program yöneticilerine programların tasarlanıp, uygulandıktan sonra deđerlendirme sonuçlarını beklemek yerine program daha düşünce/tasarı aşamasındayken tartıřma fırsatı vermektedir.
- Program devam ederken programın her boyutuyla ilgili yapılan deđerlendirme sonuçları anında yöneticilere sunulmakta ve yöneticilerden de deđerlendiricilere geribildirim verilmektedir. Dolayısıyla programın yeni gereksinimlerden, kaynaklardan, yeni geliřmelerden, günlük iřlemlerden etkilenmemesi ve deđerlendiricilerin kararsız kalması olanaksızdır.
- Programın deđerlendirmenin çevre, girdi, sđreç ve ürün deđerlendirmesi olarak ayrı ayrı ele alınması karar vericilere deđerlendirme bilgilerinin anlaşılır, sistematik ve kolay sunulmasını sađlar.

Yönetim odaklı program deđerlendirme yaklařımının sınırlılıkları ise řunlardır:

- Programın değerlendirme yargılarının yöneticilerin elinde olması değerlendiricinin yalnızca bir görevli olarak algılanmasına ve değerlendirmenin güvenilirliğinin sorgulanmasına yol açabilir.
- Değerlendiricinin kararlarının yöneticiler ile uyuşmaması durumunda değerlendiricinin yöneticiler tarafından yetersiz olarak değerlendirilmesine ve işine son verilmesine sebep olabilmekte ve değerlendiricinin adil ve demokratik bir değerlendirme yapmamasına neden olabilmektedir.
- Değerlendirmede öncelikler, değerlendirme amaçları ve değerlendirme bütçesi net olarak belirlenmemişse değerlendirme karmaşık ve maliyetli olabilmektedir.

1.3.5.3. Uzman Odaklı Program Değerlendirme Yaklaşımı ve Modelleri

Uzmanlığa dayalı program değerlendirme yaklaşımı, bir kurum, program, ürün veya etkinliği değerlendirmede alandaki uzmanların mesleki uzmanlıklarını temel alan program değerlendirme yaklaşımıdır. Değerlendirilen olguya göre değerlendiricinin uzmanlığını dikkate alan yaklaşım eğitimden farklı alanlara kadar geniş bir çerçevede kullanılmaktadır. Yaklaşım, eğitsel bir programın değerlendirilmesinde program geliştirme ve değerlendirme uzmanına gereksinim duyarken, bir hastanenin hizmetlerinin etkililiğini değerlendirme sürecinde ise, doktorların, sağlıkçıların, eczacıların vb. uzmanlıklarına gereksinim duymaktadır. Doktora yeterlik sınavların için oluşturulmuş jüriler, kurum ve kuruluşları değerlendiren akreditasyon birimleri, makale değerlendirme için oluşturulmuş hakem kurullar, programın yerinde incelenmesi için oluşturulan kurullar bu tür bir değerlendirme yaklaşımının örnekleridir. Örnekler, bu yaklaşımın daha çok formal incelemeler, informal incelemeler, paneller ve bireysel incelemeler yoluyla uygulandığını göstermektedir. Tablo 5'te bu sınıflandırmanın sahip olduğu özellikler kısaca özetlenmiştir.

Tablo 5.

Uzmanlığa Dayalı Değerlendirme Yaklaşımının Değerlendirme Türleri

Türler	Uygulama Alanı	Ölçütler	Katılım	Değerlendirme Etkisi
Formal Uzman Değerlendirmeleri	Akreditasyon Kurulları	Kurulların belirlediği standartlar veya değerlendirme ölçütleri ve belgeler	Geniş katılımlı	Değerlendirme sonuçları çoğunlukla önceki statüde bir değişikliğe neden olur.
İnformal Uzman Değerlendirmeleri	Tez Jürileri	Ölçütler genellikle belirlenmez veya nadir durumlarda belirlenir	Geniş katılımlı	Değerlendirme sonuçları çoğunlukla önceki statüde bir değişikliğe neden olur.
Panel Değerlendirmeleri	Ajanslar, Büyük Ödül Jürileri	Değerlendirmeler bir kez yapıldığı için ölçütler yoktur	Geniş katılımlı	Değerlendirme sonuçları önceki durumda nadiren değişikliğe neden olur.
Bireysel Değerlendirmeler	Danışmanlık Hizmeti	Bireysel olarak bir danışmandan program, hizmet veya ürünle ilgili değerlendirme alınması amaçlandığı için standart ya da ölçütler yoktur.	Bireysel	Değerlendirme sonuçları önceki durumda nadiren değişikliğe neden olur.

Kaynak: Fitzpatrick, Sanders ve Worthen, 2004'den sadeleştirilmiştir.

Tablo 5 incelendiğinde, formal uzman değerlendirmeleri, informal uzman değerlendirmeleri ve panel değerlendirmelerinin geniş katılımlı olduğu, bu değerlendirmelerden elde edilen sonuçların bir önceki durumda değişikliğe neden olduğu görülmektedir. Formal ve informal uzman değerlendirmelerinde dikkate alınması gerekli bazı ölçütler varken; panel değerlendirmeleri ve bireysel değerlendirmelerde ölçütlere gereksinim duyulmamaktadır. Uzmanlığa dayalı program değerlendirme yaklaşımının bu değerlendirme türlerine ilişkin ayrıntılı açıklamalar aşağıda verilmiştir.

- *Formal Uzman Değerlendirmeleri:* Formal değerlendirmeler, belirli süreler içinde, resmi bir kurum, kuruluş, yapı ve organizasyon tarafından yürütülen uzmanlığa dayalı değerlendirme türlerindedir. Geniş katılımlı bir takım çalışması eşliğinde yürütülen değerlendirmeler genellikle önceki durumda bir değişiklikle son bulur. Değerlendirme sürecinde birimin kendisinin geliştirdiği ya da farklı birimlerce geliştirilen standartlar, yönergeler kullanılır. Bu tür değerlendirmelerin en önemli örneği akreditasyondur (Fitzpatrick, Sanders ve Worthen, 2004). Akreditasyon, birçok ülkede ve birçok sektörde topluma sunulan program ve hizmetlerin niteliğinin sistematik bir yaklaşımla güvence altına alınması için geliştirilen bir yöntemdir. Laboratuvarların, belgelendirme ve muayene kuruluşlarının üçüncü bir tarafça belirlenen teknik ölçütlere göre çalıştığının bağımsız ve tarafsız kuruluşlarca onaylanması olarak da tanımlanabilen akreditasyon, bir değerlendirme tekniği olarak, söz konusu kuruluşlarda güvenilirliğin sağlanması ve sürekliliği için

önemli bir araçtır (Aktan ve Gencel, 2007). Diğer bir ifadeyle, daha önceden belirlenmiş standartlara dayalı olarak, belirli bir değerlendirme sürecinden sonra bir kurum veya programın niteliğiyle ilgili olarak resmi yargının rapor olarak sunulması sürecidir. Ükelere göre farklılıklar göstermesine karşın genel olarak akreditasyon sürecinin bazı ortak özellikleri vardır. Bu özellikler şunlardır (Aktan ve Gencel, 2007; Gencel, 2001):

- Akreditasyon, belli beklentiler karşısında (standartlar, ölçütler), bir programın niteliğinin sorgulanmasını içerir.
- Akreditasyonun temelinde programın işlevi, tarihi ve amaçlarının bulunması vardır.
- Akreditasyonda, programların özerklik ve farklılıklarının olduğu gibi benimsenmesi söz konusudur.
- Akreditasyon, akredite edilen programların kamuya beklentilerini yeterince ya da daha fazla karşıladıkları konusunda güvence sağlar.
- Akreditasyon beş ve on yıllık sürelerle dayanan bir döngüdür. Ancak bazı özel durumlarda süreç hızlandırılabilir.

Çoğu ülkede programların değerlendirilmesinde kaynak olarak yaygın kullanılan akreditasyon süreci Türkiye’de de eğitimde özellikle son yıllarda yaygın olarak başvurulan bir süreçtir. Henüz program değerlendirme çalışmalarında başvurulmamasına karşın, YÖK tarafından üniversitelerin akreditasyonuna ilişkin hazırlanmış standartlar bulunmaktadır. Diğer taraftan kimi kurumlar uluslararası akreditasyon kurumlarına başvurarak kurum düzeyinde akredite edilmektedir.

- *İnformal Uzman Değerlendirmeleri*: Informal uzman değerlendirme birimlerinin de kendilerine özgü yapıları, değerlendirme sürecine ilişkin yönergeleri ve süreçte yer alan birçok uzman değerlendiricisi vardır. Buna karşın formal değerlendirme birimlerinde geliştirilen ve yayınlanan resmi standartlarından yoksundurlar. Bu nedenle kendilerine özgü değerlendirme sistemlerini geliştirmişlerdir. Bir lisansüstü öğrencisinin seçtiği alandaki uzmanlardan oluşan tez yürütme jürisi informal uzman değerlendirmesine bir örnektir. Bu tür bir değerlendirmede genellikle her bir öğrencinin değerlendirilmesi için de farklı ölçütler geliştirilmiştir (Fitzpatrick, Sanders ve Worthen, 2004).

- *Panel Değerlendirmeleri:* Formal ve informal değerlendirmelerden farklı olarak bir makalenin, bir programın ya da bir projenin değerlendirilmesinde alan uzmanlarının yer aldığı ve tek oturumluk olarak düzenlenen panellerdir. Geniş bir uzman katılımıyla gerçekleştirilen değerlendirmelerde genellikle değerlendirme sonuçları nadiren değişikliğe neden olur. Değerlendirme süreci sonunda ulaşılan sonuçlar programın paydaşları için daha çok öneri niteliği taşırlar. Tek oturumlu ve kısa aralıklarla yapıldığı için değerlendirmede dikkate alınacak standartlara da başvurulmaz (Fitzpatrick, Sanders ve Worthen, 2004) .
- *Birey Değerlendirmeleri:* Bireysel olarak bir danışmandan program, hizmet veya ürünle ilgili değerlendirme alınmasının amaçlandığı birey değerlendirmelerinde izlenilecek standart ya da ölçütler yoktur. Bireyin bireysel uzmanlığına dayalı olarak yapılan bu değerlendirmeler öneri niteliğinden öteye geçmemekle birlikte, özellikle özel öğretim kurumlarınca eğitim programlarının değerlendirilmesinde sıklıkla başvurulmaktadır. Uygulama kolaylığı ve ekonomik olan bu değerlendirmede bireyin uzmanlığı oldukça önemlidir (Fitzpatrick, Sanders ve Worthen, 2004) .

Yukarıda açıklanan uzman odaklı program değerlendirme türlerinin yanı sıra uzman odaklı program değerlendirme yaklaşımını temel alan tek program değerlendirme modeli Eisner'in "Eğitsel Uzmanlık ve Eleştiri Modeli"dir.

Eğitsel Uzmanlık ve Eleştiri Modeli

Sanatta uzmanlık, eğitsel değerlendirmeden çok geleneksel kültürde sanatsal eserlerin değerlendirilmesi gibi olgularla ilişkilendirilen bir kavramdır. Benzer bir şekilde, eleştiri kavramı da eğitsel değerlendirme ile değil edebiyatın ya da sanatsal çalışmaların değerlendirilmesiyle ilişkilendirilen bir sözcüktür. Bununla birlikte, 1970'li yıllarda bu iki kavram eğitsel değerlendirme alanında, özellikle Elliot Eisner'in (1976; 1985) çalışmalarında sık kullanılmıştır.

"Eğitsel uzmanlık" ve "eğitsel eleştiri" kavramları modelin temel kavramlarıdır. Her iki kavram da yukarıda belirtildiği gibi sanata dayalı olarak ortaya çıkmıştır. Eisner'e (1976; 1985) göre, "takdir etme ve değerlendirme sanatı" olarak tanımlanan uzmanlığın amacı

bir süreci ya da bir nesneyi oluşturan niteliklerin farkındalığını ortaya çıkarmak ve bunların önemini vurgulamaktır. Örneğin, uzmanlık tiyatrodaki sete, konuya, oyuncuların performansına, hıza ve tempoya odaklanabilirken eğitimde ise programın niteliğine, öğrenci etkinliklerine, öğretimin niteliğine, öğrencilerin öğrenme süreçlerine, araç-gereçlere vb. odaklanabilir. Yine Eisner (1976; 1985) tarafından “söyleme ve açıklama sanatı” olarak tanımlanan eleştiri ise uzmanlıkla doğrudan ilişkili bir kavramdır. Uzmanlığın paylaşılabilmesi, ilgililere ulaştırılabilmesi için eleştiriye gereksinim duyulmaktadır. Burada eleştiri bir şey hakkında olumsuz yorumlar yapma anlamına gelmemekte, tersine bir nesnenin algısını yeniden oluşturma anlamında kullanılmaktadır. Bir sanat eleştirmeni bir heykele ya da resme farklı bakış açılarıyla bakılmasını sağlamayı, onları anlaşılır kılabilmeyi amaçlıyorsa, bir eğitsel eleştirmen de sınıf kurallarını, öğretimin niteliğini, öğrenci davranışlarındaki değişimleri vb. durumları ortaya çıkarmayı amaçlar. Bu durumda eğitimde eleştirmen öğreticinin kendisidir. Eisner’e (1976) göre, uzmanın bir eleştirmenlik rolü yoktur, yalnızca yapılan işleri değerlendirir, takdir eder. Oysa eleştirmen yapılması gerekenlerin oluşturulmasında uzmana gereksinim duyar. Kısaca programın, projenin, öğretimin etkisi ile ilgili farkındalık uzmanlığın; bu farkındalığın kamuya duyurulması ise eleştirmenin görevidir. Eisner bu iki kavramı eğitimde değerlendirme alanında kullanarak *Eğitsel Uzmanlık ve Eleştiri Modelini* ortaya koymuştur.

Öğretimin sanatsal beceri gerektiren bir etkinlik olduğunu vurgulayan Eisner, eğitimin kültürel bir sanat olduğu ve bunun bireyden bireye, çevreden çevreye değişen bir süreç olduğunu ifade etmektedir (Eisner, 1985). Eisner (1985) eğitsel değerlendirmenin amacının yalnızca ürünleri gözden geçirme ve süreç içindeki etkinlikleri ölçme değil aynı zamanda öğreticinin kazanacağı beceriyi artırma olduğunu da belirtmektedir.

Eisner’in uzmanlık ve eleştiri kavramlarını değerlendirme alanyazınına ekleme düşüncesi, 70’li yıllarda Amerika’da ve Avrupa’da değerlendirme alanındaki genel hareketin bir parçasıdır. Bu yıllar, yarı-deneysel yöntemlerin ve diğer bilimsel yöntemlerin eğitsel değerlendirme için yeterli veya uygun olmadığına farkına varıldığı yıllar olarak bilinmektedir. Geleneksel bilimsel değerlendirme yaklaşımlarının yetersizliği ya da uygun olmaması hakkında artan şüpheler birçok eğitim uzmanının değerlendirmenin şekil ve işlevlerini tekrar düşünmesini sağlamıştır. Bu uzmanlar,

antropoloji, gazetecilik, felsefe ve hukuk vb. farklı akademik alandan etkilendiler ve doğacı-yapıcı değerlendirme, ihtiyaca karşılık veren değerlendirme, amaçtan bağımsız değerlendirme, zıt savunma değerlendirmesi gibi farklı değerlendirme modelleri geliştirdiler (Mathison, 2005). Eisner de bu akımdan etkilenmiş ve değerlendirme modelini sanat ve estetik alanları üzerine oluşturmuştur.

Eisner'e (1976) göre, program değerlendirici eğitsel sanat uzmanına; değerlendirme süreci de sanat eleştirisine benzemektedir. Bu açıdan bakıldığında bir değerlendirici bir programın, bir sınıfın ya da bir okulun eğitsel eleştirisini yaparken öncelikle gördüklerini betimler, yorumlar ve son olarak değerlendirir (Eisner, 1976; 1985). Eisner de modelini üç eylemi yansıtan üç temel boyut üzerine geliştirmiştir. Bu boyutlar şunlardır:

- *Betimleyici Boyut:* Eisner'e (1976) göre, eğitsel eleştirinin betimsel yönü/boyutu değerlendirilecek programı, sınıfı, okulu vb. olduğu sadelikte tanımlama ile ilgilidir. Eisner (1985), eleştirinin "sözel ifadelerinin en keskin olması gerektiği" boyutun betimleme olduğunu belirtmektedir. Bu nedenle sanat eleştirisinde olduğu gibi, eğitsel eleştiride vurgu, değerlendirilen olgunun estetik boyutlarını ifade etmede kullanılan edebi dil ve mecazlardadır. Betimleyici eğitsel eleştiri okuyucuya sınıfta sorulan soruların sayısı ve türleri, tartışmalara ayrılan süreyi, öğretmenin ya da sınıfın öğrencide ne gibi bir izlenim bıraktığı hakkında bilgi sunar. Diğer bir deyişle betimleyici eğitsel eleştiri niteliklerin ayrıntılarına değinmeden edebi bir dil ve mecazlar yoluyla tanımlanmasını amaçlar.
- *Yorumlayıcı Boyut:* Eisner (1985), eğitsel eleştirinin yorumlayıcı yönünün/boyutunun eleştirinin sosyal ortamdaki çeşitli eylemlerin anlamını ve önemini anlama çabasıyla ilişkili olduğunu belirtmektedir. Eğitsel eleştirinin bu yönü, eğitsel ortamlardaki etkinlikleri yorumlamada eleştirinin çoklu kuramları, bakış açıları, modelleri kullanma uzmanlığını ortaya çıkarır (Koetting, 1988). Örneğin eleştirinin bir sınıf ortamında havaya kalkan istekli ellerin öğretmen ve öğrenci için ne anlam ifade ettiğini, sınıf ortamının tüm katılımcılar için ne anlama geldiğini vb. yorum gerektiren soruları yanıtlayabilmelidir. Eisner (1976) bu tür eylemlerin değerlendirici tarafından yorumlanması gerektiğini, bunun için de

eleştirmenlerin yorumlama sürecinde geniş bir kuramsal bilgiye ve ölçütler dizisine sahip olmaları gerektiğini ifade etmektedir.

- *Değerlendirici Boyut:* Eğitsel eleştirinin son yönü/boyutu değerlendirmedir. Bu boyutta eleştirmen, yorumladığı deneyimlerin/etkinliklerin eğitsel önemini ve etkisini değerlendirir. Bu süreçte eleştirmenin deneyimler hakkında bir yargıya varabilmesi için izlemesi gereken bazı eğitsel ölçütlerin olması zorunludur. Koetting'e (1988) göre bu durum eğitsel eleştirinin kuralcı /normatif özelliğini de ortaya çıkarır.

Uzman Odaklı Program Değerlendirme Yaklaşımının Üstünlükleri ve Sınırlılıkları

Uzman odaklı program değerlendirme yaklaşımı, kurumsal etkililiğin değerlendirilmesinde, değerlendirme yönergelerinin ve ölçütlerin geliştirilmesinde, kurumsal gelişimin sağlanmasında, genel olarak topluma faydalı eğitim amaçlarının oluşturulması ve düzenlenmesinde, yeni kurulan ve gelişen kurumlara yardım ve danışmanlığın sağlanmasında, akademik özgürlük ile eğitsel etkililiklerin gerçekleştirilmesinde önemli katkılar sağlar (Fitzpatrick, Sanders ve Worthen, 2004). Ayrıca yaklaşım, geniş uygulama alanına sahip olması, kolay ve uzun zaman gerektirmeden uygulanabilmesi ve değerlendirilen programın niteliği hakkında gözlemlere dayalı nitel veriler sağlaması yönünden de tercih edilen bir yaklaşımdır (Fitzpatrick, Sanders ve Worthen, 2004).

Uzman odaklı program değerlendirme yaklaşımlarının olumlu olarak uygulandığı kullanıldığı alanlardan biri akreditasyondur. Bağımsız kurumlar olan akreditasyon kurumlarının yapacağı değerlendirmeler kuşkusuz daha güvenilir ve geçerli olacaktır. Etkili ve sağlam programların resmi olarak tanınması, programın değerlendirilmesini sağlayacak bir mekanizmanın yaratılması, paydaşların programın niteliği ve sağlanan eğitimle gurur duymalarının sağlanması, program paydaşları arasında iletişimin geliştirilmesi ve katılımın desteklenmesi akreditasyonun sağlayacağı en önemli katkılardandır.

Yukarda verilen güçlü yönlerine karşın, uzman odaklı program değerlendirme yaklaşımının zayıf yönleri de bulunmaktadır. Öncelikle yaklaşımın odağında uzman

bireyler yer aldığından uzmanın seçiminde oldukça titiz davranılması bir zorunluluktur. Çünkü değerlendirme kararlarını verecek uzmanın eğitimi, deneyimi ve algısal yetenekleri süreçte çok önemli bir rol oynamaktadır. Diğer taraftan değerlendirmenin yalnızca uzmanın görüşlerine göre yapılması da sürecin nesnellliğini etkileyebilmektedir. Örneğin panel değerlendirmelerinde bir panelde yer alan uzmanların programla ilgili değerlendirmeleri ile başka bir panelde yer alan uzmanların aynı programa ilişkin değerlendirmeleri farklılık gösterebilmektedir (Fitzpatrick, Sanders ve Worthen, 2004).

1.3.5.4. Tüketici Program Odaklı Değerlendirme Yaklaşımı ve Modelleri

Tüketici odaklı program değerlendirme yaklaşımı, eğitim programları, çalıştaylar, eğitsel yazılımlar, eğitim araç-gereçleri, hizmetiçi eğitimler vb. eğitsel ürünler ve hizmetler hakkında bilgi toplamakla yükümlü bireyler ya da bağımsız kuruluşlar tarafından geliştirilmiş ve desteklenmiş bir program değerlendirme yaklaşımıdır (Fitzpatrick, Sanders ve Worthen, 2004). Günümüzde eğitsel ürünlere ve hizmetlere olan talepte büyük bir artış görülmekte, buna karşın bu ürünlerin ve hizmetlerin tüketicilerin gereksinimlerini ne kadar karşıladığının değerlendirilmesi de ihmal edilmektedir. Bu nedenle bazı değerlendiriciler, tüketici eğitimi ve ürünlerin etkililiği konusunda nesnel değerlendirmeler yapılabilmesi için acil önlemler alınması gerektiğini belirtmektedirler (Mathison, 2005). Bunun sağlanması için de ürün değerlendirme formları, ürün değerlendirme kontrol listelerinin kullanılması önerilmiştir. Bu yönüyle bakıldığında yaklaşım toplam değerlendirme özelliklerini taşıyan bir program değerlendirme yaklaşımıdır.

1960'lı yıllarda yeni eğitsel ürünlerin pazara girmesiyle birlikte tüketici odaklı program değerlendirme yaklaşımı da program değerlendirme alanında kendisine önemli bir yer bulmaya başlamıştır. Çünkü daha önceki yıllarda eğitimcilerin elindeki tek eğitsel ürünler olan ders kitaplarına eğitsel yazılımlar, eğitsel medya vb. araç-gereçler de eklenmiştir. Yıllar içinde eğitsel ürün yelpazesi genişlemiş, ürünlere olan yatırımlar artmıştır. Bu durum tüketici odaklı program değerlendirme yaklaşımının hızla gelişmesini sağlamıştır. Bu anlamda ilk büyük katkıyı da Scriven yapmıştır. Scriven (1967) eğitsel ürünlerin/programların değerlendirilmesinde kullanılacak şu ölçütleri önermiştir:

- Ürünün/programın önemli eğitsel amaçları kazandırıp kazandırmadığına ilişkin yeterli kanıtlar
- Ürünün/programın sosyal amaçlar gibi eğitsel olmayan ikincil amaçları kazandırıp kazandırmadığına ilişkin yeterli kanıtlar
- Ürüne/programa ilişkin izleme ve gözlem sonuçları
- Ürünün/programın öğretmen, öğrenci, aile, yönetici, okul vb. üzerinde ikincil ve istenmeyen etkileri ile diğer tesadüfî olumlu ve olumsuz etkileri
- Ürünün/programın kimler açısından yararlı ve kullanılabilir olduğuna ilişkin bir dağılım
- Ürüne/programa ilişkin etik durumlar (adaletsiz ceza kullanımı, tartışmalı içerik)
- Ürünün/programın maliyetleri

Eğitsel ürünlerin değerlendirilmesinde kullanılan bu ölçütler, sonraki yıllarda devlet destekli araştırma-geliştirme merkezlerinde ve bölge eğitim lâboratuarlarında yapılan değerlendirme sonuçlarına dayalı olarak Scriven (1974) tarafından yeniden gözden geçirilerek geliştirilmiş, yeni ölçütler eklenmiş ve bir ürün değerlendirme kontrol listesine dönüştürülmüştür. Doksanın üzerinde eğitsel ürünün değerlendirilmesinde kullanılarak geliştirilen bu kontrol listesi, “Gereksinim”, “Pazar”, “Performans”, “Maliyet Etkililiği” ve “Destek Alanları” olmak üzere beş temel alan ve “performans” standardı altında dokuz göstergeden oluşmaktadır. Fitzpatrick, Sanders ve Worthen (2004) bu standartları şöyle sıralamaktadırlar:

- *Gereksinim*: Üründen dolaylı ya da doğrudan etkilenen kişi, ürünün sosyal önemi ve ürüne gereksinim duyulduğuna dair kanıtlar
- *Pazar*: Ürünü yaygınlaştırma planı, ürün büyüklüğü, ürünün pazardaki önemi
- *Performans*: Performans alanı altında dokuz standart yer almaktadır. Bu standart maddeleri şunlardır:
 - ▀ Alan Denemeleri: Ürüne ilişkin son uygulamanın etkililiğine ilişkin kanıtlar
 - ▀ Doğru Tüketici: Öğrenciler, öğretmenler, okul görevlileri, kurum çalışanları,

vergi ödeyenler gibi uygun tüm tüketiciler üzerinde ürün denemeleri

- ▶ Karşılaştırma: Önemli rakipler hakkında karşılaştırmalı veriler
 - ▶ Kullanım: Ürünlerin bir haftadan bir aya ya da bir aydan bir yıla vb. kadar kullanımından sonra etkilerine ilişkin kanıtlar
 - ▶ Yan Etkiler: Uzun dönem kullanımları sırasında ve sonrasında ürünlerin istenmeyen çıktılarını araştıran bağımsız araştırma sonuçları
 - ▶ Süreç: Ürünün kullanımının etik yönden uygunluğunu belirlemede, ürünle ilgili haklı şikâyetleri ortaya koymada ve ürün tanımlamalarını gözden geçirmede kullanılan kanıtlar
 - ▶ Nedensellik: Ürünün etkililiğini deneysel ve yarı deneysel araştırmalar yoluyla sağlam bir nedene dayandırmada kullanılacak kanıtlar
 - ▶ İstatistiksel Önem: Ürün etkililiğine ilişkin istatistiksel kanıtlar
 - ▶ Eğitsel Önem: Ürünün bağımsız değerlendirmeler, uzman değerlendirmeleri, madde analizi ve testlerin ham puanlarına dayalı değerlendirmeler vb. yoluyla vurgulanan eğitsel önemi
- *Maliyet Etkililiği:* Aynı özellikteki diğer eğitsel ürünlerin fiyatlarıyla yapılan karşılaştırmalar, ayrıntılı maliyet etkililik analizi
 - *Destek:* Pazar sonrası ürün hakkında bilgi toplamaya, ürünü geliştirmeye, dönük planlar

Eğitsel ürünlerin değerlendirilmesine dönük geliştirilen yukarıdaki kontrol listesinin yanı sıra Scriven (1991a) özellikle program değerlendirme çalışmalarına yönelik de bir kontrol listesi geliştirmiştir. Değerlendirme Kontrol Anahtarı (Key Evaluation Checklist) olarak adlandırılan bu kontrol listesi, “Başlangıç”, “Altyapı”, “Değerlendirme” ve ”Sonuçlar” adı altında dört bölüm ve 15 maddeden oluşmaktadır. Bu bölümler ve altında yer alan maddeler şunlardır (Scriven, 1991a):

- *Başlangıç:* Program değerlendirme kontrol listesinin bu bölümü, program değerlendirmenin tasarım ve uygulanmasıyla ilgili olmamasına karşın bir program değerlendirme raporunda yer alması gereken özet, önsöz ve yöntem standartlarını içerir.

- *Altyapı:* Kontrol listesinin bu bölümü bir sonraki bölümde gereksinim duyulacak programın bağlamı (uygulandığı çevre) ve doğasını ortaya çıkaran bir dizi araştırmayı içerir. Programın geçmişi ve bağlamı, programın tanımları ve açıklamaları, tüketiciler, kaynaklar ve bir yargıya varmada temel alınacak değerler bu bölümün temel standartlarını oluşturur.
- *Değerlendirme:* Kontrol listesinin bu bölümü bir önceki bölüme bağlı kalınarak planlanan etkinliklerin yürürlüğe konmasıyla ilgilidir. Süreç, çıktılar, maliyetler, karşılaştırmalar ve genelleştirme standartları bu bölümün standartlarını oluşturmaktadır.
- *Sonuçlar:* Kontrol listesinin son bölümü program değerlendirme çalışması tamamlandıktan sonraki işlemlerle ilgilidir. Toplam önem, öneriler, açıklamalar ve tahminler, sorumluluklar, rapor ve destek ile meta-değerlendirme standartları bu bölümün altında yer almaktadır.

Eğitim programını bir eğitsel ürün olarak gören ve öğrencileri ve aileleri de tüketici olarak bu program değerlendirme modelinde, değerlendirme bu anlayışa dönük olarak ürün değerlendirme kontrol listeleri yoluyla yürütülmektedir.

Tüketici Odaklı Değerlendirme Modelinin Üstünlükleri ve Sınırlılıkları

Tüketici odaklı program değerlendirme yaklaşımı, değerlendirme yoluyla eğitsel ürünlerin niteliklileri, etkililiği vb. boyutlarda kamuoyunun yanlış bilgilendirilmesini ve satış hilelerinden etkilenmemesini sağlamıştır. Bu farkındalığın artması beraberinde üreticiler arasında rekabeti de arttırmış ve eğitsel ürünlerin daha nitelikli, sağlıklı, ergonomik ve ekonomik olmasını sağlamıştır. Tüketicilerin nitelikli ürün ve program geliştirme çabaları da ürünlerin ve programların değerlendirilmesinde kullanılabilecek bazı standartların ve ölçütlerin geliştirilmesine katkıda bulunmuştur. Tüm bu olumlu yönlerine karşın değerlendirme sürecinin maliyeti ve harcanan zamanın tüketiciye ekonomik yönden olumsuz yansımaları yaklaşımın zayıf yönü olarak görülmektedir (Stufflebeam, Madaus ve Kellaghan, 2002). Ayrıca ürün gelişiminde uyulması gerekli olan sıkı kurallar ve standartların yaratıcılığı engelleyebilme olasılığı da yine yaklaşımın bir diğer zayıf yönünü ortaya koymaktadır. Fitzpatrick, Sanders ve Worthen (2004) bu

olumsuzlukların ortadan kaldırılabilmesi için eğitsel ürünlerin onaylanıp, pazara sürülmeden önce denenmesi gerektiğini vurgulamaktadırlar.

1.3.5.5. Katılımcı Odaklı Program Değerlendirme Yaklaşımı ve Modelleri

Amerika Birleşik Devletleri'nde değerlendirmenin devlet tarafından denetlenmeye başlanmasından sonraki ilk yıllarda çoğu değerlendirme kuramcısı program değerlendirme çalışmalarında benimsedikleri ya da uyarladıkları geleneksel kuralcı bilim felsefesini, sosyal bilimlerde kullanılan yöntemleri ve eğitim araştırmalarını kullanmışlardır. Buna karşın 1967 yılında Stake, Scriven, Eisner, Stufflebeam vb. bazı değerlendiriciler duyarsız ve mekanik olarak nitelendirdikleri bu yaklaşımların eğitsel değerlendirmelerde kullanılmasına karşı çıkmışlardır (Mathison, 2005). Bu değerlendiriciler, değerlendiricilerin değerlendirilen programda gerçekte nelerin olup bittiğiyle değil; çoğunlukla programın amaçlarını belirleme ve sınıflama, ayrıntılı değerlendirme sistemlerini seçme ve planlama, savunulabilir nesnel veri toplama araçları geliştirme, uzun teknik raporlar hazırlama vb. işlemler ile ilgilendiklerini savunmuşlardır. Geleneksel değerlendirme yaklaşımlarını eleştiren bu değerlendiriciler, çoğu kapsamlı değerlendirme çalışmasının değerlendirici tarafından programın uygulandığı ortam görülmesizin ve o ortamda bulunulmaksızın değerlendirildiğini belirtmektedirler. Birçok değerlendirici ve eğitsel topluluk bu eleştirilere hak vermiş ve yapılan çalışmalarda geçen numaraların, figürlerin, tabloların, şekillerin gerçekte neyi yansıtmak istediği ile sorgulamalara başlamışlardır. Bu sorgulamalarda varılan sonuç ise program değerlendirme çalışmalarında insan etkeninin eksik bırakıldığıdır. Sonuç olarak değerlendirmede program paydaşlarının katılımını ve program etkinliklerinde ve uygulama ortamında aracısız deneyimleri vurgulayan yeni bir değerlendirme yaklaşımı ortaya çıkmıştır. Katılımcı odaklı program değerlendirme yaklaşımı olarak adlandırılan ve 1970'li yıllardan sonra hızla yaygınlaşan bu genel yaklaşımın temel amacı bireye yönelik hizmetlerle ilgili beklentileri, sorunları ve çözümleri gözlemlemek ve tanımlamaktır.

Fitzpatrick, Sanders ve Worthen (2004), katılımcı odaklı program değerlendirme yaklaşımını temel alarak yürütülen program değerlendirme çalışmalarının aşağıdaki özelliklere sahip olduğunu belirtmektedir:

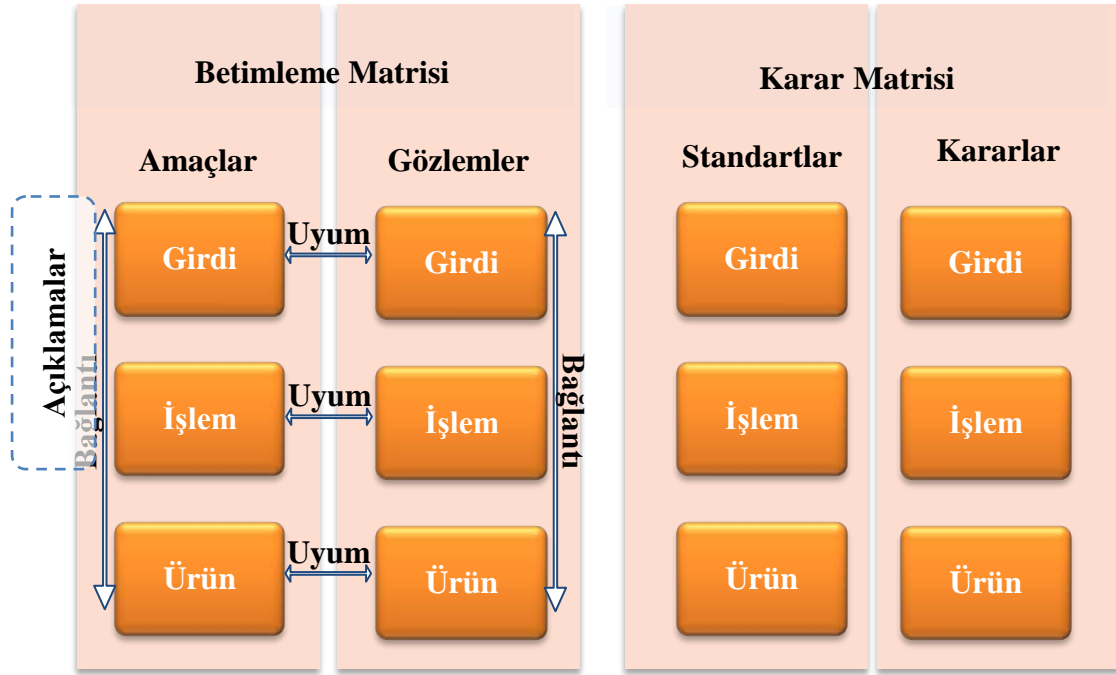
- Program değerlendirme çalışmaları tümevarımsal mantığa dayanır.
- Program değerlendirme çalışmaları birbirinden farklı birden fazla veri kullanır. Çalışmalarda değerlendirilen programın öznel ya da nesnel, nitel ya da nicel örnekleri kullanılır.
- Katılımcı odaklı program değerlendirme yaklaşımını temel alarak yürütülen program değerlendirme çalışmalarında standart bir plan takip edilmez. Değerlendirme süreci, katılımcıların etkinliklerdeki deneyimlerine bağlı olarak gelişir.
- Katılımcı program değerlendirme çalışmalarında tek bir gerçekten çok, çoğul gerçekler dikkate alınır. Bireyler olayları farklı şekilde görürler ve yorumlarlar. Hiç kimse okulda olan her şeyi tam olarak bilemez ve kimsenin bakış açısı doğru olarak kabul edilmez. Çünkü birey sadece kendi deneyimlerinin gerçekliğini bilebilir, bu nedenle tüm bakış açıları doğru olarak kabul edilir. Değerlendiricinin buradaki temel görevi bu gerçeklerin hepsini programın yapısına zarar vermeden görmek ve betimlemektir.

Bu yaklaşıma dayalı olarak yıllar içinde çeşitli program değerlendirme modelleri geliştirilmiştir. Bunlardan en önemlileri, Stake'in program değerlendirmenin genel görünüşünü ortaya koyduğu, değerlendiriciler için bir değerlendirme veri toplama çerçevesi hazırladığı modeli; yine Stake tarafından geliştirilen fakat kendisi tarafından bir değerlendirme modeli olarak kabul edilmeyen yanıtlayıcı değerlendirme modeli; program paydaşlarını değerlendirmenin merkezine alan esnek bir değerlendirme modeli olan katılımcı değerlendirme modeli ve doğal değerlendirme modelidir (Stake, 1972).

Stake'in Değerlendirme Matrisi

Katılımcı odaklı program değerlendirme yaklaşımının eğitimde hızla yaygınlaşmasına katkıda bulunan ilk değerlendirme uzmanı Stake'tir. Bu konudaki ilk çalışması 1967 yılında yayımladığı "Eğitimde Değerlendirmenin Görünüşü" adlı eseridir. Bu çalışmada Stake (1967), program değerlendirme sürecinde paydaş kararlarının önemine dikkati çekmekte ve paydaş katılımının program değerlendirme çalışmalarının vazgeçilmez parçası olduğunu vurgulamaktadır. Stake'in bu çalışması sonraki yıllarda yeni yetişen değerlendiricilerin program değerlendirme alanına ilişkin anlayışlarını büyük ölçüde

etkilemiş ve birbirinden farklı katılımcı değerlendirme modelleri geliştirilmiştir (Fitzpatrick, Sanders ve Worthen, 2004). Değerlendirmenin iki temel eyleminin *betimleme* ve *karar* olduğunu vurgulayan Stake, bu çalışmasında değerlendiricilere veri toplama ve yorumlama süreçlerinde yardımcı olması için bir değerlendirme çerçevesi de geliştirmiştir (Şekil 9).



Şekil 9. Stake'in Program Değerlendirme İçin Veri Toplama Planı

Kaynak: Stake (1967)'den alınmıştır.

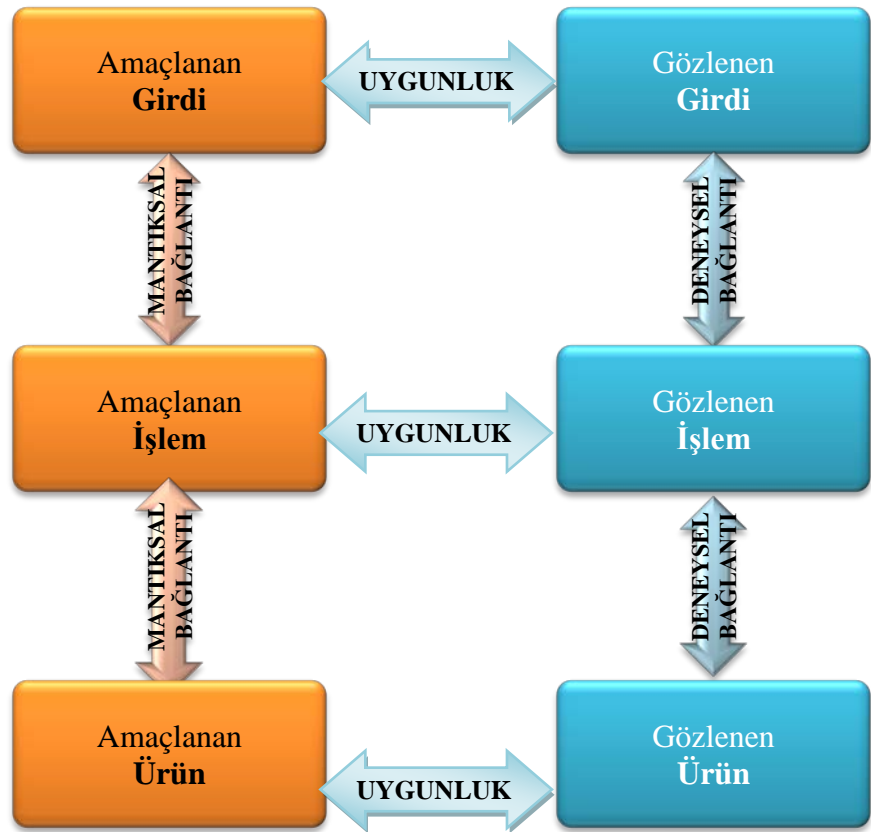
Şekil 9'da gösterilen değerlendirme çerçevesi açıklamalar, betimleme ve karar matrislerinden oluşmaktadır. Stake'in (1967) bir eğitim programının değerlendirilebilmesi için değerlendiricinin, programın uygulandığı ortama ilişkin bilgiler, programın geçmişi, tanımı vb. açıklamalar doğrultusunda programı betimlemesi/tanımlaması ve daha sonra da ortaya çıkan sonuçlara göre program hakkında bir karar vermesi gerektiğini ifade ettiği bu değerlendirme çerçevesinde her iki matrisin ortak noktası girdi, işlem ve ürün boyutlarıdır. Bu boyutlardan girdi, öğretme-öğrenme süreci başlamadan önce var olan ve ürünle bağlantılı olan durumları ifade eder. Örneğin, öğrencinin dersten önceki durumu, tutumları, ön bilgileri, ilgisi, isteği vb. durumlar girdileri oluşturur. Kısaca girdiler "giriş davranışları"dır (Stake, 1967). Değerlendiricinin görevi bu girdileri ayrıntılı bir şekilde tanımlamaktır. İşlem boyutu, kısaca eğitim sürecini

kapsayan öğretmen-öğrenci, öğrenci-öğrenci, veli-öğretmen vb. arasındaki sayısız etkileşimi ifade eder. Örneğin, bir film sunumu, bir sınıf-içi tartışma, ödevlerle ilgili bir açıklama, yazılı yoklama vb. eylemler bir işlemdir. Matrisin son boyutu olan ürün ise bir eğitsel deneyim sonucu ortaya çıkan beceri, başarı, tutum, beklenti vb. çıktıları ifade eder. Ürünler öğretimin öğretmen, yönetici ve diğerleri üzerindeki etkililiğine ilişkin ölçümleri de içerebilir. Değerlendirmede dikkate alınacak ürünler yalnızca bir ders sonunda gözle görülebilir olanları değil, uzun bir süreç sonunda ölçülebilecek uygulamaları, transferleri ve yeniden öğrenmeleri de içerebilir. Kısaca, ürünler hemen ya da uzun zamanlı, bilişsel ya da çabaya dayalı, bireysel ya da toplumsal olsun bir eğitimin sonuçlarıdır.

Matriste yer alan iki temel eylem olan betimleme ve karar süreçlerine ilişkin ayrıntılı bilgiler aşağıda verilmiştir.

- *Betimleme Süreci:* Stake (1967), bir eğitim programı için betimleyici değerlendirme verilerinin işlenmesinin iki temel yolu olduğunu belirtmektedir: Amaçlar ve gözlemler arasındaki *uyum* ile girdi, işlem ve ürün arasındaki *bağlantı*.
 - ▀ *Uyum/uygunluk:* Değerlendirici girdi, işlem ve ürün boyutlarında programın temel felsefesine (mantığı) dayalı olarak istenenleri/ amaçları ortaya koyar ve çeşitli veri toplama araçlarıyla yapılan ölçümler sonucunda gözlenen değerler ile amaçlanan arasındaki uyuma/uygunluğa bakar. Eğer iki veri arasında bir uyum varsa, istenenler/amaçlar gerçekleşmiş demektir. Şekil 10'da ilk satırda değerlendirici amaçlananlar ile gözlenenler arasındaki farklılığı ve uyumu betimlemek zorundadır. Zor da olsa tümüyle bir uyumdan söz edebilmek için girdi, işlem ve üründe uyumun sağlanması gereklidir. Ancak amaçlanan ile gözlenen arasında sağlanan bu uyum, ürünün geçerli ya da güvenilir olduğu konusunda kesin bilgiler vermez.
 - ▀ *Bağlantı:* Program değerlendirmede amacın eğitimin niteliğini arttırmak olduğu düşünülünce, değerlendiricinin temel görevlerinden biri eğitsel ürünlerin girdiler ve işlemlerle olan ilişkisini belirlemektir. Örneğin, seller ile ilgili bir film öğrencilerin doğal yaşamı koruma kanunlarını anlamalarına yardımcı olmak amacıyla programa alınabilir. Burada filmin programa alınması amaçlanan işlem, öğrencilerin doğal yaşamı koruma kanunlarını

öğrenmeleri de amaçlanan çıktıdır. Değerlendiriciye sorulabilecek “Bu eylem ile ürün arasında mantıksal bir ilişki var mı? sorusu ise aradaki bağlantıyı göstermektedir.Şekil 10’da görüldüğü gibi, burada amaçlanan/gözlenen ürün ile amaçlanan/gözlenen girdi ve işlem arasında mantıklı bir ilişkinin olup olmadığı değerlendirilmektedir. Bu değerlendirmede amaçlananlar arasındaki ilişki/bağlantı mantıksal olarak kurulur, yani bu değerlendirmede ölçüt mantıktır. Değerlendiricinin geçmiş eğitsel deneyimleri bu noktada büyük rol oynar. Gözlenenler arasındaki ilişki/bağlantının değerlendirilmesi de deneysel kanıtlara dayalıdır. Örneğin “öğretmen matematikte yeteri kadar uzman olmasa da bu sınıf matematikte hızlı bir şekilde gelişme gösterdi ” demek deneysel bir çalışmayı gerektirir.



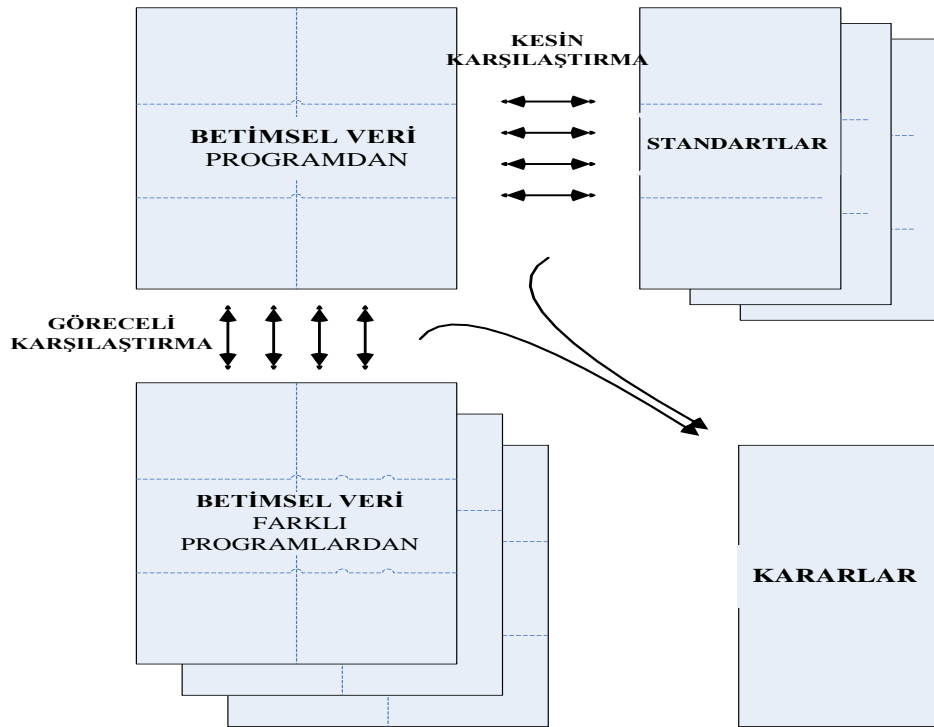
Şekil 10. Betimsel Veri İşleme Süreci

Kaynak: Stake (1967)'den alınmıştır.

Şekil 10’da görüldüğü gibi program değerlendirmede betimsel veri işleme sürecinde iki temel eylem; amaçlanan ve gözlenenler arasında uygunluğun

ve mantıksal/deneysel bağlantının araştırılmasıdır. Değerlendirici tarafından belirlenen bu bağlantı ve uygunluklar standart ve karar aşamasında kullanılırlar.

- *Karar Süreci:* Stake'e (1967) göre, eğitimin temel amacı mükemmelleşmektir. Mükemmelleşmenin ölçülmesi için de standartlara gereksinim vardır. Stake, eğitim programlarının "standart temelli" değerlendirmeye bağlı olmadığını, oysa programların değerlendirilmesinde standartların da kullanılması gerektiğini vurgulamaktadır. Bunun da iki yöntemle gerçekleştiğini belirtmektedir: Kişisel yargıları yansıtan kesin standartlar ve farklı programların özelliklerini yansıtan göreceli standartlar. Örneğin, bir değerlendirici İngilizce programını, İngilizce programının ne ve nasıl olması gerektiğine ilişkin görüşlere dayalı olarak ya da diğer İngilizce programlarının özelliklerine göre karşılaştırarak değerlendirebilir. Değerlendiricinin bu süreçteki karşılaştırmaları ve kararları Şekil 11'de gösterilmiştir.



Şekil 11. Program Değerlendirmede Karar Süreci

Kaynak: Stake (1967)'den alınmıştır.

Şekil 11’de sağ üstte değerlendirilen programla ilgili bir önceki aşamada elde edilen verilerin kesin bir şekilde karşılaştırılacağı bir dizi standart yer almaktadır. Standartların katmanlar halinde olması birden çok referans grupları ve görüşlerin olabileceğini göstermektedir. Şeklin sol altında ise değerlendirilen programla ilgili bir önceki aşamada elde edilen verilerin göreceli olarak karşılaştırılabileceği farklı programlar yer almaktadır. Her iki tür karşılaşmadan elde edilen sonuçlara dayalı olarak değerlendirici tarafından karara varılır.

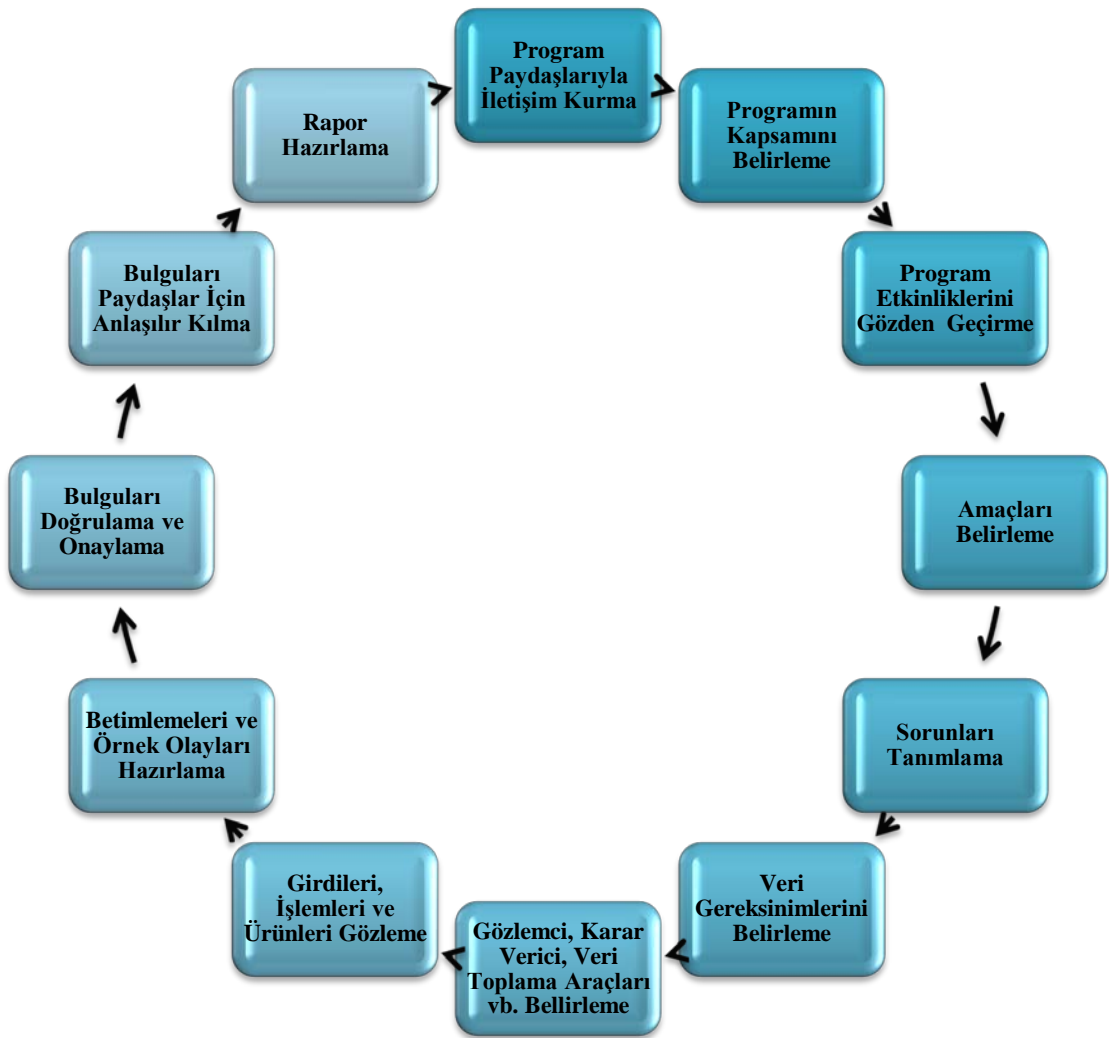
Stake’in bu değerlendirme çerçevesi dışında, geliştirilmesinde öncülüğünü yaptığı ve katılımcı program yaklaşımını temel alan diğer program değerlendirme modeli yanıtlayıcı değerlendirmedir.

Yanıtlayıcı Değerlendirme

Bireylerin bir nesneyi değerlendirdiklerinde doğal olarak gerçekleştirdikleri iki eylem vardır: Gözlemlemek ve gözlem sonucu tepkide bulunmak/karar vermek. Stake (1975) yanıtlayıcı değerlendirmeyi bu iki kavramla tanımlamaktadır: Gözlem süreci ve tepkide bulunma süreci. Çünkü birey gereksinimlerine dönük olarak gözlem yapmakta ve yine gereksinimlerine dönük olarak bir tepkide bulunmakta/karar vermektedir. Bu nedenle program da bireyin gereksinimlerini yansıtmalı ve bu anlayışla değerlendirilmelidir. Stake (1975) yanıtlayıcı bir program değerlendirmenin temel özelliklerinin değerlendirmenin program amaçlarından çok program etkinliklerine odaklanması, paydaşların gereksinimlerini dikkate alması ve raporda farklı değerlere atıfta bulunması olduğunu belirtmektedir.

Yanıtlayıcı değerlendirmenin geliştirilmesinin temel nedenlerden biri Stake’in değerlendirmenin geçerliliği ile paydaşların programı ne düzeyde anladığını ilişkilendirmek istemesidir. Bu geçerliliğin sağlanması da paydaşların programı tam olarak anlamalarına, tanımlamalarına, gereksinimlerini karşılayıp karşılamadığına bağlıdır. Bu nedenle bu değerlendirme modelinde paydaşlarla iletişim oldukça önemlidir. Bunun için de değerlendirici hedef kitle ile gereksinimleri konusunda sürekli bir etkileşim içinde olmalı, süreci yürütmek için bireylerin program hakkında gözlem yapabileceği

gözlem planları tasarlamalıdır. Değerlendirici gözlem sonuçları doğrultusunda kısa özetler, örnek olaylar, betimlemeler, çıktılarla ilgili bilgiler, grafikler vb. hazırlar ve hedef kitle için değerli olabilecek tüm verileri toplar. Programla doğrudan ilgili tüm bireyleri sürece katarak kendisinin ortaya koyduğu bulguların önemi, doğruluğu ve programla ilgililiği konusunda onlarla etkileşimde bulunur ve gerekirse raporlaştırır (Fitzpatrick, Sanders, ve Worthen, 2004; Stake, 1975; Stufflebeam ve Shinkfield, 2007). Bu süreçte yer alan ve yinelenen etkinlikleri saat görünümü bir şekilde görselleştirmiştir (Şekil 12).



Şekil 12. Yanıtlayıcı Değerlendirmede Belirgin Eylemler

Kaynak: Stake (1975)'den sadeleştirilmiştir..

Stake'in geliştirdiği bu saat figüründe değerlendiricinin tepeden (saat 12) başlaması ve saat yönünde hareket etmesi önerilmektedir. Ancak Stake, yanıtlayıcı değerlendirilmede

eylemlerdeki esnekliđi vurgulamak için saat figüründeki herhangi bir eylemin başka bir eylemi izleyebileceđini ya da sürece herhangi bir eylemden başlanabileceđini ve saat yönünde ya da tersinde eyleme devam edilebileceđini belirtmiştir. Ayrıca birçok eylemin eşzamanlı gerçekleşebileceđini ya da bir eylemin süreç boyunca birkaç kez tekrarlanabileceđini de ifade etmiştir.

Katılımcı Deđerlendirme

Katılımcı deđerlendirme karar vericiler, hedef kitle, program yürütücüleri, mali destekleyiciler vb. tüm paydaşların deđerlendirmenin geliştirilmesinde ve uygulanmasında etkin olarak görev aldığı bir deđerlendirme yaklaşımıdır. Paydaşların katılımına sürecin yalnızca bir boyutunda deđil; ilgili soruların belirlenmesi, deđerlendirme deseninin planlanması, veri toplama ve ölçme yöntemlerinin seçilmesi, verilerin toplanması ve çözümlenmesi, bulgular, sonuçlar ve öneriler üzerine görüş birliđi sağlanması, sonuçların yayınlanması, programın geliştirilmesi için bir eylem planının hazırlanması vb. tüm boyutlarda yer verilir (Zukoski ve Luluquisen, 2002).

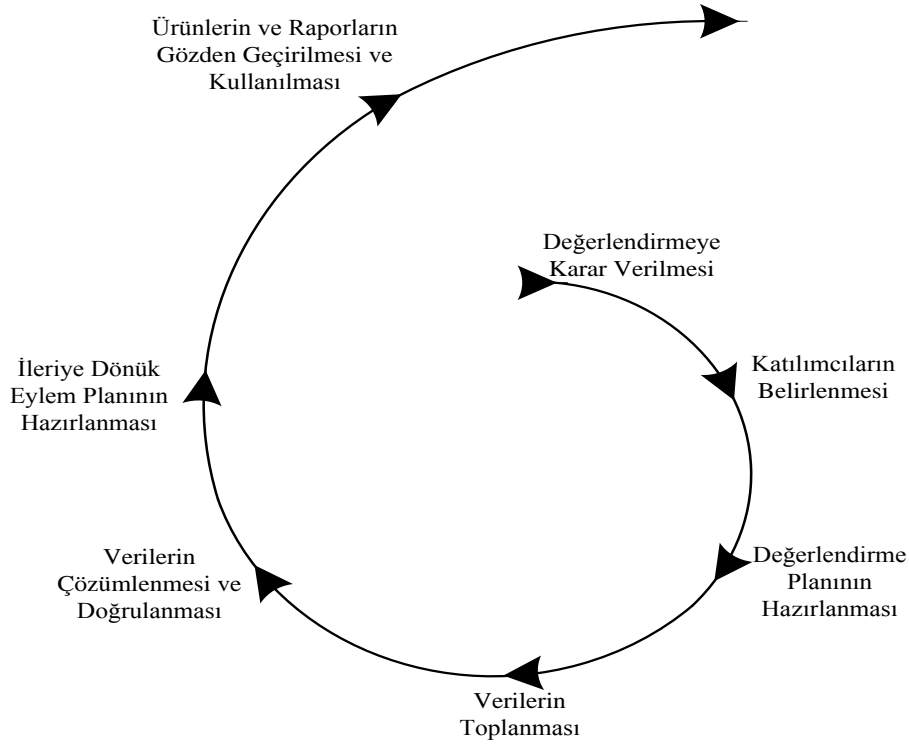
Modelin geliştirilmesine katkı sağlayan Cousins ve Whitmore (1998), katılımcı deđerlendirme modelinin üç belirgin özelliđinin olduğunu ifade etmektedir. Bunlar; deđerlendirme sürecinin kontrolü, paydaş seçimi ve katılımın derinliđidir. Cousins ve Whitmore (1998) bu üç özelliđi üç boyutlu bir şekilde görselleştirmiştir (Şekil 13).

Şekil 13. Katılımcı Değerlendirme Modelinin Temel Özellikleri

Kaynak: Cousins ve Whitmore, 1998.

Şekil 13’de de görüldüğü gibi bu özelliklerden ilkinde değerlendirmeyle ilgili kararların kontrolü değerlendiricinin (a2) ve paydaşların (a1) elindedir. Kontrol burada değerlendirmenin ne zaman ve nasıl başlatılacağı kararlarının aksine değerlendirme yürütülmesine ilişkin teknik kararlarla ilgilidir. İkinci özellik olan paydaş seçimi, değerlendirme sürecinde bireyler (b1) ile tüm yasal grupların (b2) katılımını betimler. Katılımın derinliği ise değerlendirmeye katılımın yoğunluğu ile ilgilidir. Burada değerlendirmeye katılım, kontrol ve sorumluluk içermeyen danışma (c1) amaçlı olabileceği gibi değerlendirme planlanması, verilerin toplanması, analizi, raporun oluşturulması ve sonuçların yaygınlaştırılması gibi tüm değerlendirme süreçlerine etkin katılım (c2) da olabilir.

Katılımcı değerlendirme uygulanmasında temel alınan belirli aşamalar bulunmamaktadır. Aşamalar uzmanlara göre değişiklik gösterebilmektedir. Bazı araştırmalarda (IDS, 1998; Rossman, 2008) dokuz aşamalı bir uygulama sürecinden söz edilirken Burke (1998), spiral-tasarımlı olarak geliştirdiği yedi aşamalı bir süreçten söz etmektedir (Şekil 14).



Şekil 14. Katılımcı Program Değerlendirme Sürecinin Aşamaları

Kaynak: Burke (1998)'den alınmıştır.

Şekil 14'te gösterildiği gibi, katılımcı değerlendirme sürecinin başlangıcında, değerlendirici öncelikle, değerlendirme modelinin araştırma için her yönüyle uygun olup olmadığı kararını vermelidir. Örneğin, katılımcı değerlendirme maliyeti yüksek olan bir değerlendirme modelidir. Bunun göz önünde bulundurulması gereklidir. Değerlendirmeye karar verildikten sonra değerlendirmede yer alacak bireylerin belirlenmesi gereklidir. Bu aşamada değerlendirme için iç ve dış değerlendiriciye görev verilip verilmeyeceğinin de belirlenmesi gereklidir. Değerlendirme takımı belirlendikten sonra, grup işbirliği içinde değerlendirmenin amacı, ürünleri, veri toplama ve çözümleme yöntemlerini içeren bir değerlendirme planı hazırlanır. Belirlenen plan doğrultusunda veriler toplanır ve çözümlenir. Grup, değerlendirme sonuçları ve bulguları üzerinde tartışmalar yaparak kullanılacak bilgilere son halini verir. Daha sonra bilgilerin ileriye dönük olarak program geliştirme sürecinde kullanılması için bir eylem planı hazırlanır ve tüm süreç raporlaştırılır ve paydaşlara sunulur (Burke, 1998).

Doğal Değerlendirme

Dođal deđerlendirme, katılımcıların farklı grşlerini ve deđerlerini dikkate alarak, program hakkında gereksinim duydukları bilgilendirmeye yanıt verebilmek ve bu yanıtlara gre programa iliřkin dođal genellemeler yapmayı amalayan katılımcı deđerlendirme modelidir (Alkin ve Christina, 2004). Bu alanda en kapsamlı alıřmalar Guba ve Lincoln tarafından yapılmıřtır. Guba ve Lincoln (1981) dođal deđerlendirmeyi Stake'in yanıtlayıcı deđerlendirme modeli ile iliřkilendirmiřler ve bu yaklařımın uygulanmasındaki iřlemleri tanımlamıřlardır.

Dođal deđerlendirmede yer alan katılımcılar, programın uzmanı olan bireyler deđil, daha ok programın deđerlendirilmesini gnll olarak yapan kuruluřlar ya da bireylerdir. Bu katılımcılar programın deđerlendirilmesini, teknik bilgiler iermeyen gncel hayatta karřılařılan olaylarla iliřkilendirirler, grşlerini basit bir dille aıklarlar. Bu ynyle dođal deđerlendirme formal deđerlendirmeden ok informal deđerlendirme zelliklerine sahiptir. Bařka bir deyiřle, deđerlendirici programı dođal olarak oluřtuđu asıl yerinde herhangi bir sınırlama, dzenleme yapmadan ve sreci kontrol etmeye alıřmadan gzlemeler ve deđerlendirir.

Bu deđerlendirme modeli deđerlendiriciyi bir ğrenen, katılımcıları ise deđerlendiriciye bilgi veren, ğretenler olarak ele alır. Katılımcıların bilgi verenler olarak tanımlanmasının gerekesi ise, deđerlendiriciye bakıř aılarını, kendi dnyalarını tanımlamada kullandıkları kavramları ve bu kavramları nasıl tanımladıklarını, aıkladıklarını ve bunu yařamlarına nasıl uyguladıklarını aıklamaları ve deđerlendiricinin bilgilenmesini sađlamalarıdır.

Fitzpatrick, Sanders ve Worthen (2004) deđerlendiricinin katılımcılardan bilgi toplarken dikkat etmesi gereken ařamaları ve bilgi trlerini syle sıralamaktadır:

- Deđerlendirilecek programa ve programın iinde bulunduđu ortama iliřkin bilgilerin tanımlanması
- Toplanacak bilgilerin gereksinim ve ilgilere yanıt verici olması
- Toplanacak bilgilerin sorunlara yanıt verici olması (Sorunların sınıflandırılması, bu sorunları zmeye ynelik olası faaliyetlerin tanımlanması)

- Değerler hakkındaki bilgiler (değerlerin tanımlanması, bu değerlerin kaynaklarının ve arkasındaki inançların araştırılması)
- Değerlendirmede kullanılacak standartlara ilişkin bilgiler (ölçütlerin, beklenti ve gereksinimlerin tanımlanması)

Doğal değerlendiriciler için program değerlendirme sürecinde bulguların güvenilirliği ve inandırıcılığı çok önemlidir. Bu nedenle elde edilen bulguların doğrulanmasında ve güvenilirliğinin belirlenmesinde çapraz kontrol ya da üçgenleme yöntemi kullanılır. Üçgenleme yönteminde değerlendirmenin başka alanlara ve konulara uygulanabilirliği, değerlendirme bulgularının tutarlılığı ve değerlendirmenin tarafsızlığı ayrıntılı olarak incelenmekte ve bir yargıya varılmaktadır.

Katılımcı Odaklı Program Değerlendirme Yaklaşımının Üstünlükleri ve Sınırlılıkları

Fitzpatrick, Sanders ve Worthen (2004) katılımcı odaklı program değerlendirme yaklaşımının üstünlüklerini şöyle açıklamışlardır:

- Programın değerlendirilmesi sürecinde izlenen sabit bir plan olmaması başka bir deyişle uygulanmasındaki esneklik, değerlendirme sürecinin değiştirilebilirliğini sağlamaktadır.
- Hem öznel hem nesnel hem de nitel ve nicel verilerin toplanabilmesi değerlendirilen programa ilişkin çok miktarda veri elde edilmesini sağlar.
- Değerlendirmede insan faktörüne vurgu yapılır ve tüm katılımcıların görüşleri dikkate alınır, dolayısıyla programa farklı bakış açıları getirilmiş olur.
- Değerlendirilen programdan etkilenenler ve etkileyenler değerlendirme sürecine katıldıklarında kendilerini daha rahat ifade etme, ilgilerine ve sorunlarına ilişkin görüş belirtme şansı bulmaktadırlar. Bu durumda toplumda katılım kültürünün geliştirilmesine yardımcı olmaktadır.

Katılımcı odaklı program değerlendirme yaklaşımına dayalı program değerlendirme modellerinin sınırlılıkları ise şunlardır:

- Değerlendirme sürecinde katılımcıların farklı bakış açılarına yer verilmesi, elde edilen verilerin nitel olması nedeni ile verilerin çözümlenmesinin çok zaman alıcıdır.
- Değerlendirme verilerinin daha çok nitel ve öznel olması elde edilen verilerin geçerliği ve güvenilirliğinde değerlendiricinin çok dikkat etmesini gerektirir.
- Programın değerlendirilmesinin paydaşlar ile gerçekleştirilmesi ve bu süreçte birçok katılımcının yer alması bu sürecin maliyetini arttırır.

1.4. Eğitimde Program Değerlendirme Standartlarının Tanımı, Kapsamı ve Kullanımı

Yaklaşık yüz yıllık bir tarihsel süreç içerisinde günümüze kadar gelişerek gelen eğitimde standart geliştirme çalışmaları ilk olarak 1869 yılında Amerika Birleşik Devletleri'nde fen bilimleri üzerine yapılan bir kongrede dile getirilmiştir (Edelfelt ve Raths, 1998). Bu kongrede eğitimde standart belirlemenin önemi vurgulanmış, standartlar yoluyla eğitimin niteliğinin güvence altına alınacağı ve öğretimde standartlara gidilmesi ile etkin öğretim yöntemlerinin ortaya çıkabileceği üzerine tartışılmıştır. Daha sonraki dönemlerde “Öğretim Etiği” ve “Kuramsal ve Uygulamalı Öğretim”le ilgili uygulanmak üzere birtakım ölçütler geliştirilmiştir (Edelfelt ve Raths, 1998). Geçen yüzyılın başlarında ise, bu ölçütler geliştirilerek öğretime ilişkin 40 sayfalık bir standart listesi oluşturulmuştur. Bu çalışma beraberinde belirlenen standartlara uygunluk göstermeyen program ya da kurumların kapatılması için bir süreçten geçmesi gerektiğini, yani akreditasyonu gündeme gelmiştir. Çünkü ortaya konan tartışmalarda, artık programların oldukça çeşitlilik kazandığı, nitelikli birey yetiştirmede bazı programların güçlü bazılarının ise zayıf kaldığı ve bu tür programların sistemden elenmesinin zorunlu olduğu vurgulanmıştır. Bu durum ayrıca standart geliştirmeye ne denli gereksinim duyulduğunu da ortaya koymuştur (Edelfelt ve Raths, 1998).

Bu sürece paralel olarak eğitimde standartların geliştirilmesi çalışmaları toplumsal içerikli olaylardan da doğrudan etkilenmiştir. Standart geliştirme çalışmalarını ilk olarak yürüten Amerika Birleşik Devletleri, 1930'lu yıllarda ülkede yaşanan ekonomik krizin temelinde eğitim sistemindeki nitelik eksikliği olduğunu vurgulamış, krizin nedeni olarak yetiştirilen niteliksiz bireyleri göstermiştir. Bu durum standartların daha nitelikli

hazırlanmasını, gözden geçirilmesini, eğitimde her yönden standartlaşmaya gidilmesini ve standartların ulusal çerçevede ele alınmasını zorunlu kılmıştır. Yine ülke öğrencilerinin uluslararası sınavlarda aldıkları derecelerin oldukça düşük olması ülkede bu uygulamanın önemini ortaya koymuştur (Lachat, 1994). Tüm bu çalışmalar daha sonra önemli ölçüde artmış ve günümüzde çoğu çağdaş ülke tarafından kullanılan standartlara temel oluşturmuştur.

1.4.1. Standartın Kuramsal Çerçevesi

Kullanıldığı ilk günden bu yana birçok sözcükle ifade edilen standart kavramının tüm uzmanlarca kabul gören tanımını Richardson (1994) yapmıştır. Richardson, yapmış olduğu iki farklı tanımla standart kavramının iki farklı boyutunu ortaya koymuştur. Richardson standardı, “bir otorite, gelenek ya da bir ortak anlayış tarafından belirlenmiş, izlenmesi gereken bir model ya da örnek” olarak tanımlarken, standardın bir erk tarafından belirlendiğini; standardı “belli bir amaca ulaşmada uygun ve yeterli nitelik düzeyi” olarak tanımlarken de, standardın amaca ulaşmada sahip olunması gereken düzey boyutunu ortaya koymuştur. Amerika Birleşik Devletleri’nde program değerlendirme standartlarını geliştiren The Joint Committee (Ortak Komite) ise, standardı “mesleki bir uygulamanın doğruluk düzeyinin artırılması amacıyla bu mesleki uygulamayla ilgilenen bireylerin karşılıklı olarak üzerinde anlaşmaya vardıkları ilke” olarak tanımlamaktadır (Keith, 2003). Tarihsel süreç içerisinde standart kavramıyla aynı anlamda “ölçüt”, “mikyas”, “ayar”, “mihenk” ve “test” sözcüklerini kullanmıştır. Ancak Glass (1978), bu sözcüklerden özellikle “ölçüt” sözcüğünün standart ile aynı anlamı taşımadığını ve çoğu eğitimcinin standart sözcüğünü yanlış kullandığını savunmaktadır. Glass’a göre ölçüt “karar vermede bir ilgi değişkenidir, standart ise bu değişkenin miktarını gösterir.” Örneğin, Türkiye’de polis olmanın ölçütü uzun boylu olmaktır. Ancak standardı 1.70 cm boya sahip olmaktır. Tüm bu tartışmalarla birlikte alanyazında ölçüt ile standart eş değer sözcükler olarak halen kullanılmaktadır.

Conley (1997), eğitimcilerin standart kavramını, öğrencilerin program sonunda ne yapmaları gerektiğini belirlerken kullandıklarını ifade etmektedir. Conley’e göre standartlar, öğrencilerin ve öğretmenlerin eğitim süreci içinde izlemesi gereken en değerli davranış, tutum ve becerilerdir. Standartlar genellikle belirgin bileşenlere sahiptir. Bu bileşenler öğretmenin, öğrencinin ve velinin kendilerinden beklenenleri bilebilmeleri için

ayrıntılı ve açıktır. Standartların yeteri düzeyde açık olmaması, öğretmenin ne öğreteceğini, öğrencinin ne öğreneceğini ve neyin, nasıl değerlendirileceğinin de bilinmemesine neden olur. Bu nedenle açık ve anlaşılır özellikteki çoğu standart Conley'in (1997) tanımladığı aşağıdaki bileşenlere sahiptir.

- *Standart Alanı:* Standart ifadesinin ait olduğu, üzerinde uzmanlaşma beklenen çalışma alanını temsil eder.
- *Ayrıntılı Tanım:* Standart alanının ayrıntılı tanımlarını ifade eder.
- *Yeterlikler:* Öğrencilerin program sonunda standart alanıyla ilgili sahip olmaları gereken bilgi ve becerileri ifade ederler. Yeterlik, programın ayrıntılı bir tanımı ya da öğrencilerin program sonunda bilmeleri gereken her şeyin bir listesi değil; birbirini izleyen bir dizi beceriyi tamamlayan anahtar beceridir.
- *Göstergeler:* Yeterlikleri daha ayrıntılı olarak bir ya da birkaç ifadeyle tanımlayan, program ve değerlendirme arasında bir köprü görevi gören ve davranışa dönüştürülebilen ifadelerdir.
- *Düzye:* Düzye, standart için bir derecelendirme ölçeğidir. Davranış, beceri ya da bilginin ne düzeyde olması gerektiği bu ölçekle ifade edilir. Bu ölçek, sayılar (1–5) biçiminde olabileceği gibi, görev tanımlarıyla (deneyimsiz, yeterli ve uzman) ve harflerle (A-F) de gösterilebilir.
- *Bilgi Yoğunluğu:* Bu bileşen, değerlendirilsin ya da değerlendirilmesin, öğrencinin öğrenmesi gereken bilginin toplam miktarını belirtir.
- *Değerlendirme Yöntemi:* Yeterliklerin göstergelerinin süreç sonunda kazanılıp kazanılmadığının bilinmesi için belirli ölçme yöntemlerine gereksinim vardır. Her bir yeterlik ifadesi için özel değerlendirme yöntemi olabileceği gibi, yeterliklerin tümünün değerlendirilmesinde yalnızca bir değerlendirme yöntemi de kullanılabilir. Bazı durumlarda düzey, değerlendirme yönteminin yerine kullanılabilir.

1.4.2. Standart Geliştirme Aşamaları

Standartlar yukarıda incelenen bileşenleri kapsamak koşulu ile eğitimin tüm basamaklarında ve alanlarında ilgili uzman görüşleri doğrultusunda birkaç oturumluk süreçler halinde geliştirilir. Ancak bir standardın tam olarak standart özelliği kazanabilmesi bazen yıllarca süren çalışmaları da gerektirebilir. Genel olarak standart geliştirme aşamaları ülkeden ülkeye, hatta aynı ülke içinde kurumdan kuruma farklılıklar gösterebilmektedir. Bu konuda Amerika Ulusal Standartlar Enstitüsü için standart geliştiren bir kurum olan Ulusal Bilgi Standartları Organizasyonu (National Information Standards Organization -NISO, 2009) ve Uluslararası Standartlaştırma Organizasyonu (International Organization for Standardization -ISO) beş, Plake (2008) altı aşamalı bir standart geliştirme sürecinden söz ederken, Kanada Standartlar Birliği (The Canadian Standards Association -CSA, 2009) yedi ve Cheng (2003) oniki aşamadan söz etmektedir. Bu kaynaklarda geçen aşamalar özetlenerek şu aşamalar altında sınıflandırılmıştır (Şekil 15).

- *Başlangıç Aşaması:* Programa girişte bireylerin izlemesi gereken yönergeler niteliğinde olan standartlar program süreci içinde ve sonunda bireylerin gereksinimlerini karşılamalıdır. Bu nedenle standartların geliştirilmesi süreci, program geliştirme sürecinde olduğu gibi gereksinimlerin belirlenmesiyle başlamaktadır. Belirlenen bu gereksinimlerin güncel gereksinimlerle de bağlantısı kurularak ortaya standartların temelini oluşturacak ifadeler çıkarılır.



Şekil 15. Standart Geliştirme Süreci

Kaynak: NISO (2009), Plake (2008), CSA (2009) ve Cheng (2003) den sadeleştirilmiştir.

- **Öneri Aşaması:** Standartlara olan talebin belirtilmesi ve gereksinimlerin belirlenmesinden sonra yapılan tartışmalar sonucu standartların kimler tarafından geliştirileceği açıklığa kavuşturulur. Bu oturumlar sonucu alan uzmanları, standartları talep eden kurum ve kuruluşlardan temsilciler ve ilgili bireylerden oluşan standart geliştirme kurulu belirlenir. Kurulun belirlenmesiyle birlikte ilk olarak standart geliştirme sürecinin temel amaçları belirlenir.
- **Geliştirme Aşaması:** Standart geliştirme kurulu gereksinimler ve amaçlar doğrultusunda belirli aralıklarla bir araya gelip standart olabilecek ifadeler üzerinde tartışılır. Tartışmalarla ortaya çıkan sorunlar ve öneriler doğrultusunda kurul bilgi

toplar ve bu bilgileri etkin olarak kullanır. Çünkü standartların kuramsal hiçbir eksikliğinin olmaması ve bireylerde karar yanılıklarına neden olmaması gerekmektedir. Bu nedenle standart oluşturma süreci sıkıcı ve uzun süreli olabilmektedir. Bu süreç sonunda kurul tarafından ortaya bir taslak standartlar listesi çıkar.

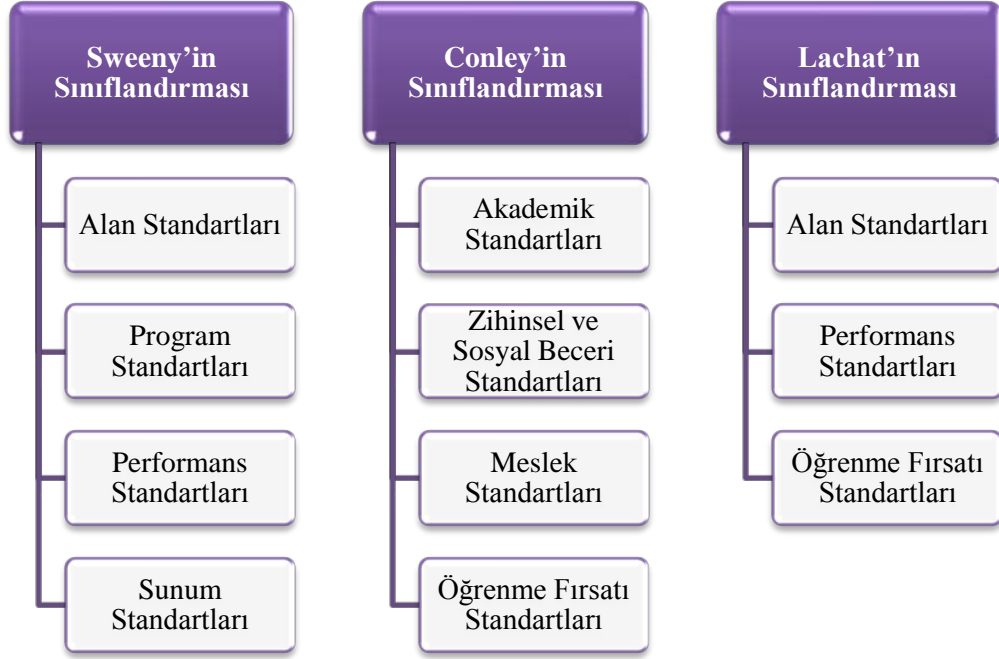
- *İnceleme Aşaması:* Standart geliştirme kurulu tarafından geliştirilen taslak standart ifadeleri, standartların uygulanacağı tüm kurum, kuruluş ve bireylerin görüşlerine sunulur. Bu süreç sonunda gelen tüm öneri ve eleştiriler kurul tarafından dikkatle incelenir, tartışılır. Bu tartışmalardan çıkan sonuçlarla birlikte her bir standart üzerinde görüş birliği sağlanır ve standartlar onaya sunulur.
- *Onay Aşaması:* Bu aşamada standartların içerikleri kurul tarafından son kez gözden geçirilir. Yazım ve yönergeleri dikkatli bir biçimde incelenir ve daha sonra standart geliştirme süreci baştan sona kadar değerlendirilerek varsa sorunlar çözülür ve eksiklikleri giderilir.
- *Yayınlama Aşaması:* Kurul tarafında her yönüyle gözden geçirilen standartlar yayınlanır ve katılımcı ve ilgililerin gereksinimleri doğrultusunda tüm baskı türlerinde çoğaltılarak yaygınlaştırılır.
- *Sürdürme Aşaması:* Yayınlanan standartların kalıcılığı sağlanır ve tıpkı bir program geliştirme sürecinde olduğu gibi ilgili bireylerin talepleri doğrultusunda günün koşullarına uygun olarak yeniden geliştirilir.

1.4.3. Standart Türleri

Standartlara ilişkin yapılan tüm tanımlamalar, standardı oluşturan bileşenlerdeki çeşitlilik ve standartların oluşturma süreçlerinin farklı amaçlarla bir araya gelmesi beraberinde farklı türde standartların ortaya çıkmasına neden olmuştur. Alanyazında farklı uzmanlar tarafından sınıflandırması yapılan ve en çok karşılaşılan standart türleri aşağıda verilmiştir (Tablo 6).

Tablo 6.

Farklı Uzmanlara Göre Standartların Sınıflandırılması



Kaynak: Sweeny (1999), Conley (1997) ve Lachat (1994) den özetlenmiştir.

Tablo 6'da yapılan sınıflandırmalar ayrıntılı olarak incelendiğinde, Sweeny (1999) ve Lachat (1994) tarafından yapılan sınıflandırmaların bir standart türü dışında birbirine benzediği, Conley'in sınıflandırmasının ise içerik olarak incelendiğinde her iki sınıflandırmayı da kapsadığı görülmektedir. Conley (1997) yaptığı sınıflandırmada diğer sınıflandırmalarda yer verilen "*alan standartları*"nı ve "*performans standartları*"nı "*akademik standartlar*" adı altında incelemiş, program standartlarını ise farklı bir standart türü olarak onaylamasına karşın, sınıflandırmasına almamıştır. Bu çalışmada Conley'in yaptığı sınıflama temel alınmış ve ilgili standart türleri aşağıda açıklanmıştır.

- *Akademik Standartlar*: Eğitimde en sık karşılaşılan ve kullanılan standart türü akademik standartlardır. Bu standartlar genel olarak bireylerin sayısal (matematik, fizik kimya vb.), sosyal (tarih, coğrafya, psikoloji vb.) dil ve görsel sanatlar alanlarında sahip oldukları akademik bilgi ve beceriler üzerine yoğunlaşır. Her alana ilişkin standartlar birbirinden farklı olmakla birlikte yapı olarak tüm

alanlardaki standartlarda iki boyut vurgulanmaktadır: Alan bilgisi ve bilişsel beceriler.

Akademik standartların alan bilgisi boyutu öğrencilerin alanla ilgili ne bilmeleri ve ne yapmaları gerektiği üzerinde durur. Bu standartlar, üst düzeyde bir yeterliğe sahip olabilmek için öğrencinin öğrenmesi gerekli bilgi, beceri ve tutumları içerir. Kısaca bireyin alanla ilgili kuramsal yeterliğini ifade etmektedir. Bilişsel beceriler ise analiz, sentez ve değerlendirme gibi üst düzey bilişsel becerileri içermektedir. Bu nedenle alan bilgisi ile bilişsel beceriler bir bütündür (Conley, 1997).

- *Zihinsel ve Sosyal Beceri Standartları:* Akademik standartlardan farklı, ancak bu standartları tamamlayan, geliştirilmesi ve belirlenmesi oldukça güç olan standartlardır (Conley, 1997). Bu standartlar genellikle bir alanla ilgili içerikten çok bu içeriğin daha iyi öğretimi, öğrenci tarafından daha iyi öğrenilmesi için gerekli temel becerileri içermektedir. Bu yönüyle akademik standartların tamamlayıcısı olan standartlardır. Örneğin, sorun çözme becerisine ilişkin standartlar belli bir akademik alanla ilişkilendirilmediği sürece bir anlam taşımaz. Okuma, yazma, iletişim yeterliği, analitik düşünme, sorun çözme, takım çalışması vb. beceriler zihinsel ve sosyal beceri standart alanlarına ilişkin örneklerdir.
- *Meslek Standartları:* Bu standartlar, bireylerin mesleği seçmeden önce bir meslek için sahip olması gerekli temel yeterlikleri ve mesleği seçip göreve başladıktan sonra mesleği yürütmek için gerekli temel bilgi ve becerileri içeren standartlardır. Günümüzde işverenler yalnızca alanında uzman ve işini iyi bilen bireyleri değil, bu özelliklerinin yanı sıra kendini yazılı ve sözlü ifade edebilen, akıcı ve net konuşabilen, yönergeleri uygun bir şekilde izleyebilen, hangi alandan olursa olsun temel matematik becerilerini kusursuz yapan, yaratıcı ve eleştirel düşünebilen, sorun çözen, öğrenme sorumluluğunu üstlenen bireylere de gereksinim duymaktadır. Bu nedenle mesleğe girmeden önce bireylere bu becerilerin, daha sonra da meslekle ilgili ayrıntılı bilgi ve becerilerin kazandırılması gereklidir. Bu yönüyle bakıldığında meslek standartları çok boyutlu olabilmektedir. Örneğin, akademik standartlar ve/veya zihinsel ve sosyal beceri standartları meslek standartlarının bir boyutunu oluşturabilmektedir. Öğretmenlik mesleği standartları

dikkate alındığında akademik standartlar belirli bir alanı, öğretmenlik mesleğine ilişkin bilgi ve beceriler belirli bir alanı oluşturmaktadır (Conley, 1997).

- *Öğrenme Fırsatı Standartları:* Öğrenme fırsatı standartları, her öğrenciye alan ve performans standartlarını gerçekleştirebilmeleri için nitelikli öğretim elemanı, nitelikli dersler, program, araç-gereçler, öğretim zamanı, yardım sağlanması vb. boyutlar üzerinde durur. Bu yönüyle bu standartlar yalnızca öğrencinin öğrenmesi ile ilişkili değil, öğrenci öğrenmesini çevreleyen durumlarla da ilişkilidir (Lachat, 1994).

Conley (1997) daha önce açıklandığı gibi, program standartlarını ayrı bir tür olarak değerlendirmiş ancak standart türlerini incelediği çalışmasında yer vermemiştir. Daha çok akademik standartlar altında alana dönük standartlar olarak tanımlanan program standartlarına, bu bağlamda program değerlendirme standartları da eklenebilir. Belirli bir alana özgü olmayan ancak programla iç içe olan program değerlendirme standartlarına ilişkin bilgiler aşağıda verilmiştir.

1.5. Program Değerlendirme Standartlarının Kuramsal Çerçevesi

Program değerlendirme standartları, program değerlendirme çalışması yapacak araştırmacılara yol göstermede, yapılan bir program değerlendirme çalışmasını değerlendirmede ya da değerlendirmenin sonuçlarını kullanacak ve buna dayalı karar verecek yetkililere araştırmanın geçerliği ve güvenilirliği hakkında sağlam bilgi sunmada kullanılacak ölçütleri kapsar (Sağlam ve Yüksel, 2007). Bu ölçütler, bir değerlendirmenin sahip olması gereken yararlık, yürütülebilirlik, uygunluk ve doğruluk özelliklerini yansıtmaktadır (JC, 1994).

- *Yararlık:* Yararlık, değerlendirmenin hem yeterliliğinin hem de kazancının açık göstergesidir. En basit yapısıyla bir değerlendirmenin yararlılığı, değerlendirmenin gereksinimlere ne kadar yanıt verdiği, amaçlara ne kadar hizmet ettiği ile ilgilidir (JC, 1994; Beywl ve Speer, 2004). Alanda yeni olanlar için, bir program değerlendirme çalışması yürütmenin temel amacı programın geliştirilmesi ya da programla ilgili bir yargıya varılması için gerekli bilginin sağlanmasıdır. Genel olarak bu amacı taşıyan değerlendirmeler, program değerlendirmeyi mali olarak

destekleyenlerin ve değerlendirme sonuçlarını kullananların gereksinimleri üzerine odaklanır. Böyle bir durumda değerlendirme, programla ilgili kaynakların ve taleplerin doğru olup olmadığının belirlenmesinde bir mekanizma olarak kullanılmaktadır. Bu tür bir değerlendirme programının niteliğiyle ilgili doğru yargıları sağladığı sürece yararlıdır.

- *Yürütülebilirlik/yapılabilirlik*: Yürütülebilirlik/yapılabilirlik, herhangi bir girişimin yönetim ve ekonomi yönlerinden durumunun önceden belirlenmesine karşılık gelen bir ifadedir (Kabukçuoğlu, 2005). Program değerlendirme alanında ise, yürütülebilirlik bir programın değerlendirilebilirliğinin saptanması, programın içinde bulunduğu sosyal, politik ve ekonomik ortamın tanımlanması, programın uygulandığı okul, hapishane, sağlık merkezi vb. fiziki ortamın betimlenmesi ve son olarak programın politikacılar, paydaşlar, değerlendiriciler için ne önem arz ettiğinin betimlenmesi ve değerlendirme kaynaklarının etkili ve verimli kullanılması gibi eylemleri içeren bir ifadedir (JC, 1994; Beywl ve Speer, 2004). Yürütülebilirlik genel olarak değerlendirme çalışmasına üç yolla katkıda bulunur:
 - Olası değerlendirme süreçlerini, gerçek uygulama süreçlerine dönüştürür. Değerlendirici öncelikli olarak tasarladığı sürecin programın bulunduğu ortamda ya da durumda kullanılabilir olup olmadığına, tasarladığı olası sürecin yapılabilir olup olmadığına karar verir. Buna karar verirken yürütülebilirlik/yapılabilirliğin “bir devam et” ya da “bırak” süreci değil; değerlendirme boyunca sürekliliği olan esnek ve yinelenen bir süreç olduğuna dikkat eder. Çünkü programdaki olası değişikliklerden dolayı değerlendirme sürecinin başında yürütülebilir/yapılabilir olan daha sonra olmayabilir.
 - Yürütülebilirlik/yapılabilirlik değerlendirilmenin yararlık, doğruluk ve uygunluk gibi temel özelliklerine de temel oluşturur. Başka bir deyişle yürütülebilirlik/yapılabilirlik yokluğunda, doğru, yararlı ve uygun bir değerlendirmenin olması olası değildir.
 - Yürütülebilirlik/yapılabilirlik değerlendirme kaynaklarının etkili ve verimli kullanılmasını gerektirir.

- *Uygunluk:* Değerlendirmede uygunluk, bir değerlendirme uygulamasında neyin uygun, dürüst, yasal, doğru, kabul edilebilir ve mantıklı olduğu ile ilgilidir (JC, 1994). Dolayısıyla bir değerlendirme çalışmasının uygunluğu birbirini kapsayan üç temel temayla açıklanabilmektedir: Etik, yasal ve mesleki yönden uygunluk. Her bir tema değerlendiricinin, katılımcıların hak ve sorumlulukları ile değerlendirme sürecinde ortaya çıkan sistem kuralları, rolleri ve görevleri ile doğrudan ilgilidir.
 - Etik yönden uygunluk, değerlendirme çalışmasında değerlendirici ve katılımcıların davranışları, sorumlulukları ve hakları ile ilgili konuları içerir. Değerlendirmeye katılan katılımcıların temel hak ve özgürlüklerinin değerlendiriciler tarafından dikkate alınması ve saygı gösterilmesi sürecin etik yönden uygunluğu açısından önemlidir. Bireylerin dil, din, ırk ve cinsiyet yönünden temel haklarına saygı gösterilmesi temel bir zorunluluktur. Bu saygı değerlendirmenin yalnızca bir bölümünde değil sürecin tümünde gösterilmelidir.
 - Yasal yönden uygunluk, değerlendirmede tüm süreçlerin devletin hali hazırda uygulanan yasalarıyla uyumlu olmasını gerektirir. Bu nedenle değerlendiriciler değerlendirme çalışmalarının yasalarla uyumlu bir şekilde yine yasalarla uyumlu eğitsel ortamlarda yürütmelidirler. Yasal yönden uygunluğun bir diğer boyutu da değerlendirme çalışmaları için gerekli uygulama izinlerinin alınmasıdır.
 - Mesleki yönden uygunluk, değerlendirme uygulamasında değerlendiricinin görev ve sorumlulukları ile ilgilidir. Değerlendiricinin süreç içinde bu görev ve sorumlulukları katılımcıların hak ve özgürlüklerini korumaları ve değerlendirmenin mali süreçlerini dikkatle izlemeleridir.
- *Doğruluk:* Doğruluk bir programın ya da bileşenlerinin niteliği hakkındaki yargıları destekleyen değerlendirme bulgularının, önerilerinin, süreçlerinin ve planlarının güvenilirliğini belirtmektedir (JC,1994). Bir program değerlendirme çalışmasının doğruluğu için; değerlendirme bulgularının, yorumlarının, sonuçlarının, önerilerinin ve sonuçta varılan kararlarının güvenilirliğinin, değerlendirme desen ve yapılarının geçerliğinin (kapsam ve yapı geçerliği vb.), başvuru bilgilerin güvenilirliğinin, değerlendirilecek programlarla ilgili kaynakların doğruluğunun,

programın uygulandıđı çevrenin iyi çözümlenmesinin ve raporun dış etkilerden bağımsız olmasının sağlanması gereklidir. Öte yandan değerlendirmenin doğruluđunu sınırlayan etmenlerin de giderilmesi önemlidir. Bu engeller genel olarak tutarsızlık, çarpıklık ve kavram yanlışlıklarıdır.

- *Tutarsızlık*, değerlendirme çalışmasının tutarsız olması, değerlendirmenin güvenilir sonuçlar üretememesi, sonuçlar ile verilen kararların birbiriyle çelişmesidir. Bunun önlenmesi için öncelikle güvenilir kaynaklardan güvenilir bilgiler edinilmesi, değerlendirme süreçlerinde kullanılan tüm yapıların geçerliklerinin ve güvenilirliklerinin sağlanması gereklidir.
- *Çarpıklık*, değerlendirmenin tutarsız olmasının da temel etkenlerinden biri olan bilgilerin güvenilirliğiyle doğrudan ilgidir. Bilgi çarpıklığı değerlendirme sonuçları ve kararlarının da çarpık olmasına yol açar. Değerlendirme sonuçlarının geçerli ve güvenilir olması ve bu sonuçlara dönük doğru kararlar alınması için güvenilir ve sağlam bilgilerin toplanması ve bu bilgilerin yine uygun yöntem ve yollarla işlenmesi, yanlış düzenlenmesinin ve kodlanmasının önüne geçilmesi, analizlerinin doğru yapılması, yanlış değerlendirme deseni kullanımının engellenmesi, eksik raporlaştırmaya izin verilmemesi gereklidir.
- *Kavram yanlışlığı*, değerlendirme sürecinde doğruluđu sınırlayan diđer bir etmendir. Genelde kavram yanlışlığı iki farklı kavramsallaştırma sorunundan kaynaklanmaktadır. Bunlardan ilki bir kavramın mantıksız olması, ilgisiz durması, kullanıldıđı kuramsal çerçeve içinde geçersiz olması ve verilerle desteklenmemesi; diđeri ise ilkinden daha önemli ve giderilmesi zor olan, bireyin “kavramla ilgili farklı bilgilendirilmiş” olmasıdır. Diđer bir deyişle, bireyin kavramı olması gerekenden farklı bir anlamda kullanmasıdır. Çünkü bireyin geldiđi ortam, sahip olduđu değerleri, ön bilgileri, algıları farklılaşabilmektedir. Değerlendirme sürecinde bu durum doğal olarak sonuçlara ve dolayısıyla da kararlara yansıdıđı için, paydaşlar ve değerlendirme kullanıcıları tarafından farklı yorumlanabilmektedir. Genellikle bu sorun, süreç içinde değerlendiricinin paydaşların, değerlendirme kullanıcılarının sıkı bir iletişim içinde olmaları ile çözülebilmektedir.

Program değerlendirme standartları bir değerlendirme çalışmasının sahip olması gereken bu dört temel özelliği yansıtacak şekilde geliştirilmiştir.

1.5.1. Program Değerlendirme Standartlarının Amacı ve İşlevi

Program değerlendirme standartlarının temel amacı eğitim programlarının, projelerinin ve ürünlerinin doğru, yararlı, yürütülebilir, uygulanabilir, etik ve sağlam olmasını sağlamaktır. Değerlendirme standartları aracılığı ile değerlendirme için işe yarar bir değerlendirme felsefesinin ortaya konması amaçlanmıştır. Öte yandan, standartlar ayrıca program değerlendirme çalışmalarının yönetilmesi ve yönlendirilmesini de amaçlar. Bu yönüyle hem değerlendiricilere hem de değerlendirmeyi kullanan kesime yardımcı olur.

Sağlam bir program değerlendirme uygulaması için geliştirilen bu standartlar değerlendirmenin, yararlık, uygunluk, yürütülebilirlik ve doğruluk gibi standartları içeren temel özellikleri dikkate alınarak düzenlenmiştir (Keith, 2003). Düzenlenen bu standartlar genel olarak öğretmenler, yöneticiler, okul müdürleri, değerlendirme eğiticileri, program geliştirme uzmanları, karar vericiler, personel yöneticileri, danışmanlar, toplum liderleri, iş ve eğitim örgütleri vb. tarafından kullanılmak üzere alana kazandırılmıştır. Ayrıca bir değerlendirme çalışmasını yürüten ya da yönetenler ile değerlendirme sonuçlarını okullarda, üniversitelerde vb. alanlarda yürütülen eğitimin niteliğini arttırmada kullanmak isteyen bireyler için değerlendirme standartları yararlı ve kullanışlı ölçütler dizisidir (Stufflebeam, 1992).

Stufflebeam (1992) program değerlendirme standartlarını kayda değer ilkeler, yönergeler olarak ifade ederken, standartların;

- program değerlendirmenin işlevsel bir tanımını yaptığını,
- eğitsel değerlendirmenin anlamlı ve uygun yöntemleri ilgili tarafların yaptığı bir sözleşme, bir kanıt sağladığını,
- çeşitli değerlendirme sorunları ile baş etmede genel ilkeler sunduğunu,
- değerlendirmenin planlanmasında pratik bir yönerge sağladığını,
- değerlendirme planları ve raporları hakkında bir yargıya varmada herkesçe kabul edilen ölçütlerden oluştuğunu,

- değerlendirme yapacaklar için bir kuramsal çerçeve niteliği taşıdığı,
- değerlendirmenin meslekleştirilmesinde kullanılabilecek bir ilerleme kanıtı olduğunu,
- değerlendirme öğretiminde için bir içerik olarak kullanılabileceğini belirtmektedir.

Patton (1994) standartların bilmenin değerlendirmeyi bilme ile eşdeğer olduğunu vurgulayarak “çağdaş bir değerlendirme öğrencisinin standartları çalışmadan yetiştirilmesinin düşünülemediğini” ifade etmiştir. Fournier (1994) ise, değerlendirme önerilerini, planlarını gözden geçirenler, değerlendirme çalışmalarını yürütenler, yönetenler ve tamamlanmış bir değerlendirme çalışmasının değeri ve kazancı hakkında bir yargıya varacaklar için program değerlendirme standartlarının sahip olmaları gerekli “başucu kaynağı” olarak tanımlamaktadır. Fournier (1994) bu ifadelerini biraz daha genişleterek görüşlerini şöyle açıklamaktadır:

- Deneyimli araştırmacılar için standartlar, başarılı değerlendirme uygulamasının yönetimi için bir değerler ve ilkeler dizisidir.
- Bir değerlendirme çalışmasında yer alan veya bir değerlendirme çalışmasını kullanan, değerlendirme alanında yeni olanlar için değerlendirme planı ile ilgili değerlendirme sorularının ve amaçlarının oluşturulmasında başvurulacak yararlı bir çerçevedir.
- Alana temkinli girmiş lisansüstü öğrenciler için alanla ilgili yapılacaklara ilişkin eşsiz bir “nasıl” kaynağıdır.
- Sorumlu bir değerlendirici olabilme yönergisidir.
- Değerlendirme sonuçlarından etkilenen ve değerlendirme sonuçlarını kullanan tüketiciler için değerlendirmeden ne beklenmesi gerektiğine ilişkin bilgilerin bulunduğu temel bir kaynaktır.

Fournier (1994) program değerlendirme standartlarının değerlendirme alanı (eğitim, sağlık, hukuk vb.), değerlendirme ortamları (işletme, endüstri vb.), değerlendirme rolleri (iç-dış) gibi geniş bir çeşitliliğe ve farklılığa sahip olduğunu belirterek standartların çok işlevli özelliğine dikkat çekmektedir. Öte yandan program değerlendirmede temel olarak

standartların biçimlendirici ve değerlendirici olmak üzere iki temel işlevinden söz edilebilir.

- *Program Değerlendirme Standartlarının Biçimlendirici İşlevi:* Program değerlendirme standartları, bir program değerlendirme çalışması yapılmadan önce değerlendirme sürecinin başlangıcını ve başarısını etkileyen değerlendirme amaçlarının belirlenmesi, değerlendirme sınırlıklarının ve çerçevesinin belirlenmesi, değerlendirme sorularının ve ölçütlerinin belirlenmesi, değerlendirme uygulamasının planlanması, desenin (deneysel-betimsel), evren-örneklem belirlenmesi, veri toplama yöntemlerinin tanımlanması, verilerin seçimi, hangi ölçme araçlarının kaç kez, ne zaman, nasıl, hangi amaçla kullanılacağı belirlenmesi, bu araçların geçerlik ve güvenilirlik çalışmalarının yapılması gibi boyutlarda değerlendiricinin neleri nasıl yapması gerektiğine ilişkin yönlendirmede bulunabilir. Bu durum sürecin başlangıcında olabileceği gibi süreç içinde de gerçekleşebileceğinden standartlar biçimlendirici ve yol gösterici bir işleve sahiptir. Kısaca standartlar değerlendirme çalışması yapacak değerlendiricinin yol haritasıdır.
- *Program Değerlendirme Standartlarının Değerlendirici İşlevi:* Standartların değerlendirici işlevi ise değerlendiricinin tamamlamış olduğu program değerlendirme çalışmasını yararlık, yürütülebilirlik, uygunluk ve doğruluk yönlerinden değerlendirmektir. Diğer bir ifadeyle standartlar, biten bir program değerlendirme çalışmasının amaçları, desenleri, uygulamaları ve sonuçları bakımından değeri ve kazancı hakkında bir yargıya varmada kullanılmaktadır.

1.5.2. Program Değerlendirme Standartlarının Gelişimi

Amerika Birleşik Devletleri'nde program değerlendirmenin tarihine ilişkin birçok araştırma yapılmıştır (Rossi, Lipsey, ve Freeman, 2004; Fitzpatrick, Sanders, ve Worthen, 2004; King, 2003). Bu araştırmalar incelendiğinde, program değerlendirmenin yirminci yüzyılın ilk yarısında yalnızca programların ve program ürünlerinin tanımlanmasında kullanıldığını görülmektedir. Yüzyılın ikinci yarısında ise federal, merkezi ve yerel düzeyde eğitim programlarındaki sayısal artış, ölçme ve değerlendirme sistemlerindeki gelişmeler, birey ve toplum gelişimine ilişkin farklı düşünceler, resmi harcamaların

denetlenmesi talebi vb. birçok etken programların daha nitelikli yöntemlerle değerlendirilmesini ve değerlendirmenin de tüm kesimler tarafından benimsenen birtakım ölçütler doğrultusunda yürütülmesini zorunlu kılmıştır. Bu zorunluluk programların değerlendirilmesinde standartların geliştirilmesini de beraberinde getirmiştir (Yarbrough, Shulha ve Caruthers, 2004).

Program değerlendirme standartlarına ilişkin ilk çalışmalar 1960'lı yıllarda değerlendiriciler tarafından yayınlanmayan ve bireysel olarak farklı değerlendirme standartları içeren kontrol listelerinin oluşturulması ile başlamıştır. 1975 yılında ise Amerika Birleşik Devletlerinde, Amerikan Eğitsel Araştırmalar Derneği (American Educational Research Association -AERA), Amerikan Psikoloji Derneği (American Psychological Association -APA) ve 1975 yılında "Eğitsel ve Psikolojik Test ve Yönerge Standartları"nı geliştiren Eğitimde Değerlendirme Ulusal Kurulu (National Council on Measurement in Education -NCME) gibi birimlerden temsilcilerin yer aldığı Ortak Komite (The Joint Committee) kurulmuştur (Gould, 1995). Bu komitenin temel amacı eğitimde programların, projelerin ve ürünlerin değerlendirilmesine yönelik kullanılacak standartları geliştirmektir. Temelde, "Eğitsel ve Psikolojik Test ve Yönerge Standartları" içerisinde değerlendirmeye ilişkin bir bölüm bulunmaktaydı, ancak komite uzmanları bu standartların amaca uygun olmadığını ve yeni ve ayrı standartların bu üç kurumun yansıra geniş katılımlı bir çevre içinde tartışılarak geliştirilmesi gerektiğini önerdiler.

Yeni kurulan komite 1975 yılında çalışmalarına başlamıştır. Kurul, derinlemesine yaptığı çalışmalarda eğitsel değerlendirmelerde kullanılacak uygun standartların bulunmadığını belirlemiş ve geliştirilecek yeni standartların eğitsel değerlendirmede yaşamsal bir öneme sahip olacağını ifade etmiştir. Ortak Komitenin bu çabaları sonucu 1981 yılında "Eğitsel Programların, Projelerin ve Gereçlerin Değerlendirmesi Standartları" yayınlanmıştır (JC, 1994). Kurul standartları geliştirmesine karşın, standartların sürekli olarak gözden geçirilmesi, alanda denenmesi, eğitimde diğer değerlendirme türlerine (öğrenci-öğretici) göre geliştirilmesinin de önemli olduğunu vurgulamıştır. Bu nedenle komite, farklı kurumlardan gelen azami on sekiz üye ve bir başkandan oluşan bağımsız bir kurum haline gelmiştir.

1980'lerde program değerlendirme uygulamalarındaki nicel artış, program değerlendirici olmak isteyen öğrencilerin ve öğretmenlerin eğitimi, değerlendirmeye ilişkin alanyazının genişlemesi, nitel değerlendirme, meta-değerlendirme, içerik, ölçüt ve yapı geçerlik kavramlarının sık vurgulanması vb. durumların etkisiyle 1989 yılında kurul, 1981 yılında geliştirilen standartların gözden geçirilmesine karar vermiştir. Yine eski baskının yirmi binlerin üzerinde bir satış yakalaması da eski standartların gözden geçirilmesinde etkili olmuştur (JC, 1994).

Standartların gözden geçirilme çalışmaları 1990 yılında başlamıştır. Çalışmalar ilk olarak standartların nasıl bir içerikle ve düzende yer alması gerektiği üzerine yoğunlaşmıştır. Bu ilk çalışmanın sonucu olarak bir taslak standart dizisi oluşturuldu ve farklı kurum, kuruluş ve ülkelerden değerlendiricilerin görüşlerine sunulmuştur. Değerlendiricilerden genel olarak standardın tanımı ve kullanımı ile ilgili öneri ile standardın kullanımına dönük örnek bir olay yazmaları istenmiştir. Gelen öneriler doğrultusunda otuz standart belirlenmiş ve bu standartlar sağlam ve doğru bir program değerlendirme için gerekli ve yeterli olan dört özellik adı altında sınıflandırılmıştır: Yararlık, yürütülebilirlik, uygunluk ve doğruluk (JC, 1994).

- *Yararlık Standartları:* Yararlık standartları, program değerlendirme çalışmalarının, belirlenmiş değerlendirme amaçları ve ilgililerin bilgi gereksinimleri doğrultusunda yürütülmesini amaçlar (JC,1994). Bunlar:
 - ▀ *Paydaşların Tanımlanması:* Program değerlendirme sürecine katılan ve değerlendirme sonuçlarından etkilenmesi düşünülen birey ya da gruplar tanımlanır, gereksinimleri belirlenir ve planlama aşamasında bunlar dikkate alınır.
 - ▀ *Değerlendirici Güvenirliği:* Değerlendirmeler süreçte ve kullanıcılarla güvenilirliği kuran ve sürdüren bireylerce yürütülmelidir. Değerlendirmeyi yürüten bireyler araştırma yöntemleri ve konu alanı uzmanları olmaları yanında güvenilir olmalıdırlar.
 - ▀ *İlgili Verilerin Kapsamı ve Seçimi:* Değerlendirme, belirlenmiş değerlendirme amaçlarını ve katılımcıların ve kullanıcıların belirli ve ani gereksinimlerini karşılayan bilgilere erişmelidir.

- *Değerlerin Belirlenmesi:* Değerlendirme amaçlarının, sürecinin, bulgularının ve yorumlarının altında yatan değer yargıları net bir biçimde ifade edilir.
 - *Raporun Anlaşılabilirliği:* Değerlendirme raporu, değerlendirme amacını, içeriğini, araştırma sorularını, süreçte kullanılan yöntemleri ve değerlendirme bulgularını açık ve kapsamlı olarak betimler.
 - *Rapor Zamanlaması ve Yayımları:* Değerlendirme sonuçlarının karar vericiler ve geliştiriciler için kullanılabilirliği açısından, değerlendirme belirli bir takvime uygun olarak başlatılmalı, bitirilmeli ve yayımlanmalıdır.
 - *Değerlendirme Etkisi:* Değerlendirmenin planlanması, uygulanması ve raporlaştırılması paydaşların değerlendirme sürecine katılımlarına ve değerlendirme sonuçlarını kullanmalarına yardımcı olmalıdır.
- *Yürütülebilirlik Standartları:* Bu standartlar program değerlendirme çalışmasının mantıklı, titiz, ustaca ve uygun maliyetli bir biçimde planladığının ve yürütüldüğünün güvence altına alınmasını amaçlar (JC,1994).
- *Uygulanabilir Süreçler:* Program değerlendirme çalışmaları, programın normal olarak değerlendirildiğinin güvence altına alınması, program değerlendirme çalışmasının onaylanması ve bu çalışmalara olan katılımın artırılması, değerlendirme çıktılarının geçerliliğinin ve yararlılığının sağlanması için yürürlükteki örgütsel kurallara ve süreçlerle uyumludur.
 - *Politik Yeterlik:* Program değerlendirme çalışmaları birey ve grupların çıkarlarını tanır, izler ve dengeler.
 - *Maliyet Etkinliği:* Program değerlendirme çalışmaları kaynakları etkili ve verimli biçimde kullanılır.
- *Uygunluk Standartları:* Uygunluk standartları program değerlendirme çalışmalarının uygun, adil, yasal, doğru, kabul edilebilir olmasını amaçlar (JC,1994).
- *Duyarlı ve Kapsamlı Yönlendirme:* Program değerlendirme çalışmaları katılımcıların ve toplumun değerlerine karşı duyarlıdır.
 - *Resmi Sözleşme:* Program değerlendirme çalışması, katılımcıların gereksinimlerinin ve beklentilerinin resmi/gayri resmi olarak ifade edildiği

anlaşmalara dayanır.

- **İnsan Hakları ve Saygı:** Program değerlendirme çalışması yasal ve bireysel hakları korur ve katılımcıların saygınlığına ve etkileşimine saygı gösterir.
 - **Birey Etkileşimleri:** Değerlendiriciler, değerlendirmeye katılan bireylerle etkileşimlerinde onların onur ve değerlerine saygılıdır.
 - **Eksiksiz ve Dürüst Değerlendirme:** Program değerlendirme çalışması kullanıcılar için yeterince açık, anlaşılır ve eksiksizdir ve programdan ve değerlendirme süreç, uygulama ve ürünlerinden kaynaklanan yanlış etkilerden bağımsızdır.
 - **Bulguların Açıklanması:** Program değerlendirme bulgularının, sınırlıklarının ve sonuçlarının tarafsız, eksiksiz ve güvenilir bildirim program değerlendirme çalışmasının tüm katılımcılar için şeffaf olduğunu güvence altına alır.
 - **Çıkar Çatışmaları:** Program değerlendirme çalışmaları çıkar çatışmalarının değerlendirme süreçleri ve sonuçlarını tehlikeye atabileceği durumları belirler, sınırlar ve gerektiğinde uzlaştırıcı bir işlev yüklenir.
 - **Finansal Sorumluluk:** Program değerlendirme çalışmaları tüm harcamaların nedenlerini açıklar, sağlam işlem ve süreçlere uyumludur ve katılımcıların tüm harcamalar konusunda bilgi sahibi olmasını güvence altına alır.
- **Doğruluk Standartları:** Doğruluk standartları program değerlendirmede doğruluğu destekleyen süreçlerin, yöntemlerin ve desenlerin nasıl uygulanabileceğini ve değerlendirme bulgu ve sonuçlarında doğruluğu azaltıcı ve zayıflatıcı tutarsızlıkların, çarpıtmaların ve yanlış anlamaların nasıl giderilebileceğini açıklar (JC,1994).
- **Programın Tanımlanması:** Değerlendirilen program açık ve doğru bir biçimde tanımlanır.
 - **Çevrenin Analizi:** Programın içinde bulunduğu koşullar ayrıntılı biçimde ortaya konulur ve bu koşulların ve çevrenin program üzerindeki olası etkileri tanımlanır.
 - **Amaç ve İşlemlerin Tanımlanması:** Değerlendirmenin amaç ve işlemlerinin tanımlanır ve gözden geçirilir.

- ◆ *Savunulabilir Bilgi Kaynakları:* Program değerlendirmede kullanılan bilgilerin kaynakları ayrıntılı bir biçimde tanımlanır.
- ◆ *Geçerli Bilgi:* Değerlendirme yorumunun istenen kullanımının sağlayabilmesi için, değerlendirme amacıyla bilgi toplamada yapılacak işlemlerin seçimi ya da geliştirilmesinin ardından uygulamaya koyulmalıdır.
- ◆ *Güvenilir Veriler:* Bulgular, yorumlar ve sonuçların temelini oluşturan veriler kesin, güvenilir ve tutarlıdır.
- ◆ *Sistemik Bilgiler:* Değerlendirme için seçilen, uygulanan ve rapor edilen bilgi, sistemik olarak gözden geçirilir ve herhangi bir hata durumunda düzeltilir.
- ◆ *Nicel Bilginin Analizi:* Değerlendirmede yer alan nicel bilgiler uygun bir biçimde ve sistemik olarak analiz edilir.
- ◆ *Nitel Bilginin Analizi:* Değerlendirmede yer alan nitel bilgiler uygun bir biçimde ve sistemik olarak analiz edilir.
- ◆ *Doğrulanabilir Sonuçlar:* Değerlendirmede ulaşılan sonuçlarının paydaşlarca değerlendirilmesi sağlanır
- ◆ *Tarafsız Raporlaştırma:* Raporlaştırma işlemleri kişisel duygular ve çıkarılardan bağımsız hazırlanır.
- ◆ *Meta-değerlendirme*

Amerika'da Ortak Komite tarafından geliştirilen bu standartlar dışında yine Amerikan Değerlendirme Derneği (AEA) tarafından program değerlendiricilerin izleyebileceği ilkeler geliştirilmiştir (AEA, 2005). Altı alanda geliştirilen ilkeler şunlardır:

- *Sistemik Sorgulama:* Program değerlendiriciler sistemik ve veri-tabanlı sorgulamalar yürütürler (AEA, 2005). Bunu sağlamak için;
 - ◆ Kullandıkları yöntemlere uygun üst düzeyde teknik standartları dikkate alırlar.
 - ◆ Değerlendirme sonuçlarını kullanacaklar ile araştırma sorularının ve değerlendirme yaklaşımının güçlük ve eksikliklerini araştırırlar.
 - ◆ Katılımcıların çalışmayı anlamaları, yorumlamaları ve eleştirebilmeleri için değerlendirici, değerlendirmenin yaklaşımını, yöntemini ve sınırlıklarını tam ve ayrıntılı olarak aktarır.

- *Yeterlik:* Program değerlendiriciler program paydaşlarına yeterli performans sunar (AEA, 2005). Bunun için şu ilkeleri benimserler.
 - Program değerlendirici değerlendirme ekibinin değerlendirmeye uygun eğitim, bilgi, beceri ve deneyime sahip olduğunu güvence altına alır.
 - Program değerlendirici değerlendirme takımının kültürel yeterliklerini süreçte kullanmalarını ve kültürel olarak farklı gruplar ile çalışmak için uygun değerlendirme stratejilerini ve becerilerini kullanmalarını güvence altına alır.
 - Program değerlendiriciler kendi sınırlıklarını bilirler ve bu sınırlıkları aşan çalışmalara başlamaya çekinirler.
 - Değerlendirme çalışmalarında üst düzey bir performans sağlamak için program değerlendirme yeterliklerini sürdürme ve geliştirme yollarını ararlar.
- *Doğruluk:* Program değerlendiriciler dürüst davranışlar sergiler ve bu davranışları tüm değerlendirme süreci boyunca sürdürürler. Bunun için program değerlendiriciler şu ilkelere eksiksiz uyarlar (AEA, 2005).
 - Maliyet, sorunlar, yöntemin sınırlıkları, sonuçların kapsamı, verilerin kullanımı gibi konularda ilgililer ile dürüst bir anlaşma içindedirler.
 - Program değerlendiriciler görevi kabul etmeden önce bir çıkar çatışmasını çıkarabilecek görevleri ve ilişkileri açıklar.
 - Program değerlendiriciler üzerinde anlaşılmış orijinal proje planlarıyla ilgili yapılan tüm değişiklikleri, bu değişikliklerin nedenlerini ve olası etkilerini kaydeder ve raporlaştırır.
 - Program değerlendiriciler değerlendirme ile ilgili kendilerinin ve paydaşlarının çıkarlarını ve değerlerini açıkça ortaya koyar.
 - Program değerlendiriciler yapılan çalışmanın başkaları tarafından yanlış kullanımını engellemek için değerlendirme sürecini, verilerini ve bulgularını eksiksiz olarak tanımlar.
 - Hatalı değerlendirme bilgileri üretmeye yatkın etkinlik ve süreçlerle ilgili her bir kaygıyı azaltırlar, kaygıların giderilemediği durumlarda çalışmayı durdurur veya farklı yollar denemek için meslektaşlarına ve paydaşlara danışırlar.

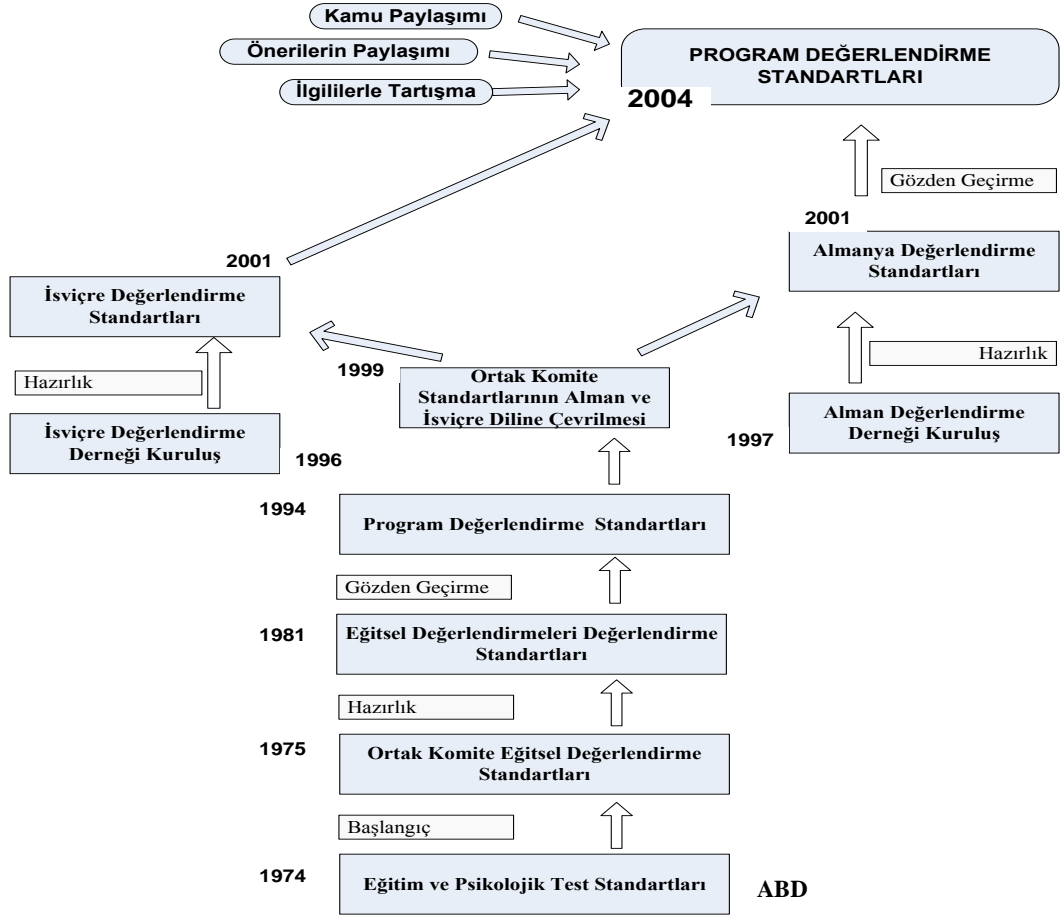
- ▀ Değerlendirme için ayrılan tüm finansal destekleri ve değerlendirme çalışmasının yürütülmesini isteyen kaynağı açıklar.
- *İnsan Hakları:* Program değerlendiriciler tüm katılımcıların (değerlendirici, paydaş) güvenliğine, itibarına ve öz-değerlerine saygı gösterir. Bunun için şu ilkelere uyarlar (AEA, 2005).
 - ▀ Katılımcılar, etkileyen zararlar, olası riskler, gizlilik, onaylar ile ilgili var olan mesleki etik kurallarına, standartlara ve düzenlemelere uyarlar.
 - ▀ Yararları arttırmanın ve değerlendirmeden kaynaklanabilecek her tür gereksiz zararı azaltmanın yollarını arar, olası risklerden dolayı değerlendirme veya süreçlerin sağlayacağı yararların önceden belirlenmesi gerektiğinde durumu dikkatli bir şekilde yargırlar.
 - ▀ Paydaşların itibarına ve öz-değerlerine saygılı bir biçimde değerlendirmeyi yürütür ve sonuçlarını aktarırlar.
 - ▀ Değerlendirmeye katkıda bulunanlara bir karşılık olarak değerlendirmede-uygun olduğunda- sosyal eşitliği teşvik ederler.
 - ▀ Paydaşlara arasındaki kültür, din, yetersizlik, yaş, cinsel yönelim ve etnik durumlar farklılıkları anlar, dikkate alır onlara saygı gösterirler.
- *Genel ve Kamu Refahına Karşı Sorumluluk:* Program değerlendiriciler genel ve kamunun ilgi ve değerleri arasındaki farklılığı dikkate alır ve aktarırlar. Bunun için şu ilkelere dikkat ederler (AEA, 2005).
 - ▀ Tüm paydaşların ilgi ve bakış açılarını dikkate alır.
 - ▀ Yalnızca anlık eylemleri ve değerlendirme sonuçlarını değil, geniş varsayımlar, çıkarımlar, olası yan etkileri de dikkate alır.
 - ▀ Paydaşların katılımına izin verir, değerlendirme bilgilerini paylaşır ve değerlendirme sonuçlarını bireylere ve gizliliğe saygılı bir biçimde sunar.
 - ▀ Program değerlendirme sonuçlarını uygulayacak bireyler ile diğer paydaşların gereksinimleri ve çıkarları arasında bir denge oluşturur.
 - ▀ Bir bütün olarak toplumun refahını düşünüp paydaşların çıkarlarını derinlemesine çözümleyerek kamu ilgi ve çıkarını dikkate alır.

Amerika’da geliştirilen bu standartlar dışında İsviçre, Almanya ve Kenya, Ortak Komite Program Değerlendirme Standartlarını (OKPDS) kendi ülkelerine uyarlamış; Avustralya ve Kanada da bu standartları onaylamakla birlikte kendi ülkelerine özgü yeni standartlar da geliştirmişlerdir. Fransa ise yalnızca kendisinin geliştirdiği standartları kullanmaktadır. Aşağıda bu ülkelerde kullanılmakta olan (uyarlanmış veya geliştirilmiş) program değerlendirme standartları açıklanmıştır.

İsviçre Program Değerlendirme Standartları: Birçok uzman İsviçre’yi Avrupa’da değerlendirme konusunda lider bir ülke olarak görmese de 1980lerin son ve 1990ların ilk yıllarında İsviçre Federasyonu Komitesi tarafından İsviçre Değerlendirme Derneğinin kurulmasıyla sonuçlanan ulusal değerlendirme araştırmaları programı başlatılmıştır (Widmer, 2004). Nitelik sorunu programın temel konusuydu ve ilk olarak bu programda program değerlendirme standartları iki meta-değerlendirme çalışmasında uygulandı. Ortak Komitenin geliştirmiş olduğu değerlendirme standartlarının ilk baskısında yer alan standartların kullanıldığı bu iki meta-değerlendirme çalışmasında endüstri, enerji, sosyal politika, dış ilişkiler, çevre vb. alanlardan 15 çalışmanın niteliği değerlendirilmiştir. Sonuçlar standartların İsviçre ortamında eğitim alanı dışında da kullanılabilirliğini ortaya koymuştur. Bununla birlikte, bazı standartlar da önerilmiştir (Widmer, 2004).

1996 yılında kurulan İsviçre Değerlendirme Derneğinin ilk etkinliği bu meta-değerlendirme çalışmalarının sonuç ve önerilerini inceleyerek gerekli nitelik sorunlarını belirleyecek ve standartların ayrıntılı incelemesini yapacak bir birim kurmak olmuştur. Kurul yaptığı toplantılar sonucu İsviçre’ye özgü değerlendirme standartlarının geliştirilmesi gerektiğini ve geliştirilecek yeni standartların da Ortak Komitenin geliştirmiş olduğu değerlendirme standartlarının gözden geçirilmiş baskısının temel alınarak geliştirilmesini önermiştir. Bu öneriler doğrultusunda, İsviçre Değerlendirme Derneği (SEVAL) üyelerinden bir grup ile geliştirme birimi birçok kez bir araya gelmiş ve gerekli düzenlemeleri yaparak 1999 yılında İsviçre’ye özgü dört temel alanda 27 standarttan oluşan ilk standartları ortaya koymuştur. 2001 yılında ise genel kurul bu standartları onaylanmıştır. Böylece İsviçre kendi program değerlendirme standartlarını geliştiren ilk ülke olmuştur (Widmer, 2004). Dört temel alanda 26 standarttan oluşan İsviçre program değerlendirme standartları Tablo 7’de verilmiştir.

Almanya Program Değerlendirme Standartları: ABD ve İsviçre'den sonra program değerlendirme standartlarını geliştirmeye başlayan Almanya, bu çalışmalardan önce program değerlendirme alanında başarılı bir ülke olarak görülmemiştir (Derlien, 2002; Stockmann, 2004; Struhkamp, 2005). Bu tarihe kadar ülkede program değerlendirmenin kapsamının oldukça sınırlı olduğu görülmektedir. Örneğin, üniversitelerde değerlendirmeye yönelik herhangi bir programın olmadığı, değerlendirmeye ilişkin basılı bir kitap, kitapçık ya da akademik dergiye rastlanmadığı ve eğitimin birçok alanı için yapılmış meslek tanımlamasının değerlendiriciler için yapılmadığı görülmektedir. Değerlendirme yalnızca programın bir ögesi olarak ele alınmakta, değerlendirme yöntemleri, modelleri, desenleri, değerlendirmede dikkat edilecek etik ilkeleri ihmal edilmektedir (Beywl, 2000). 1990'lı yılların sonunda İsviçre'nin değerlendirme standartları üzerine yaptığı çalışmalar ve OKPDS'nin Alman diline çevrilmesiyle birlikte Alman Değerlendirme Derneği (DeGEval) tarafından Almanya'ya özgü program değerlendirme standart geliştirme çalışmaları başlamıştır. Bu amaçla dernek üyeleri arasından seçilen bir grup, genel kurul tarafından belirlenen bireylerin de katılımıyla bir standart geliştirme komitesi oluşturmuştur. Bu komite OKPDS ve İsviçre değerlendirme standartlarını inceleyip, gerekli uyarılama çalışmalarını yaptıktan sonra Almanya için taslak bir program değerlendirme standartları listesi oluşturmuştur. Bu taslak standartlar 2001 yılında genel kurulun da onayı ile Almanya için program değerlendirme standartları olarak kabul edilmiştir. (Beywl ve Speer, 2004; Struhkamp, 2005). OKPDS ve İsviçre program değerlendirme standartları ile aynı yapıda olan Alman program değerlendirme standartları, dört temel standart alanı altında 25 standarttan oluşmaktadır (Tablo 7). Her üç ülkenin program değerlendirme standartlarının tarihsel gelişimi Şekil 16'da gösterilmiştir.



Şekil 16. OKPDS, İsviçre ve Almanya Program Değerlendirme Standartlarının Tarihsel Gelişimi

Kaynak: Beywl, 2000'den sadeleştirilmiştir.

Afrika Program Değerlendirme Standartları: Afrika'da değerlendirme konusunda ilk gelişmeler 1990'lı yıllarda başlamıştır. İlk olarak 1990 yılında geniş bir katılımı ve önemli kuruluşların sponsorluğu altında gerçekleştirilen bir seminerde Afrika ülkelerinin temel değerlendirme gereksinimlerinin neler olduğu, ülkelerin kendilerine özgü değerlendirme yeterliklerini nasıl kazanacakları ve bu yeterlikleri nasıl geliştirilecekleri tartışılmıştır (Patel ve Russon, 1999). Bu seminerden sonra durgun bir döneme giren süreç, 1990'lı yılların sonlarına doğru tekrar hareketlenmiş ve birbiri ardına önemli değerlendirme konferansları ve seminerleri düzenlenmiştir. Bunlardan biri de 1998 yılında UNICEF'in sponsorluğunda düzenlenen bir değerlendirme çalıştaydır. Bu çalıştayın ilk bölümünde 1994 yılı OKPDS'nin ayrıntılı öğretimi yapılmış ve uygulama örnekleri gösterilmiştir. İkinci bölümde ise, katılımcılardan Afrika kültürünü dikkate alarak, bu standartların uygunluğunu tartışmaları istenmiştir. Bu çalıştay sonucunda

ortaya çıkan bulgular bir sonraki yıl bir arařtırmacı tarafından yayına dönüřtürülmüř ve “Afrika Deęerlendirme Yönergesi” haline dönüřtürülmüřtür (Patel ve Russon, 1999). Bu yönerge, OKPDS temel alınarak Afrika költürüne uygun biçime getirilmiř, uzmanlar tarafından uygun bulunmayan bazı standartlar yeniden gözden geçirilmiř ve geniř katılımlı bir seminerde tartıřılarak son biçimini almıřtır. OKPDS’den uyarlanan Afrika program deęerlendirme standartları, dört standart alanı altında 30 standarttan oluřmaktadır (Rouge, 2004; Patel ve Russon,1999).

OKPDS, İsviçre, Almanya ve Afrika program deęerlendirme standartlarının karřılařtırmalı incelemesi Tablo7’de verilmiřtir.

Tablo 7.

Program Değerlendirme Standartlarının Karşılaştırılması

		ABD (OKPDS)	İsviçre	Almanya	Afrika
Yararlık Standartları	1.	Paydaşların Tanımlanması	Paydaşların Tanımlanması	Paydaşların Tanımlanması	Paydaşların Tanımlanması
	2.	Değerlendirici Güvenirliği	Değerlendirme Amaçlarının Belirlenmesi	Değerlendirme Amaçlarının Belirlenmesi	Değerlendirici Güvenirliği
	3.	İlgili Verilerin Kapsamı ve Seçimi	Değerlendirici Güvenirliği ve Yeterliği	Değerlendirici Güvenirliği ve Yeterliği	İlgili Verilerin Kapsamı ve Seçimi
	4.	Değerlerin Belirlenmesi	İlgili Verilerin Kapsamı ve Seçimi	İlgili Verilerin Kapsamı ve Seçimi	Değerlerin Belirlenmesi
	5.	Raporun Anlaşılabilirliği	Değerlerin Açıklığı	Değerlerin Açıklığı	Raporun Anlaşılabilirliği
	6.	Rapor Zamanlaması ve Yayımı	Raporun Kapsamlılığı ve Anlaşılabilirliği	Raporun Kapsamlılığı ve Anlaşılabilirliği	Rapor Zamanlaması ve Yayımı
	7.	Değerlendirme Etkisi	Rapor Zamanlaması	Rapor Zamanlaması	Değerlendirme Etkisi
	8.	-----	Değerlendirme Kullanımı	Değerlendirme Kullanımı	-----
Yürütülebilirlik Standartları	1.	Uygulanabilir Süreçler	Uygulanabilir Süreçler	Uygun Süreçler	Uygulanabilir Süreçler
	2.	Politik Yeterlik	Politik Yeterlik	Diplomatik Yürütülebilirlik	Politik Yeterlik
	3.	Maliyet Etkinliği	Maliyet Etkinliği	Değerlendirme Verimliliği	Maliyet Etkinliği
Uygunluk Standartları	1.	Duyarlı ve Kapsamlı Yönlendirme	Resmi Sözleşme	Resmi Sözleşme	Duyarlı ve Kapsamlı Yönlendirme
	2.	Resmi Sözleşme	İnsan Hakları ve Saygı	Birey Haklarının Korunması	Resmi Sözleşme
	3.	İnsan Hakları ve Saygı	Birey Etkileşimleri	Eksiksiz ve Dürüst Araştırma	İnsan Hakları ve Saygı
	4.	Birey Etkileşimleri	Eksiksiz ve Dürüst Değerlendirme	Yansız Uygulama ve Raporlaştırma	Birey Etkileşimleri
	5.	Eksiksiz ve Dürüst Değerlendirme	Bulguların Açıklanması	Bulguların Açıklanması	Eksiksiz ve Dürüst Değerlendirme
	6.	Bulguların Açıklanması	Çıkar Çatışmaları	-----	Bulguların Açıklanması
	7.	Çıkar Çatışmaları	-----	-----	Çıkar Çatışmaları
	8.	Finansal Sorumluluk	-----	-----	Finansal Sorumluluk
Doğruluk Standartları	1.	Programın Tanımlanması	Programın Tanımlanması	Programın Tanımlanması	Programın Tanımlanması
	2.	Çevrenin Analizi	Çevrenin Analizi	Çevrenin Analizi	Çevrenin Analizi
	3.	Amaç ve İşlemlerin Tanımlanması	Amaç ve İşlemlerin Tanımlanması	Amaç ve İşlemlerin Tanımlanması	Amaç ve İşlemlerin Tanımlanması
	4.	Savunulabilir Bilgi Kaynakları	Savunulabilir Bilgi Kaynakları	Savunulabilir Bilgi Kaynakları	Savunulabilir Bilgi Kaynakları
	5.	Geçerli Bilgi	Geçerli ve Güvenilir Bilgiler	Geçerli ve Güvenilir Bilgiler	Geçerli Bilgi
	6.	Güvenilir Bilgiler	Sistemik Bilgi İnceleme	Sistemik Bilgi İnceleme	Güvenilir Bilgiler
	7.	Sistemik Bilgiler	Nicel ve Nitel Bilginin Analizi	Nicel ve Nitel Bilginin Analizi	Sistemik Bilgiler
	8.	Nicel Bilginin Analizi	Doğrulanabilir Sonuçlar	Doğrulanabilir Sonuçlar	Nicel Bilginin Analizi
	9.	Nitel Bilginin Analizi	Meta-değerlendirme	Meta-değerlendirme	Nitel Bilginin Analizi
	10.	Doğrulanabilir Sonuçlar	-----	-----	Doğrulanabilir Sonuçlar
	11.	Tarafsız Raporlaştırma	-----	-----	Tarafsız Raporlaştırma
	12.	Meta-değerlendirme	-----	-----	Meta-değerlendirme
		30 Standart	26 Standart	25 Standart	30 Standart

Tablo 7 incelendiğinde, Afrika program değerlendirme standartlarının OKPDS ile aynı olduğu, ancak ayrıntılı incelemede bazı standart ifadelerinin tanımlamalarının yeniden oluşturulduğu görülmektedir. Almanya ve İsviçre tarafından uyarlanan standartların da OKPDS ile karşılaştırıldığında, standart alanlarının aynı olduğu ancak bazı standart maddelerinin çıkarılıp yerine yenilerinin eklendiği görülmektedir. Örneğin, Almanya “Yararlık” standartları alanı altında “Değerlendirme Amaçlarının Belirlenmesi” adı altında yeni bir standarda, “Uygunluk” standartları alanı altında “Yansız Uygulama ve Raporlaştırma” standardına yer vermiştir. Öte yandan, OKPDS “Uygunluk” standartları alanı altında yer alan “Duyarlı ve Kapsamlı Yönlendirme”, “Birey Etkileşimleri”, “Çıkar Çatışmaları” ve “Finansal Sorumluluk” standartları ile “Doğruluk” standartları alanı altında yer alan “Tarafsız Raporlaştırma” standartlarını çıkarmıştır. Bazı standartlar ise tanımları değiştirilmeden yalnızca adları değiştirilerek kabul edilmiştir. İsviçre tarafından uyarlanan standartlar incelendiğinde durumun Almanya ile aynı olduğu görülmektedir. İki ülke standartları arasındaki tek fark İsviçre’nin “Uygunluk” standartları alanı altında yer alan “Çıkar Çatışmaları” standardına yer vermesidir.

Fransa Program Değerlendirme Standartları: Fontaine ve Monnier (2002)’e göre, Fransa son çeyrek asırda yaşadığı kayda değer gelişmelere karşın, henüz değerlendirme alanında istediği yerde değildir. 1990’lı yılların ilk yıllarından itibaren değerlendirmeye gereken önem verilmeye başlanmış ve birçok kitap, bildiri yayınlanmış, değerlendirme dergileri çıkarılmıştır. Bu yıllarda Bilimsel Değerlendirme Kurulu (Conseil scientifique de l’évaluation) değerlendirme uygulamalarının niteliğinin değerlendirilmesi için bir dizi standart geliştirmiştir. Sonraki yıllarda Fransa Değerlendirme Derneğine (Société française d’évaluation, SFE) dönüştürülen bu kurul, geliştirmiş oldukları eski standartların dışında “bildirge” (charter) olarak tanımladıkları bir değerlendirme seti geliştirmeye başlamıştır. SFE’nin içinde kurulan Standartlar ve Etik Çalışma Grubu, öncelikle farklı ülkelerdeki değerlendirme standartlarını incelemiş, uzun süren görüşme ve tartışmalardan sonra ülkeye özgü genel bir dokümanın hazırlanmasına karar vermiştir. 2003 yılında çalışmalar tamamlanmış ve “kamu politikaları ve programları için değerlendirme yönerge ilkeleri” adı altında geliştirilen standartlar derneğinin genel kurulu tarafından onaylamıştır. İki sayfadan oluşan bildirgede bir giriş ve kısaca tanımlanmış altı standart yer almaktadır (Widmer, 2004). Standartlar şunlardır:

- *Çoğulculuk:* Değerlendirme, değerlendirme çalışması ile ilgili paydaşların yasal bakış açılarını dengeli bir biçimde yansıtır.
- *Bağımsızlık:* Değerlendirme çalışması tarafsız bir şekilde yürütülür. Değerlendirme uzmanları süreçte karşılaşılabilecek çıkar çatışmaları ile ilgili olarak tüm katılımcıları bilgilendirir. Değerlendirme süreci program yönetiminden ve kara-alma sürecinden bağımsız olarak gerçekleştirilir. Bu bağımsızlık karar-alıcıların özgür seçimleri için bir garanti işlevi görür.
- *Yeterlik:* Değerlendirme çalışmasını tasarlama, yönetme, değerlendirmede kullanılacak verileri toplama, yöntemleri seçme, bulguları yorumlama ve öneriler geliştirme değerlendiricinin süreçte özel yeterlikler kullanmasını gerektirir. Değerlendiriciler bu nedenle mesleki becerilerini sürekli olarak geliştirmeli ve güncellemelidir.
- *Bütünlüğe Saygı:* Katılımcılar değerlendirme sürecinde tüm bireylerin haklarına, gizliliğine, güvenliğine, birey bütünlüğüne saygılıdır.
- *Şeffaflık:* Değerlendirme bulgularının sunumu; değerlendirilecek konunun, amacın, araştırma sonuçlarının, değerlendirmede kullanılacak yöntemlerin, sınırlıkların, bulguların, öneri ve tartışmaların açık ve net bir biçimde ortaya konulmasıyla birlikte yapılmalıdır. Bulguların bütünlüğüne tüm katılımcılar tarafından saygı gösterilmelidir.
- *Sorumluluk:* Değerlendirme sürecinin başlangıcında katılımcılar arasında süreç yönetimi, araştırma ve çözümlenme, bulguların yorumlanması ve yaygınlaştırılması gibi sorumluluklar paylaşılmalıdır.

Altı standarttan oluşan Fransa program değerlendirme standartlarının uygulamaya dönük herhangi bir alt standart ya da göstergesi bulunmamaktadır.

Kanada Program Değerlendirme Standartları: Program değerlendirmenin önem kazandığı II. Dünya Savaşı sonrası yıllarda Amerika Birleşik Devletleri'nden sonra program değerlendirmeye büyük önem veren ülke Kanada olmuştur (Greene, 2002). İlk

önemli program değerlendirme çalışmaları 1960'lı yıllarda yerel hükümet tarafından yürütülen eğitim programlarının değerlendirilmesidir.

1980'li yıllarda Kanada Değerlendirme Derneği'nin de kurulmasıyla program değerlendirme çalışmaları daha sistematik yürütülmeye başlamıştır. Bu çalışmalarla birlikte ülkede yeni program değerlendirme kitapları, akademik dergiler çıkmış ve değerlendiricilere yönelik dersler açılmıştır. Doğal olarak yapılan çalışmalarda nitelik aranmaya başlanmış ve standartlaşmaya gidilmiştir. Bu konuda yapılan ilk çalışmalar değerlendirme derneğinin de sponsor olarak aralarında yer aldığı Ortak Komite tarafından geliştirilen standartların kabul edilmesi, diğeri ise 1996 yılında "Program Değerlendirmede Etik İlkeler" in geliştirilmesidir (Keith, 2003). Bu ilkeler şunlardır:

- *Yeterlik:* Değerlendiriciler değerlendirme çalışmalarında aşağıda verilen yeterliklere sahip olmalıdır.
 - ▀ Değerlendiriciler değerlendirme çalışmasında uygun sistematik araştırma yöntemleri uygulamalıdır.
 - ▀ Değerlendiriciler değerlendirme çalışması için yeterli konu alanı bilgisine sahip olmalı ve onu işe koşmalıdır.
 - ▀ Değerlendiriciler sürekli olarak yöntemsel ve uygulama bilgi ve becerilerini geliştirmelidirler.
- *Bütünlük:* Değerlendiriciler diğer paydaşlarla olan ilişkilerinde bir bütünlük içinde olmalıdır. Değerlendiricinin bu bütünlüğün sağlanmasında aşağıdaki ilkelere dikkat etmesi gereklidir.
 - ▀ Değerlendiriciler değerlendirme çalışmasına ilişkin bilgi ve beceri düzeylerini net olarak açıklamalıdır.
 - ▀ Değerlendiriciler değerlendirme çalışmasından önce değerlendirme sonuçlarını talep eden bireylere çıkar çatışması ile ilgili bilgileri sunmalı ve bu çatışmaların nerede gerçekleşebileceğini ifade etmelidir.
 - ▀ Değerlendiriciler tüm paydaşların kültürel ve sosyal çevrelerine karşı duyarlıdır ve bu ortamlara uygun davranışlarda bulunur.
 - ▀ Değerlendiriciler değerlendirme sonuçlarını talep eden bireylerle gizlilik,

özerklik, iletişim, bulguların ve raporun mülkiyeti gibi konularda sözleşme yaparlar.

- **Sorumluluk:** Değerlendiriciler için çalışma sürecinde gösterdikleri performans ve ortaya konan ürünün niteliği önemlidir. Bu nedenle süreçte değerlendiricinin aşağıdaki sorumluluklarını eksiksiz yerine getirmesi gereklidir.
 - Değerlendiriciler değerlendirme sonuçlarını uygulayacak bireyleri, onların uygun değerlendirme yöntem ve stratejilerinin seçimi ile ilgili karar almalarını kolaylaştırmak amacıyla bilgilendirmekle sorumludurlar.
 - Değerlendiriciler çalışma bulgularının, sınırlıklarının ve önerilerinin açık, doğru, yazılı ve/veya sözel sunumlarından sorumludurlar.
 - Değerlendiriciler değerlendirme süreci ile ilgili yapılan harcamalarla ilgili bilgiler düzenli olarak incelendiği için kendi mali kararlarından sorumludurlar.
 - Değerlendiriciler değerlendirme çalışmasının önceden tüm katılımcılarla belirlenen zamanda bitirilmesinden sorumludurlar. Bu tür anlaşmalar değerlendiricinin kontrolü dışında gerçekleşen ve önceden kestirilemeyen gecikmelerin onaylanmasını kolaylaştırır.

Görüldüğü gibi Kanada program değerlendirme standartlarının diğer ülke standartlarından farklı olarak değerlendiricinin sahip olması gereken standartlar üzerinde durduğu görülmektedir.

Avustralya Program Değerlendirme Standartları: Avustralya'da program değerlendirme çalışmalarının geçmişi 1950'li ve 60'lı yıllara dayanır. Uzun soluklu bu değerlendirme çalışmaları teknik eğitim ve zirai genişleme programlarını kapsamaktaydı. Tam anlamıyla eğitim programı değerlendirme 1980'li yıllarda gerçekleşmiştir. Devletin de desteğiyle birlikte birçok konferans düzenlenmiş ve değerlendirme raporu hazırlanmıştır (Straton, 2001). Program değerlendirme standartlarının geliştirilmesine dönük ilk çalışmalar 1990'lı yıllarda Avustralya Değerlendirme Kurulu (AES) tarafından yapılmıştır. Program değerlendirmenin özellikle etik boyutu üzerinde duran standartlar 2000 yılında tamamlanmış ve 2006 yılında gözden geçirilerek yayınlanmıştır (AES, 2009). Avustralya ve Yeni Zelanda'da

kullanılmak üzere geliştirilen “Değerlendirmenin Etik Uygulanmasına İlişkin Yönerge”si üç temel alanda 21 alt standarttan oluşmaktadır. Bazı standart maddelerinin göstergeleri de bulunmaktadır (AES, 2009).

- Onaylama ve Hazırlama
 - ▀ *Açıklayıcı Belge:* Değerlendirmeye önderlik yapan bireyler, çalışmanın ardındaki gerekçeyi, mantığı, kapsamı, çalışmada yanıtı aranacak soruları, kullanılacak yaklaşımları, dikkate alınacak sorunları ve değerlendirme sonuçlarının kimler tarafından kullanılacağını ifade eden bir açıklayıcı doküman hazırlamalıdır.
 - ▀ Değerlendirme çalışmasında paydaşlar tanımlanmalı ve değerlendirme sürecinin paydaşlar üzerindeki olumlu ve olumsuz etkileri ortaya konmalıdır.
 - ▀ *Sınırlıkların ve Çıkarların Belirlenmesi:* Değerlendiriciler değerlendirme çalışmasının olası yöntemsel ve etik sınırlıklarını belirlemeli ve bu sınırlıkların değerlendirme süreci ve sonuçları üzerindeki etkisini ortaya koymalıdır. Değerlendirme çalışmasını yürütenler ile paydaşların ve sonuçları kullanacak kurum, kuruluş veya bireyler arasındaki çıkarlar net olarak ifade edilmelidir.
 - ▀ *Yazılı Anlaşma:* Değerlendirme çalışması değerlendirme komisyonu ile değerlendiriciler arasında bir yazılı ve sözlü anlaşmaya dayandırılmalıdır. Anlaşma şartları, kaynak kullanımı, değerlendirmenin tamamlanma süresi, hakların, gizliliğin korunması, verilerin toplanması ve açıklanmasına ilişkin haklara saygı, değerlendirme sonuçlarının sunulması ve paylaşılması gibi konuları içermelidir.
 - ▀ *Değişken Şartlar:* Değerlendirme çalışmasında yazılı/sözlü anlaşmanın şartlarına göre hareket edilmesi beklenir, ancak önceden fark edilemeyen durumlarda veya gereksinimler doğrultusunda yapılması gereken değişikliklerde her iki taraf da önerilerde bulunma ve sözleşmeyi yenileme sorumluluğuna sahiptir.
 - ▀ *Olası Riskler veya Hasarlar:* Değerlendirmenin temelindeki düşünce /felsefe veya değerlendirme içinde özel süreçler değerlendiricileri, paydaşları veya değerlendirme sonuçlarını bekleyen bireyleri etkileyebilecek olası hasarlar

veya riskler ışığında incelenmelidir. Bu sorunlar önceden kestirilmeli ve tartışılmalıdır.

- **Değerlendirme Yeterliği:** Değerlendirici veya değerlendirme takımı değerlendirmenin yürütülmesi için gerekli bilgi, beceri, yeterlik ve deneyimlere sahip olmalıdır.
- **Çıkar Çatışması:** Değerlendirici veya değerlendirme takımı, değerlendirmenin yürütülmesi sürecinde çıkar çatışması ile sonuçlanabilecek görevini/görevlerini belirmelidir.
- **Dürüst Rekabet:** Değerlendirici bir değerlendirme çalışması yürütmek için diğer adaylarla rekabet ettiğinde, kendilerini açık, dürüst ve uzmanlıklarını gösteriri bir biçimde ifade etmelidir. Değerlendirme komisyonunu ve/veya değerlendiricileri seçenler gelen tüm önerileri açık ve dürüstçe dikkate almalıdırlar.
- **Yürütme**
 - **Farklılıklar ve Eşitsizlikler:** Irk, din, dil, yaş, cinsiyet, fiziksel ve zihinsel kapasite, sosyal, ekonomik ve etik değişkenler vb. farklılıklar değerlendirme sürecinde ve değerlendirmenin raporlaştırılma aşamasında dikkate alınmalıdır.
 - **Değerlendirme Amacı ve Komisyonunun Belirlenmesi:** Değerlendirme çalışmasında görev alacak bireyler kendilerini ve değerlendirme komisyonunda görev alan bireyleri ilgililere tanıtmalı ve değerlendirme çalışmasının amacı ile ilgili önerilerini paylaşmalıdır.
 - **Bilgilendirilmiş Onaylama:** Değerlendirme çalışmasında kullanılacak verilerin sağlandığı bireylere bu bilgilerin nerde, nasıl ne amaçla kullanılacağı ve değerlendirme çalışması içinde bu verilerin işlevi ile bilgiler sunulduktan sonra bireylerden onay alınmalıdır.
 - **Yeterli Ayrıntı:** Değerlendirme çalışması özellikle araştırma deseni, veri toplama yöntemleri ve verilerin çözümlenmesi teknikleri ile ilgili olarak bilgileri ayrıntılı olarak içermelidir.
 - **Sınırlıkların Açıklanması:** Değerlendirme çalışmasında değerlendiricinin yeterliğini aşan durumları değerlendirme komisyonuyla paylaşması gereklidir.

- **Gizlilik:** Değerlendirme süreci, değerlendirme bulguları veya sonuçları ile ilgili bilgiler komisyon ve katılımcılar tarafından uygun görülmedikçe açıklanmamalıdır.
- **Önemli Sorunların Bildirimi:** Değerlendirici değerlendirilen programla ilgili önemli bir sorunla veya beklenmeyen bir durumla karşılaştığında değerlendirme komisyonunu bilgilendirmelidir.
- **Aykırı Tutumların Öngörülmesi:** Değerlendiriciler süreç içinde yasal olmayan bir etkinlik veya etkinliğe teşebbüs gibi durumlarla (ayrımcılık, ırkçılık, cinsel taciz vb.) karşılaştığında etik ve yasal olarak şunlardan sorumludur:
 - Bu sorunları giderici veya azaltıcı önlemler almalıdır.
 - Sorunu ilgili makamlara bildiren yasal bir rapor hazırlamalıdır.
 - Gizlilikle ilgili paydaşlarla yapılan sözleşmeyi devam ettirmekle sorumludur.
- **Raporlaştırma**
 - **Sadelik ve Açıklık:** Paydaşların ve değerlendirme sonuçlarını talep eden tüm ilgililerin değerlendirme süreci ve değerlendirme bulgu ve sonuçları ile bilgileri kolaylıkla anlayabilmeleri için değerlendirme raporu sade, açık ve basit bir dille hazırlanmalıdır.
 - **Doğruluk ve Kapsam:** Değerlendirmede kullanılan veri toplama yöntemlerinin, araştırma desenlerinin, değerlendirme bulgularının, sınırlıklarının açıklanmasında değerlendirme raporu eksiksiz, kapsamlı ve güvenilir olmalıdır.
 - **Kaynaklar ve Alıntılar:** Değerlendirme çalışmasında kullanılan çalışmalar net bir biçimde ortaya konmalı ve değerlendirme ile ilgili takdir ve teşekkürler çalışmaya önemli katkılar getirenlere yapılmalıdır.
 - **Eksiksiz Bulgular ve Sonuçlar:** Son değerlendirme raporu değerlendirici tarafından ortaya konmuş sonuçları ve bulguları eksiksiz olarak yansıtmalı ve değerlendiricinin izni dışında değiştirilmemelidir.
 - **Raporun Bütünlüğü:** Değerlendirme raporu ile ilgili bilgilerin sunulmasında komisyon üyeleri raporunun bütünlüğünü korumakla sorumludurlar.

1.6. Sorun

Günümüzde her alanda olduğu gibi, eğitim alanında da nitelik ve verimliliğin artırılması önemli bir gerekliliktir. Her tür ve düzeydeki eğitim kurumlarında istenilen niteliğin sağlanmasında temel koşullardan biri nitelikli programların geliştirilmesidir. Programların niteliği ise programda öngörülen davranışları kazandırmasına bağlıdır. Bunun belirlenebilmesi içinde programların sürekli olarak değerlendirilmesi ve değerlendirme sonuçlarının programın geliştirilmesinde kullanılması gereklidir.

Programın etkililiğinin sorgulanması ve değerlendirilmesi, programların geliştirilmesi için bir başlangıç noktasıdır (Gözütok, 2001). Ancak, program geliştirme sürecinde son derece önemli bir yere sahip olan program değerlendirme, ne yazık ki Türkiye’de de en zayıf kalınan alanlardan biri olarak karşımıza çıkmaktadır. Bu konuda yapılan çalışmalarda alınan sonuçların ya tek bir ölçüte dayalı ya da oldukça sınırlı bulgular üzerine oluşturulmuş yeni yöntemlerin başarısı hakkında olduğu görülmektedir (Yaşar ve diğ., 2005). Bu nedenle, program değerlendirmenin etkili olarak gerçekleştirilebilmesi için tıpkı program geliştirme sürecine benzer süreçlerden geçirilmesi; plânlanması, uygulanması ve değerlendirilmesi gerekmektedir. Çünkü değerlendirmenin tasarlanmasından sonuçlandırılmasına kadar olan süreçte birbiriyle ilişkili pek çok öge ve işlem yer almaktadır. Bu işlemlerden bazıları değerlendirme sürecinde alınması gereken kararları yansıtırlar ve o nedenle de program değerlendirmenin temelini oluştururlar.

Son yıllarda eğitim programlarının değerlendirilmesine yönelik alanyazında oldukça fazla sayıda program değerlendirme çalışmalarının varlığı görülmektedir. Yaşar ve diğ. (2005) Türkiye’de yapılan program değerlendirme çalışmalarına ilişkin yaptıkları meta-değerlendirme araştırmasıyla şu çarpıcı sonuçlara ulaşmışlardır:

- Program değerlendirme çalışmaları genellikle programın bir boyutuna odaklanmıştır.
- Program değerlendirme çalışmalarının çoğu nicel yöntemlerle ve genellikle yalnızca anketler yoluyla yürütülmüştür.
- Program değerlendirme çalışmalarının program değerlendirme sürecini tamamen yansıtmadığı görülmüştür.

- Program değerlendirme çalışmalarının genellikle akademik kaygılarla derslerin etkililiğine yönelik ve okul, bölüm, fakülte veya il düzeyinde yürütülen çalışmalar niteliğinde olduğu saptanmıştır.
- Yapılan çalışmaların bulgularının genellikle yetersiz olduğu görülmüştür.

Türkiye’de program değerlendirme çalışmalarının genel görünümüne ilişkin bu çalışmanın sonuçları gerek bir ya da birkaç konu alanı uzmanı tarafından gerekse de disiplinler arası uzmanlar ve konu alanı uzmanlarının oluşturduğu geniş katılımlı bir grup tarafından gerçekleştirilsin, yürütülen program değerlendirme çalışmalarının, değerlendirmenin temel öğeleri olan planlama, uygulama ve değerlendirmeye uygun olarak birtakım ilke ve ölçütler doğrultusunda yapılmasının zorunlu olduğunu ortaya koymaktadır. Program değerlendirme çalışmalarında uyulması gereken bu ilke ve ölçütler genel olarak uzman kurumlarca geliştirilen standartlara dayanmaktadır. Bu açıdan Türkiye’de de gerçekleştirilen program değerlendirme amaçlı çalışmaların program değerlendirme standartlarına uygun olarak yürütülmesi gerekmektedir. Özellikle geniş kapsamlı ve katılımlı bir anlayışla ele alınan, merkezden hazırlanan ve eğitim sistemine önemli yenilikler getirdiği vurgulanan çalışmaların program değerlendirme standartlarını karşılayıp karşılamadığının bilinmesi, bu çalışmaları kullanacak ve buna dayalı karar verecek kesime değerlendirme hakkında oldukça sağlam kanıtlar sunma olanağı verecektir. Yine değerlendirme araştırmalarının standartlara uygun olup olmadığının belirlenmesi, bu değerlendirmelerin geçerliliği ve güvenilirliği hakkında ilgililere yardımcı olacaktır. Ancak Türkiye’de bu amaca hizmet edebilecek standartların olmayışı büyük bir eksiklik olarak karşımıza çıkmaktadır. Bu eksikliğin giderilmesi için ilk olarak ABD’de Ortak Komite tarafından geliştirilmiş olan ve daha sonra İsviçre, Almanya, Avustralya, Kanada ve Kenya (Afrika ülkelerinde kullanılmak üzere) tarafından kendi ülkelerine uyarlanan program değerlendirme standartları temel alınarak Türkiye’deki program değerlendirme çalışmaları için kullanılacak “Program Değerlendirme Standartları”nın oluşturulmasına gereksinim vardır. Bu araştırma bu gereksinimi karşılamaya dönük olarak hazırlanmıştır.

1.7. Araştırmanın Amacı

Bu çalışmanın genel amacı, ABD’nde Ortak Komite tarafından geliştirilmiş ve kimi ülkelerde aynen benimsenmiş ve kimi ülkelerde de ülkeye göre uyarlanmış program

değerlendirme standartlarına dayalı olarak Türkiye için alan uzmanlarının görüşlerine dayanılarak program değerlendirme standartları oluşturmaktır. Bu temel amaç kapsamında araştırmada şu sorulara yanıt aranmıştır:

1. Alan uzmanlarının program değerlendirmede “*Yararlık*” standartlarının göstergelerinin uygunluğuna ilişkin görüşleri nelerdir?
2. Alan uzmanlarının program değerlendirmede “*Yürütülebilirlik*” standartlarının göstergelerinin uygunluğuna ilişkin görüşleri nelerdir?
3. Alan uzmanlarının program değerlendirmede “*Uygunluk*” standartlarının göstergelerinin uygunluğuna ilişkin görüşleri nelerdir?
4. Alan uzmanlarının program değerlendirmede “*Doğruluk*” standartlarının göstergelerinin uygunluğuna ilişkin görüşleri nelerdir?
5. Alan uzmanlarının görüşlerine göre belirlenen program değerlendirme standartları uygunluk düzeyi *standart alanlarına* göre anlamlı farklılık göstermekte midir?
6. Alan uzmanlarının program değerlendirme standartlarına ilişkin önerileri nelerdir?
 - a. “*Yararlık*” standartları ve göstergelerine ilişkin önerileri nelerdir?
 - b. “*Yürütülebilirlik*” standartları ve göstergelerine ilişkin önerileri nelerdir?
 - c. “*Uygunluk*” standartları ve göstergelerine ilişkin önerileri nelerdir?
 - d. “*Doğruluk*” standartları ve göstergelerine ilişkin önerileri nelerdir?

1.8. Araştırmanın Önemi

Program değerlendirme çalışmalarının kullanıcılar ve karar vericiler yönünden geçerli ve güvenilir bilgiler sağlayabilmesi için daha geniş kapsamda ve planlama, uygulama ve değerlendirme aşamalarında belirli ölçütlere göre yürütülmesi gereklidir. Oysa Türkiye’de yürütülen program değerlendirme çalışmalarında bu tür ölçütlerin eksikliğiyle karşılaşılmaktadır. Bu araştırma ile Türkiye için geliştirilecek olan program değerlendirme standartlarının, gerek program değerlendirme çalışması yapacak araştırmacılara yol göstermede gerekse de tamamlanmış bir program değerlendirme çalışmasını değerlendirmede ya da değerlendirmenin sonuçlarını kullanacak ve buna

dayalı karar verecek yetkililere araştırmanın geçerliği ve güvenilirliği hakkında sağlam bilgi sunmada yardımcı olması beklenmektedir. Ayrıca bu çalışma, diğer ülkelerde geliştirilen değerlendirme standartlarından farklı olarak standart göstergelerine sahip olması yönüyle ayrıca bir öneme sahiptir. Bu göstergeler, program değerlendirmeyi yapacak araştırmacılar için yönlendirici ipuçları ve ilkeler niteliğindedir. Yine bu araştırma, alanda çalışan araştırmacıların yapabilecekleri ileri araştırmalara yön vermede ve zaman içerisinde güncelleştirilebilecek program değerlendirme standartlarına temel olması bakımından önem taşımaktadır.

1.9. Sayıtlar

Bu araştırma şu temel sayıtlara dayanmaktadır:

- Bu araştırma için eğitim fakültelerinde görevli eğitim programları ve öğretim uzmanlarına yönelik olarak geliştirilen “Türkiye İçin Program Değerlendirme Standartları Oluşturma Çalışması Anketi”nin uzman görüşüne dayalı olarak belirlenen kapsam geçerliği yeterli düzeydedir.
- Türkiye’de program değerlendirme standartların oluşturulması eğitim programları ve öğretim alanında doktorasını yapmış uzmanların görüşlerine dayalı olarak belirlenebilir.
- Araştırmaya katılan uzmanlar, “Türkiye İçin Program Değerlendirme Standartları Oluşturma Çalışması Anketi”ndeki soruları kendi görüşlerini doğru olarak yansıttıkları biçimde yanıtlamışlardır.

1.10. Araştırmanın Sınırlılıkları

Bu araştırma, doktorasını eğitim programları ve öğretim alanında yapmış ve 2008–2009 öğretim yılında 94 devlet üniversitesine bağlı 64 eğitim fakültesinde görev yapan alan uzmanlarının görüşleriyle sınırlıdır.

1.11. Tanımlar

Bu başlık altında yer alan kavramlar araştırmadaki kullanım amaçlarına uygun olarak tanımlanmışlardır.

Değerlendirme: Bir ölçme işlemi sonucunda elde edilen sonuçların bir ölçütle karşılaştırılması ve buna bağlı olarak bir değer yargısına ulaşma işlemidir (Erden, 1998)

Eğitsel Değerlendirme: Okul ile ilgili tüm değişkenlerin birlikte değerlendirilmesini temel alan değerlendirme türüdür (Oliva, 1988).

Öğretimsel Değerlendirme: Öğrencilerin değerlendirilmesini, öğreticinin performansını ve öğretim sürecindeki belli bir uygulamanın, yaklaşımın ya da yöntemin etkililiğinin değerlendirilmesini içeren değerlendirmedir (Oliva, 1988).

Program Değerlendirme: Bir eğitim programının değeri ve etkililiğinin sistematik araştırılması sürecidir (JC, 1994).

Değerlendirici: Bir program değerlendirme çalışmasını planlayabilecek, yürütebilecek ve değerlendirebilecek yeterliklere sahip kişi (JC, 1994).

Paydaş: Program değerlendirmeye katılan ya da program değerlendirme sonuçlarından etkilenen birey ya da gruplar (JC, 1994).

Standart: Bir otorite, gelenek ya da bir ortak anlayış tarafından belirlenmiş, izlenmesi gereken model ya da örneklerdir (Richardson, 1994).

Gösterge: Belli bir durumun var olduğunu yada belli sonuçların kazanılıp kazanılmadığını ortaya koymaya dönüştürülebilir ifadelerdir (Conley, 1997). Araştırmada göstergeler standartların gerçekleşip gerçekleşmediğinin kontrol edilmesinde kullanılacak ifadeler olarak tanımlanmıştır.

Program Değerlendirme Standartları: Program değerlendirme çalışması yapacak araştırmacılara yol göstermede, yapılan bir program değerlendirme çalışmasını değerlendirmede ya da değerlendirmenin sonuçlarını kullanacak ve buna dayalı karar verecek yetkililere araştırmanın geçerliği ve güvenilirliği hakkında sağlam bilgi sunmada kullanılacak ölçütlerdir (Sağlam ve Yüksel, 2007).

Yararlık Standartları: Program değerlendirme çalışmalarının, belirlenmiş değerlendirme amaçları ve ilgililerin bilgi gereksinimleri doğrultusunda yürütülmesini amaçlayan standartlardır (JC, 1994).

Yürütülebilirlik Standartları: Program değerlendirme çalışmasının mantıklı, titiz, ustaca ve uygun maliyetli bir biçimde planladığının ve yürütüldüğünün güvence altına alınmasını amaçlayan standartlardır (JC, 1994).

Uygunluk Stadnartları: Program değerlendirme çalışmalarının uygun, adil, yasal, doğru, kabul edilebilir olmasını amaçlayan standartlardır (JC, 1994).

Doğruluk Stadnartları: Program değerlendirmede doğruluğu destekleyen süreçlerin, yöntemlerin ve desenlerin nasıl uygulanabileceğini ve değerlendirme bulgu ve sonuçlarında doğruluğu azaltıcı ve zayıflatıcı tutarsızlıkların, çarpıtmaların ve yanlış anlamaların nasıl giderilebileceğini amaçlayan standartlardır (JC, 1994).

Alan Uzmanı: Bir bilim dalında, belli bir teknik alanda yoğun bir öğrenim görerek ya da türlü uygulamalara katılarak yüksek derecede yeterlik kazanmış olan kimse (Araştırmada doktorasını tamamlamış öğretim üyeleri alan uzmanı olarak tanımlanmıştır)

1.12. Kısaltmalar

Bu araştırmada kullanılan önemli kısaltmaların açıklamaları ve Türkçe karşılıkları aşağıda verilmiştir.

AERA: American Educational Research Association (Amerikan Eğitsel Araştırmalar Derneği)

AES: The Australasian Evaluation Society (Avustralya Değerlendirme Kurulu)

CSA: The Canadian Standards Association (Kanada Standartlar Birliği)

ÇGSÜ: Context-Input-Process-Product (Çevre-Girdi-Süreç-Ürün)

DeGEVAL: Gesellschaft für Evaluation (Alman Değerlendirme Derneği)

ISO: International Organization for Standardization (Uluslararası Standartlaştırma Organizasyonu)

JC: The Joint Committee (Ortak Komite)

NCME: National Council on Measurement in Education (Eđitimde Deđerlendirme Ulusal Kurulu)

NISO: National Information Standards Organization (Ulusal Bilgi Standartları Organizasyonu)

OKPDS: Ortak Komite Program Deđerlendirme Standartları

SEVAL: Swiss Evaluation Society (İsviçre Deđerlendirme Derneđi)

SFE: Société Française D'évaluation (Fransa Deđerlendirme Derneđi)

2. İLGİLİ ARAŞTIRMALAR

1970'li yıllarda Amerika Birleşik Devletleri'nde başlayan eğitimde program değerlendirme standartları geliştirme çalışmaları günümüzde birçok ülkede de kapsamlı olarak sürdürülmektedir. Birçok ülke kendi kültürüne özgü program değerlendirme standartları, ilkeleri ve yönergeleri geliştirmekte veya başta Amerika Birleşik Devletleri olmak üzere diğer ülkelerdeki çalışmaları kendilerine uyarlamakta ve zaman içinde de geliştirilen veya uyarlanan program değerlendirme standartlarını geliştirmeler ve değişen koşullar ile bu konuda yapılan araştırmaların sonuçları doğrultusunda geliştirmektedir. Türkiye'de de eğitim programlarının değerlendirilmesine dönük, özellikle programın bir boyutuna odaklanılarak çeşitli araştırmalar yapılmıştır. Ancak Türkiye'de yurtdışındaki örneklerinde olduğu gibi eğitimde program değerlendirme standartları, ilkeleri veya yönergelerinin belirlenmesi amacıyla yapılmış her hangi bir çalışmaya rastlanamamıştır.

Araştırmanın bu bölümünde, araştırmanın kuramsal çerçevesinin oluşturulması amacıyla yapılan alanyazın incelemesi çalışmaları sırasında ulaşılan eğitimde program değerlendirme standartları ile ilgili araştırma bulgularına yer verilmiştir. Bu araştırmalar, yurtdışında ve Türkiye'de gerçekleştirilenler olmak üzere iki grupta ele alınmıştır.

2.1. Yurtdışında Yapılan Araştırmalar

Chatterji'nin (2005) "*OKPDS 1994 Yılı Standartlarının Uluslararası Ortamda Uygulanması: Bangladeş'te Bir Eğitsel Değerlendirme Örneği*" adlı çalışmasında, Ortak Komite program değerlendirme standartlarının geliştirmekte olan ülkelerdeki program değerlendirme çalışmaları için uygulanabilir olup olmadığı ve program değerlendirme çalışmalarının değerlendirilmesi için izlenebilecek bir yönergenin geliştirilmesi amaçlanmıştır. Bu amaçlar doğrultusunda Bangladeş'te yayınlanan yedi program değerlendirme raporu, iki kişiyle yapılan görüşmeler ve bir örnek olay, program değerlendirme standartlarına uygun olarak değerlendirilmiştir. Bu değerlendirmeler araştırmacı tarafından OKPDS temel alınarak hazırlanan iki kontrol listesi ile yapılmıştır. Araştırmada şu sonuçlara ulaşılmıştır:

- İncelenen çalışmaların “Yararlık” standartlarını büyük ölçüde karşıladığı görülmüştür.
- İncelenen çalışmaların “Yürütülebilirlik” standartlarını karşıladığı görülmüştür.
- İncelenen çalışmaların “Uygunluk” standartlarının temel özelliklerini karşılama yönünden yeterli olmadığı sonucuna ulaşılmıştır.
- İncelenen çalışmaların araştırma yöntem ve tasarım yönünden sağlam olması bu çalışmaların “Doğruluk” standartlarını yeterli düzeyde karşıladığı göstermiştir.

Sonuç olarak, varsayımları dikkate alındığında program değerlendirme standartlarının Bangladeş gibi yeni gelişen ülkelerde de uygulanabilir olduğu görülmüştür.

Torres (2005) “*İkinci Dil Dersinin Değerlendirmesinin Değerlendirmesi*” adlı yüksek lisans tezi çalışmasında, orta düzey ikinci dil dersinin değerlendirilmesini ve bu değerlendirilmenin OKPDS yönünden meta-değerlendirmesini amaçlamıştır. Torres çalışmasında, öncelikle ikinci dil dersinin öğretim araç-gereçlerini, derste kullanılan yöntemleri ve teknikleri ve uygulamaları değerlendirmiştir. Bu bağlamda Torres programın güçlü ve zayıf yönlerini ortaya çıkarmayı amaçlamış ve sonuçların program geliştirme sürecinde kullanılabilirliğini incelemiştir. Araştırmanın ilk bölümünde yazar programın öğrencileri ikinci dildeki üst düzey derslere ve sınavlara ne kadar iyi hazırladığı sorularına yanıt aramıştır. Kısaca, programın etkililiğine baktığı bu bölümde yazar, verileri odak grup görüşmeleri yoluyla toplamıştır. Toplanan verilerin çözümlenmeleri nitel veri çözümlenme yöntemleriyle gerçekleştirilmiştir. Çözümlenme sonrası sonuçlar iki farklı rapor halinde hazırlanmıştır. Yazarın araştırmanın birinci bölümünde ulaştığı sonuçlar şunlardır:

- Programın içeriği oldukça yoğundur.
- Program öğrencilerin dil yeterliklerini geliştirmemektedir.
- Öğreticiler program izlencesini düzenli olarak takip etmemektedirler.
- Öğreticiler öğrencilere zamanında dönüt vermemektedirler.
- Ders kitapları yalnızca ana dil bilenlere yönelik olarak hazırlanmıştır.

- Program ile sınavlar arasında bir ilişki yoktur. Sınavlara dönük bir içeriğe programda yer verilmemektedir.

Araştırmanın ikinci bölümünde yazar, değerlendirdiği ve raporlaştırdığı araştırmasının program değerlendirme standartlarına uygunluğuna bakmıştır. Sonuçlar, değerlendirici ile paydaşların iletişiminin yeterli olmadığını, bu nedenle değerlendiricinin değerlerini ve değerlendirmeye bakış açısını paydaşlara yansıtamadığını; değerlendirmenin politik destekten yoksun olduğunu ve maliyet etkililiğinin yeterli olmadığını göstermiştir.

Jang (2000) "*Batı Dışı Ülkelerde OKPDS'nin Uygunluğu: Güney Kore Örneği*" adlı çalışmada, OKPDS'nin batı dışındaki ülkelerde uygunluğunu belirlemeyi amaçlamıştır. Güney Kore'de gerçekleştirilen bu araştırmanın amacı, Kore değerlendirme çalışma ve projelerinde uygulanabilecek bir öneri çerçevesi oluşturmak ve Kore'de çalışma yapacak batılı değerlendiricilere Kore değerlendirme kültürü hakkında bilgi vermektir. Bu amaç kapsamında standartların Kore kültürüne uygunluğunu tartışmak için bir program değerlendirme çalışmasına ya da projesine katılmış 21 mezun öğrenci, 11 üst düzey Kore kamu hizmetlisi ve yedi şirket müdürünün katılımıyla bir günlük çalıştay düzenlenmiştir. Çalıştayda standartlar uzmanlara tanıtıldıktan sonra standartlar temel alınarak hazırlanan beşli likert türünde bir anketle uzmanların standartların uygunluğuna ilişkin görüşleri toplanmış ve çalıştay sonunda betimsel istatistik kullanılarak veriler çözümlenmiş ve katılımcılara hemen sunulmuştur. Bulgular tekrar tartışılmıştır. Araştırmada ulaşılan sonuçlar şunlardır:

- Katılımcıların çoğu OKPDS'nin Kore değerleriyle çatıştığını belirtmiştir.
- "*Yürütülebilirlik*" ve "*Doğruluk*" standartlarının diğer standartlara göre Kore kültürüne daha yakın olduğu ve kabul edilebilir olduğu belirtilmiştir.
- "*Uygunluk*" standartlarının hiçbirisi Kore kültürü için *yeterli düzeyde uygun* bulunmamıştır.
- "Paydaşların tanımlanması", "duyarlı ve kapsamlı yönlendirme", "resmi sözleşme", "birey etkileşimleri", "insan hakları ve saygı", "bulguların açıklanması", "çıkar çatışmaları", "finansal sorumluluk", "savunulabilir bilgi kaynakları" ve "meta-değerlendirme" standartları katılımcılar tarafından Kore kültürüne *uygun* bulunmamıştır

- Dört standart *orta düzeyde uygun* bulunmuştur. Bu standartlar şunlardır: “değerlerin belirlenmesi”, “maliyet etkinliği”, “eksiksiz ve dürüst değerlendirme” ve “programın tanımlanması” standartları katılımcılar tarafından Kore kültürüne *orta düzeyde uygun* bulunmuştur.

Sonuçlara dayanarak yazar, OKPDS'nin Kore kültürüne uyarlanmasını ya da Kore kültürü dikkate alınarak yeni standartların geliştirilmesini önermiştir.

Mitchell (1998), “*Teksas Eğitim Ajansı Bölge Etkililik ve Uyum Değerlendirme Modelinin (District Effectiveness and Compliance Review-DEC) Değerlendirilmesi*” adlı doktora tez çalışmasında, Teksas Eğitim Ajansı tarafından kamu okullarının değerlendirilmesinde kullanılan DEC modelinin süreç yönergesini ve bu modelde gerçekleştirilen altı değerlendirme raporunu değerlendirmeyi ve bölgedeki 77 öğretim elemanının görüşlerine dayalı olarak bu standartların DEC yönergesiyle ne düzeyde bir uygunluk içinde olduğunu belirlemeyi amaçlamıştır. Bu amaç kapsamında yönerge ve raporlar OKPDS temel alınarak hazırlanan bir veri toplama aracıyla değerlendirilmiştir. OKPDS ile DEC arasındaki uygunluğun belirlenmesinde ise katılımcılarla görüşmeler yapılmıştır. Araştırmada veriler, betimleyici istatistik teknikleri kullanılarak çözümlenmiştir. Araştırmada şu sonuçlara ulaşılmıştır:

- DEC modelinin yönergesi, 30 standardın 17'sini “mükemmel” düzeyde, yedisini “iyi” düzeyde ve üçünü “kabul edilebilir” düzeyde karşılamıştır. “maliyet etkinliği”, “mali sorumluluk” ve “meta-değerlendirme” standartları “uygulanamaz” nitelikte görülmüş ve değerlendirmeye alınmamıştır.
- Altı DEC değerlendirme raporunun her biri, “raporun zamanı ve yayınlanması” standardı dışında tüm değerlendirme standartlarını “mükemmel”, “iyi” ve “kabul edilebilir” düzeyde karşılamıştır.
- Katılımcılar, “maliyet etkinliği”, “mali sorumluluk”, “resmi sözleşme” ve “meta-değerlendirme” standartlarının DEC ile uyumlu olmadığı, diğer tüm standartların “mükemmel”, “iyi” ve “kabul edilebilir” düzeyde olduğu görüşündedirler.

Giraffe (1997), “*Hükümet Destekli Yetişkin Eğitimi Programlarına Uygulanan Değerlendirme Standartlarının Meta-Değerlendirmesi*” adlı doktora tez çalışmasında,

yetişkin eğitim programlarından bilgisayar programcılığı programının OKPDS, öğrenci, yönetici, öğretmen, destek hizmetleri yetkilileriyle yapılan görüşmeler yoluyla değerlendirmesini amaçlamıştır. Çalışmanın giriş bölümünde, ABD hükümetinin devletten yardım alan işçilerin sayısını azaltma amacıyla yetişkin mesleki eğitimini finansal yönden destekleyeceği, ancak bu desteğin kesinleştirilmesinin programlarının değerinin ve etkinliğinin ortaya konulmasına bağlı olduğu vurgulanmıştır. Bunun için de programların değerlendirilmesi gereği açıklanmıştır. Bu temel gerekçeden hareketle araştırmada yetişkin eğitim programlarından bilgisayar programcılığı programının kuramsal yapısının ve uygulamasının değerlendirilmesi amaçlanmıştır. Programın kuramsal yapısı OKPDS temel alınarak hazırlanmış bir kontrol listesi ile uygulama süreci ise yönetici, öğrenci ve öğretmenlerden oluşan 49 program paydaşıyla yapılan görüşmeler yolu ile değerlendirilmiştir. Araştırmada şu sonuçlara ulaşılmıştır:

- Bilgisayar programcılığı programı kuramsal yönden;
 - “*Yararlılık*” standartlarından yalnızca “değerlendirici yeterliği”, “raporun zamanlılığı ve yayımı” ve “değerlendirme etkisi” standartlarını karşılamıştır.
 - “*Yürütülebilirlik*” standartlarından ikisini karşılamış, birini karşılamamıştır. Programın karşıladığı standartlar “uygulanabilir süreçler” ve “maliyet etkinliği” standartlarıdır
 - “*Uygunluk*” standartlarından “resmi sözleşme”, “insan hakları”, “bulguların açıklığı”, “tam ve dürüst değerlendirme”, “çıkar çatışması” ve “mali sorumluluk” standartlarını karşılamış diğer standartlar uygulama dışı bırakılmıştır.
 - “*Doğruluk*” standartlarından “amaç ve işlemlerin tanımlanması”, “savunulabilir bilgi kaynakları”, “güvenilir bilgiler”, “sistemik bilgiler” ve “tarafsız raporlaştırma” standartlarını karşılamış diğer yedi standartı karşılamamıştır.
- Bilgisayar programcılığı programı;
 - öğrenci görüşlerine göre içerik ve program süresi yönünden yeterli bulunmamıştır.
 - öğretmen görüşlerine göre gereksinimleri karşılamada yeterli bulunmamıştır.
 - yönetici görüşlerine göre gereksinimleri karşılamaktadır, eğitsel ve sosyo-

ekonomik yönden değerlidir ve programa olan mali yatırım devam etmelidir.

Maher (1997), “Devlet Okulunda Gelişimsel Engelli Gençlere Yönelik Eğitim Programının Değerlendirilmesi” başlıklı çalışmada, devlet okullarında engelli öğrencilerin sayılarının artmasının eğitimcileri ve okul yöneticilerini bu öğrencilere yönelik program tasarlamaya ve uygulamaya yönelttiğini vurgulamakta ve bu programların sağlamlığının ve etkinliğinin belirlenmesi için de program değerlendirmeye özel bir önem verilmesi gerektiğini belirtmektedir. Ancak Maher, özel eğitim programlarının değerlendirilmesine ilişkin çok az sayıda çalışma olduğunu ve bunların yetersiz sayılabileceğini ifade etmektedir. Maher, bu çalışmada devlet okulunda gelişimsel engelli çocuklara yönelik uygulanan davranış temelli bir programın değerlendirilmesinin planlanması ve yürütülmesi sürecini araştırmıştır. Değerlendirmenin planlama ve yürütme süreci, OKPDS dikkate alarak gerçekleştirilmiştir. Sistem yaklaşımı çerçevesinde değerlendirilen programla ilgili veriler okul bölgesindeki tüm paydaşlardan elde edilmiştir. Çalışmada şu sonuçlara ulaşılmıştır:

- Program öğrencilerin eğitsel ve yaşamsal gereksinimlerini karşılamaktadır.
- Öğrenciler program kazanımları dikkate alındığında beklentilerin üzerinde ilerleme kaydetmiştir.
- Programda öğrencilerin gelişimlerine uygun öğretim yöntemleri ve süreçleri kullanılmıştır
- Paydaşlar programın bu tür öğrenciler için yararlı ve olumlu olduğunu düşünmektedirler.

Ayrıca araştırmada program değerlendirme planı ve süreci program değerlendirme standartları çerçevesinde değerlendirilmiş ve şu sonuçlara ulaşılmıştır:

- Devlet okullarında özel eğitim program değerlendirmeleri kolaylıkla uygulanabilir.
- Bu çalışmadan elde edilen sonuçlar karar vericiler için önemli öneri niteliği taşır.

- Program değerlendirme süreci yasal yükümlülükler ve etik kurallar bağlamında uygundur.

Lee (1995) “*Tayvan Eğitim Bakanlığı Ulusal Teknoloji Enstitülerinin Meta-değerlendirmesi: Değerlendirme Takımı ve Paydaşlarının Değerlendirme Etkinlikleri ile İlgili Kararları Üzerine Bir Çalışma*” adlı doktora tez çalışmasında, Tayvan Eğitim Bakanlığı Ulusal Teknoloji Enstitülerinin değerlendirme etkinliklerinin bir meta-değerlendirmesini yapmıştır. Yazar bu meta-değerlendirme çalışmasında değerlendirme etkinliklerini, değerlendirme ekibi ve paydaş görüşlerine dayalı olarak yararlık, yürütülebilirlik, uygunluk ve doğruluk yönünden incelemiştir. Değerlendiriciler ve paydaşlardan ayrıca değerlendirme standartlarının önemi ile ilgili görüşlerini de belirtmeleri istenmiştir. Araştırmada veriler, 117 değerlendirici ve 303 paydaştan e-posta yoluyla gönderilen anketler ve görüşmeler yoluyla elde edilmiştir. Araştırmada şu sonuçlara ulaşılmıştır:

- Değerlendiriciler ve paydaşlar, 30 program değerlendirme standardının Tayvan’daki program değerlendirme çalışmalarının süreci ve uygulaması için değerli ve önemli ilkeler olduğu konusunda aynı görüşü paylaşmaktadırlar.
- Değerlendiriciler ve paydaşlar bir standart (paydaşların tanımlanması) dışında diğer tüm standartları önemli olarak kabul etmişlerdir.
- Tayvan Eğitim Bakanlığınca yürütülen değerlendirme etkinliklerinin “paydaşların tanımlanması”, “içerik analizi”, “nitel verilerin analizi”, “savunulabilir bilgi kaynakları” standartları dışında tüm değerlendirme standartlarını karşıladığı görülmüştür.
- Tayvan Eğitim Bakanlığınca yürütülen değerlendirme etkinliklerinde kullanılan değerlendirme sisteminin uygun olduğu değerlendiriciler ve paydaşlar tarafından onaylanmıştır.

Sonuçlar doğrultusunda şu öneriler geliştirilmiştir:

- Değerlendirme etkinlikleri değerlendirme modeli, politikası, süreçleri ve uygulamaları yönünden geliştirilmelidir.

- Değerlendirme bağımsız kuruluşlar tarafından yürütülmelidir
- Bakanlık öğrenciler, aileler, kuruluşlar vb. diğer tüm paydaşlara yönelik daha kapsamlı raporlar hazırlamalıdır.
- Bakanlık değerlendirici eğitimi için çalıştaylar, hizmetiçi eğitim programları vb. uygulamalara gitmelidir.

Mann (1991) “*Nebraska Liseleri Etkinlik Programları İçin Eğitim Standartları*” adlı doktora tez çalışmasında, Nebraska liselerinde öğrenci etkinlik programlarında kullanılan standart türlerinin belirlenmesini amaçlamıştır. Bu amaç kapsamında, araştırmacı OKPDS’nin etkinlik programlarının değerlendirilmesinde kullanımının ne düzeyde uygun olduğunu ve hangi standartların değerlendirmede kullanıldığını saptamak istemiştir. Araştırmada veriler 120 lise yöneticisinin etkinlik programlarının değerlendirilmesi ile ilgili ankete verdikleri yanıtlardan elde edilmiştir. Anket program değerlendirme standartlarına dayalı hazırlanmış ve anketten elde edilen veriler betimsel istatistik kullanılarak çözümlenmiştir. Araştırma bulguları katılımcıların, program değerlendirme standartlarının Nebraska liselerinin etkinlik programlarının değerlendirilmesi için uygun olduğu, öte yandan ancak değerlendirilen etkinlik programlarının standartları yeterince karşılamadığı yönünde görüş bildirdiklerini ortaya koymuştur.

Wahba (1990) “*Uygulamada Meta-değerlendirme- Dört Gelişen Ülkede Ebe Yetiştiriciliği Program Değerlendirme Raporunun OKPDS’ye Göre Değerlendirmesi*” adlı çalışmasında, Hindistan, Sierra Leone, Burma ve Gana ülkelerinde anne ölümlerinin oldukça sık yaşandığını ve bunun nedenleri arasında yeterli eğitim düzeyine sahip doğum ebesinin olmamasının yer aldığını belirterek ebelerin yetiştirilmesine dönük programların daha etkili bir biçimde değerlendirilmesinin gerekliliğini vurgulamıştır. Wahba çalışmasında, bu dört ülkede ebelik eğitimi için hazırlanan program değerlendirme raporlarının OKPDS’ye göre sağlamlığını; panel üyelerinin standartlara ilişkin görüşlerini ve program değerlendirme standartlarının sağlık alanında yararlılığını ve kullanılabilirliğini araştırmıştır. Bu amaç kapsamında araştırmacı ilgili ülkelerin eğitim komisyonlarınca hazırlanan değerlendirme raporlarını değerlendirmek için dört program değerlendirme uzmanının katılımıyla bir panel düzenlemiştir. Panelde uzmanlara değerlendirme standartlarına ilişkin kısa bir eğitimi verilmiş, daha sonra da

uzmanlara 1981 Yılı OKPDS'nin bir kontrol listesine dönüştürülmüş biçimi verilerek, onlardan ülke raporlarını bu kontrol listesine göre değerlendirmeleri istenmiştir. Araştırmada panele katılan uzmanların bu değerlendirmelerinden ulaşılan sonuçlar şunlardır:

- Uzmanların değerlendirmelerine göre raporlardan hiçbirisi 30 değerlendirme standardının tamamını karşılayamamıştır.
- Uzmanlar Sierra Leone ve Gana tarafından hazırlanan raporların 20 standardı; Hindistan tarafından hazırlanan raporun 15 standardı ve Burma tarafından hazırlanan raporun da yalnızca üç standardı karşıladığını bildirmişlerdir.
- Uzmanlara göre, raporlarda en çok “Yararlık” ve “Doğruluk” standartları karşılanmış; “Yürütülebilirlik” ve “Uygunluk” standartları ise yeterince karşılanamamıştır.
- Uzmanlar “Uygunluk” standartları dışındaki standartların çoğunun gelişen ülkeler için kullanılabilir olduğu yönünde görüş bildirmişlerdir.
- Uzmanlar meta-değerlendirmenin değerlendirme çalışmalarının teknik uygunluğu ve sağlamlığını değerlendirmede güvenilir bir yöntem olduğu yönünde görüş bildirmişlerdir.
- Uzmanlar OKPDS'nin değerlendirme çalışmalarının niteliğini değerlendirmede iyi bir araç olduğu yönünde görüş bildirmişlerdir.

Larsen'in (1985), “*Yönetici Yetiştirme Programını Değerlendirmede Kullanılan Deneysel Bir Yöntemin Uygulaması ve Meta-değerlendirmesi*” adlı çalışmasının amacı, yönetici yetiştirme programını, programın başarılı yönlerini ortaya çıkarmayı amaçlayan “Başarılı Olay Yöntemi” ile değerlendirme ve bu yöntemi “Uygunluk”, “Yürütülebilirlik”, “Yararlık” ve “Doğruluk” standartları yönünden deneme ve değerlendirmedir. Araştırmada program değerlendirmenin yetiştirme programlarında sık kullanıldığını ancak, değerlendirmelerin çoğunluğunun uygulanabilir ve ekonomik olmadığı vurgulanmıştır. Bu nedenle araştırmacı, gerek ekonomik gerekse de uygulaması basit olan “Başarılı Olay Yöntemi”nin programın değerlendirmede etkili olup olmadığını saptamaya çalışmıştır. Araştırmada öncelikle bir kurumda uygulanan yönetici yetiştirme programı “Başarılı Olay Yöntemi” ile uygulanmış ve program

sonunda programa katılan 37 katılımcıdan bu yöntemle verilen eğitimi değerlendirmeleri istenmiştir. Araştırmanın bu bölümünün verileri katılımcılarla yapılan yapılandırılmış görüşmelerle toplanmıştır. Toplanan veriler çözümlendikten sonra yöntemin etkililiğine ilişkin bir rapor hazırlanmış ve yöntemin OKPDS'ye uygunluğu incelenmiştir. Araştırmada şu sonuçlara ulaşılmıştır:

- Katılımcılar “Başarılı Olay Yöntemi”nin yönetici yetiştirme programında başarılı olduğunu, program değerlendirmede yürütülebilir, ekonomik, doğru ve güvenilir bilgi sağlayan bir kaynak olduğunu, ancak program değerlendirmede yalnızca bu yöntemin kullanılmasının yeterli olmadığını düşünmektedirler.
- “Başarılı Olay Yöntemi”, “Uygunluk”, “Yürütülebilirlik”, “Yararlık” ve “Doğruluk” standartlarını yeterli düzeyde karşılamaktadır.

Johnson (1983), “*Guilford Okul Sistemi Okuma Programının Değerlendirilmesi*” adlı çalışmasında değerlendirmenin program geliştirme sürecinin önemli bir bölümünü oluşturduğu, yöneticiler, öğretmenler ve aileler ve öğrencilerin değerlendirme sürecinde yer alması gerektiğini ve meta-değerlendirme çalışmalarının değerlendirme çalışmalarının nesnelliği, güvenilirliğini ve geçerliliğini geliştirmeye dönük olduğunu vurgulayarak Guilford Okul Sistemi okuma programının etkililiğini belirlemeyi amaçlamıştır. Bu amaç kapsamında öncelikle okuma programını değerlendirerek etkililiğini saptamış, daha sonra da programın niteliğini etkileyen etmenlerin programdan ya da program dışı etkenlerden kaynaklanıp kaynaklanmadığını belirlemiştir. Programın etkilerini belirlemek için yazar öncelikle programı 1979 ve 1983 yıllarında iki kez yönetici, öğretmen, ailelerin ve öğrencilerin görüşlerine göre değerlendirmiştir. Yazar bu bağlamda katılımcılara iki kez anket uygulamış ve katılımcılardan programı etkileyen etmenleri belirtmeleri istenmiştir. Değerlendirmede ayrıca öğrencilerin üçüncü ve altıncı sınıf okuma programına ilişkin başarı puanları da dikkate alınmıştır. Değerlendirme sonucunda yazar katılımcılar tarafından saptanan programı etkileyen etmenlerin kaynağını araştırmıştır. Yazar bunun için meta-değerlendirme yöntemini kullanmıştır. Araştırmada şu sonuçlara ulaşılmıştır:

- Yönetici ve öğretmen görüşlerine göre, 1983 programında önceki programa göre daha fazla okuma etkinliğine, öğrenci katılımına, öğretmen esnekliğine yer verilmiştir.
- 1983 programı, önceki programa göre katılımcı beklentilerini ve gereksinimleri daha fazla karşılamaktadır.
- 1979 ve 1983 yılı programlarından sonra öğrencilere uygulanan okuma testleri sonuçları karşılaştırıldığında, öğrencilerin 1983 yılı programından sonra katıldıkları testlerden daha iyi puan aldıkları görülmüştür. Bu durum 1983 programının öğrenci başarısını önceki programa göre arttırdığını göstermiştir.

Martin (1981), “*İki Dilli Eğitimin Değerlendirilmesi-Title VII Projesinin Meta-değerlendirmesi*” adlı yüksek lisans tezinde, 1978-1980 yıllarında Kaliforniya bölgesinde uygulanan ve Title VII projesi kapsamında yer alan iki dilli eğitim programlarına ilişkin bir meta-değerlendirme çalışması yürütmüştür. Yazar meta değerlendirme çalışmasıyla, değerlendirme planlarını, ABD Washington Eğitim Bölümüne sunulmuş değerlendirme raporlarını ve bu raporlarda yer alan değerlendirme etkinliklerinin değerlendirilmesini gerçekleştirmiştir. Araştırma kapsamında yazar ilk olarak değerlendirme planlarını incelemiştir. Bu kapsamda 43 değerlendirme planı incelenmiş ve şu sonuçlara ulaşılmıştır:

- Planlarda önerilen değerlendirme yaklaşımlarının daha çok öğretimsel amaçları, öğrenci başarısı ve program öğelerini dikkate aldığı görülmüştür.
- Değerlendirmede kullanılacak veri toplama araçlarının öğretim elemanlarına, ailelere uygulanabilecek anketler olabileceği vurgulanmıştır.
- Planların değerlendirme etkinliklerinin zamanının belirtilmesi, değerlendirme kaynaklarının tanımlanması, değerlendirme maliyetinin kesinleştirilmesi vb. boyutlarının yetersiz olduğu görülmüştür.

Araştırma kapsamında ikinci olarak değerlendirme raporları incelenmiştir. Bu kapsamda 47 proje raporu değerlendirme desenleri, modelleri, veri toplama teknikleri, örnekleme teknikleri vb yönlerden değerlendirilmiştir. Meta-değerlendirme sonucu ulaşılan sonuçlar şunlardır:

- İncelenen raporların amaca dayalı program değerlendirme modellerini temel aldıkları görülmektedir.
- Raporlarda değerlendirmede öntest-sontest kontrol gruplu deneysel desenlerin daha sık kullanıldığı görülmüştür.
- Raporlarda kullanılan istatistiksel işlemlerin yetersiz olduğu, değerlendirmelerde korelasyon, ki-kare, varyans analizi gibi tekniklerin nerdeyse hiç kullanılmadığı, daha çok betimsel istatistiklerin kullanıldığı, ancak bunların yetersiz olduğu ortaya çıkmıştır.
- Raporlarda yine evren ve örneklem bölümlerinin yeterince ayrıntılı ele alınmadığı, örneklem seçimlerinin nasıl yapıldığına ilişkin çok fazla bilginin olmadığı da varılan bulgular arasındadır.
- Son olarak raporlarda varılan sonuçların ve önerilerin daha iyi yapılandırılabilmesinin de altı çizilmiştir.

Martin, bu meta-değerlendirme çalışmasının sonuçlarına göre, yapılan çalışmaların genel olarak değerlendirme desenleri ve raporlaştırma konusunda yetersiz olduğunu, geliştirilmesi gerektiğini ve bu haliyle veri tabanlarında yer almaması gerektiğini ifade etmektedir. Martin ayrıca çalışmalarda veri toplama araçlarının ve gerekli istatistiksel işlemlerin daha ayrıntılı ele alınması gerektiğini, uygulama etkinliklerinin daha iyi planlanması gerektiğini de vurgulamaktadır.

2. 2. Yurtiçinde Yapılan Araştırmalar

Yurtiçinde program değerlendirme standartlarına dönük yalnızca bir çalışmaya rastlanmıştır.

Yaşar ve diğerleri (2005), "*Türkiye'de İlköğretim Öğretmeni Yetiştirme Programlarının Meta-Değerlendirmesi*" adlı çalışmalarında, 1997-2004 yılları arasında sınıf öğretmeni yetiştirme programları ile ilgili yapılmış program değerlendirme çalışmalarının değerlendirilmesi kapsamında sekiz program değerlendirme çalışmasını incelemişlerdir. İncelenen çalışmaların değerlendirilmesinde, OKPDS temel alınmıştır. Ancak, araştırmacılar bu 30 standartın 12'sinin Türkiye ortamında uygulanabilir olmadığına karar vererek değerlendirmeyi 18 standart üzerinden yapmışlardır. Değerlendirmede kullanılacak standartlar araştırmacılar tarafından bir kontrol listesine dönüştürülerek,

belirlenen çalışmaların yararlık, yürütülebilirlik, uygunluk ve doğruluk yönünden standartları ne düzeyde karşıladığına bakılmıştır. Standartların karşılanma düzeyleri, “yeterli”, “kısmen yeterli” ve “yetersiz” olarak belirlenmiştir. Araştırmada şu sonuçlara ulaşılmıştır:

- İncelenen değerlendirme çalışmaları “*Yararlık*” standartlarını “yeterli” düzeyde, “*Yürütülebilirlik*”, “*Uygunluk*” ve “*Doğruluk*” standartlarını ise “kısmen yeterli” düzeyde karşılamaktadırlar.
- Değerlendirme çalışmalarının “değerlerin belirlenmesi”, “güvenirli bilgiler” ve “programın tanımlanması” standartlarını karşılamada “yetersiz”dir.
- Program değerlendirme çalışmaları genellikle programın bir boyutuna odaklanmışlardır.
- Program değerlendirme çalışmalarının çoğu nicel yöntemler ve genellikle yalnızca anketler yoluyla yürütülmüştür.
- Program değerlendirme çalışmaları program değerlendirme sürecini tamamen yansıtmamaktadır.
- Program değerlendirme çalışmaları genellikle akademik kaygılarla derslerin etkililiğine yönelik ve okul, bölüm, fakülte veya il düzeyinde yürütülen çalışmalar niteliğindedir.
- Yapılan çalışmaların bulguları genellikle yetersizdir.

3. YÖNTEM

Araştırmanın bu bölümünde araştırma modeli, araştırmanın evren ve örnekleme, araştırmada kullanılan veri toplama aracı, verilerin toplanması ve elde edilen verilerin çözümlenmesinde kullanılan istatistiksel teknikler açıklanmıştır.

3.1. Araştırma Modeli

Araştırma, tarama modellerinden tekil ve ilişkisel tarama modelleri kullanılarak gerçekleştirilmiştir.

Tekil tarama modeli, değişkenlerin tek tek tür ya da miktar olarak oluşumlarının belirlenmesi amacıyla yapılan araştırma modelidir (Karasar, 2008). Bu yaklaşımda, “ilgilenilen olay, madde, birey, grup, kurum, konu vb. birim ve duruma ilişkin değişkenler ayrı ayrı betimlenmeye çalışılır (Karasar, 2008). Bu araştırmada tekil tarama modeliyle, araştırmacı ve tez danışmanları tarafından oluşturulan ifadelerin Türkiye için program değerlendirme standartları olarak uygun olup olmadığı eğitim programları ve öğretim alanında üniversitelerde görevli uzmanların görüşlerine dayalı olarak betimlenmiştir.

İlişkisel tarama modeli ise, iki ya da daha çok sayıdaki değişken arasında birlikte değişimin varlığını ve / veya derecesini belirlemeyi amaçlayan bir araştırma modelidir (Karasar, 2008). Bu araştırmada ilişkisel tarama modeliyle, program değerlendirme standartlarının uygunluk düzeylerinin standart alanlarına göre farklılık gösterip göstermediğine bakılmıştır.

3.2. Çalışma Evreni

Bu araştırmanın çalışma evrenini, doktora eğitim programları ve öğretim alanında tamamlamış ve 2008–2009 öğretim yılında devlet üniversitelerine bağlı eğitim

fakültelerinde görev yapan öğretim üyeleri oluşturmaktadır. Çalışma evrenindeki öğretim üyelerinin belirlenmesi için Yükseköğretim Kurulu'nun 2008–2009 öğretim yılı istatistiklerine dayalı olarak Türkiye'deki 94 devlet üniversitesi bünyesinde yer alan 64 eğitim fakültesinin web sayfaları ve öğretim elemanı veri tabanları incelenmiştir. Sınırlılıklar dahilinde yapılan inceleme sonucu doktorasını eğitim programları ve öğretim alanında yapmış ve halen devlet üniversitelerinin eğitim fakültelerinde görev yapmakta olan 158 öğretim üyesi araştırmanın çalışma evrenini oluşturmuştur. Evren ulaşılabilir nitelikte olduğu için ayrıca örneklem alınmamıştır. Çalışma evrenini oluşturan eğitim programları ve öğretim alanındaki öğretim üyelerinin akademik unvanlarına göre dağılımı Tablo 8'de verilmiştir.

Tablo 8.

Araştırmanın Çalışma Evrenini Oluşturan Uzmanların Akademik Unvanlarına Göre Dağılımı (n=158)

	Bayan	Erkek	Toplam
Profesör	8	10	18
Doçent	5	7	12
Yardımcı Doçent	49	64	113
Doktor	8	7	15
	70	88	158

Tablo 8'de de görüldüğü gibi, araştırmanın çalışma evrenini oluşturan öğretim üyelerinin %11'i profesör, % 8'i doçent, % 72'si yardımcı doçent ve % 9'u da doktordur. Araştırmanın çalışma evreni belirlendikten sonra her uzmanın e-posta adresine veri toplama aracının yer aldığı web sitesi bilgileri gönderilmiştir. Uzmanlara gönderilen ve uzmanlardan geri dönen anket sayısı Tablo 9'da verilmiştir.

Tablo 9.

Uzmanlara Gönderilen ve Uzmanlardan Geri Dönen Anket Sayısı (n=158)

	Çalışma Evrenindeki Uzman Sayısı	Gönderilen Anket Sayısı	Geri Dönen Anket Sayısı	Değerlendirme Dışı Bırakılan Anket Sayısı	Değerlendirmeye Alınan Anket Sayısı
Profesör	18	18	12		12
Doçent	12	12	10	1	9
Yardımcı Doçent	113	113	105	1	104
Doktor	15	15	11		11
Toplam	158	158	138	2	136

Tablo 9’da da görüldüğü gibi, araştırmanın çalışma evrenini oluşturan toplam 158 uzmana gönderilen veri toplama aracından 138’i geri dönmüştür. Geri dönen 138 veri toplama aracından 2’sinin eksik doldurulduğu belirlenmiştir. Bu nedenle bu veri toplama araçları değerlendirme dışı bırakılmıştır. Buna göre, 136 veri toplama aracı değerlendirmeye alınmıştır. Buna göre araştırmanın çalışma evrenini oluşturan 158 kişilik uzman grubunun %87’sine (138) ulaşılmıştır.

3.3. Verilerin Toplanması

Araştırmanın amacı doğrultusunda, araştırma için gerekli verilerin toplanması için anket tekniği kullanılmıştır. Araştırmada, eğitimde program değerlendirme standartlarının ve göstergelerinin belirlenmesi için eğitim programları ve öğretim alanında görev yapan uzmanların görüşlerine gereksinim duyulduğu için anket tekniğinin kullanılması uygun bulunmuştur.

3.4. Veri Toplama Aracının Hazırlanması

Araştırma verilerini toplamak üzere, Türkiye’deki program değerlendirme uzmanlarına yol gösterecek ve program değerlendirme çalışmalarında kullanılabilecek program değerlendirme standartlarını oluşturmak amacıyla doktorasını eğitim programları ve öğretim alanında tamamlamış ve devlet üniversitelerinde görev yapmakta olan uzmanların görüşlerini almaya yönelik bir veri toplama aracının hazırlanmasına karar verilmiştir. Bu nedenle, önce veri toplama aracının oluşturulabilmesi için gerekli olan alanyazın taraması yapılmış ve konuyla ilgili ulusal ve uluslararası program değerlendirme standartlarının neler olduğu belirlenmeye çalışılmıştır. İnternet yoluyla yapılan aramalarda birçok ülkenin eğitimde program değerlendirme standartları, yönergeleri ve ilkeleri ile ilgili geliştirme ve uyarlama çalışmalarına ilişkin bilgiler toplanmıştır. Ulaşılan standart geliştirme ve uyarlama çalışmaları incelenip, gerekli Türkçe çevirileri yapıldıktan sonra bu çalışmalarda yer alan standartlar, yönerge maddeleri ve ilkeleri bir havuzda toplanmıştır. Tüm ifadeler araştırmacı ve tez danışmanı ile birlikte incelenmiş ve kullanım amaçlarına göre, ABD’de Ortak Komite tarafından eğitimde program değerlendirme amaçlı geliştirilen program değerlendirme standartlarının yapısı temel alınarak bir sınıflandırma yapılmıştır. Bu sınıflamada yer alan standartlar genel ifadeler niteliğinde olduğundan tüm standartlar için göstergeler geliştirilmiş ve göstergelerle ilgili tez danışmanının onayı alınmıştır. Araştırmacı

tarafından önerilen tüm standartlar ve bu standartlar altında yer alan göstergeler Tez İzleme Komitesi toplantılarında komite üyeleriyle birlikte uygulanabilirlik, anlaşılabilirlik, işlevsellik, açıklık vb. yönlerden ayrıntılı olarak incelenmiş ve bu kapsamda “Türkiye İçin Program Değerlendirme Standartları Oluşturma Çalışması Anketi” adı altında uzmanlara yönelik deneme formu niteliğinde bir veri toplama aracı hazırlanmıştır (EK 2).

Hazırlanan veri toplama aracının kapsam geçerliği çalışmasını yapmak üzere bir deneme formu oluşturularak, veri toplama aracının, araştırmanın amacına uygunluğu, içeriğin amaca uygunluğu, yönergelerin, içeriğin ve yanıtlama biçiminin anlaşılır olup olmadığını belirlemek için üçü profesör, beşi yardımcı doçent, beşi öğretim görevlisi olmak üzere toplam 13 alan uzmanının görüşüne sunulmuştur. Uzmanlardan veri toplama aracının maddelerini incelemeleri, tutarsız, ilgisiz ve binişik ifadeleri belirtmeleri istenmiştir. Uzmanların veri toplama aracı taslağından çıkarılmasını ve taslağı eklenmesini önerdiği maddeler Tablo 10’da verilmiştir.

Tablo 10’da da görüldüğü gibi, “*Yararlılık*” standart alanına ilişkin olarak uzmanlar altı, “*Uygunluk*” standart alanına ilişkin bir, “*Doğruluk*” standart alanına ilişkin olarak yedi göstergenin anketten çıkarılması gerektiğini belirtmişler, “*Yürütülebilirlik*” standart alanına ilişkin madde çıkarılması önerisinde bulunmamışlardır. Bununla birlikte bazı ifadelerin de düzeltilmesini önermişlerdir. Uzmanlar ayrıca bazı standartların göstergeleri ile birlikte farklı standart alanı altında yer alması gerektiğini ifade etmişlerdir. Yine uzmanlar “*Yararlılık*” standart alanına ilişkin olarak yedi, “*Uygunluk*” standart alanına ilişkin olarak iki ve “*Doğruluk*” standart alanına ilişkin olarak da beş yeni gösterge önerisinde bulunmuşlardır.

Tablo 10.

Kapsam Geçerliğine Dayalı Olarak Çıkarılması ve Eklenmesi Önerilen Maddeler

Standart Alanları	Değiştirilmesi/Çıkarılması Önerilen Maddeler	Eklenmesi Önerilen Maddeler
“Yararlık” Standartları	<ul style="list-style-type: none"> • Paydaşları tanımlama • Değerlendirmeyi destekleyenleri tanımlama • Değerlendirme sonuçlarının uygulanması ve yayınlanması ile ilgili sorumlu grup ya da bireyleri belirtme • Raporun tüm ilgililere zamanında ulaşmasını sağlama • Değerlendirme sonuçlarının eğitsel etkilerini belirtme • Değerlendirme sonuçlarının politik, ekonomik, sosyal ve yönetsel etkilerini belirtme 	<p>YA.1.1. Program değerlendirme araştırmasını destekleyenleri tanımlama</p> <p>YA.1.5. Program değerlendirme araştırmasında sonuçların uygulanmasından sorumlu grup ya da bireyleri belirtme</p> <p>YA.1.6. Program değerlendirme araştırmasında sonuçların yayınlanmasından sorumlu grup ya da bireyleri belirtme</p> <p>YA.2.6. Program değerlendirme sürecinde tüm ilgililer karşısında tarafsızlığını koruma</p> <p>YA.5.1. Program değerlendirme sonuçlarının nasıl, nerede ve hangi amaçla kullanılacağını belirleme</p> <p>YA.5.2. Program değerlendirme sonuçlarının paydaşlar yönünden olası yararlarını/etkilerini belirleme</p> <p>YA.5.3. Program değerlendirme sonuçlarının paydaşlarda yeni bakış açısı/anlayış geliştirmesi</p>
“Yürütülebilirlik” Standartları	Değiştirilmesi Önerilen Madde yoktur	Eklenmesi Önerilen Madde yoktur
“Uygunluk” Standartları	<ul style="list-style-type: none"> • Sözleşmede değerlendirme bütçesini ayrıntıları ile sunma • Çıkar Çatışması standartı ve göstergeleri • Finansal Uygunluk Standardının Göstergelerinin bu standart altından çıkarılıp Resmî Sözleşme Standardının altına alınması önerilmiştir. <ul style="list-style-type: none"> • Değerlendirme sürecinde kullanılacak hizmet (personel, danışmanlar, seyahat, kaynaklar, posta ve telefon, verilerin çözümlenmesi, baskı ve yayınlama) harcamaları ayrıntılı olarak belirtme • Süreç içinde hizmetlerin ve harcamaların ayrıntılı kayıtlarını tutma 	<p>U.2.7. Verileri yalnızca bilimsel süreç ve sonuçlar için kullanma</p> <ul style="list-style-type: none"> • “Doğruluk” Standart Alanı altındaki Şeffaf Raporlaştırma standardının göstergeleriyle birlikte “Uygunluk” Standart alanına alınması önerilmiştir.) <ul style="list-style-type: none"> U.4.1. Raporun paydaş taleplerinden bağımsız olması U.4.2. Raporun değerlendiricinin nesnellliğini yansıtması U.4.3. Raporun tarafsız bir değerlendirmeci tarafından incelenmesi U.5.6. Sonuçları ilgili kesimlere ulaştırma • Yeniden oluşturulan Finansal “Uygunluk” Standardının altına aşağıdaki göstergelerin eklenmesi önerilmiştir. <ul style="list-style-type: none"> U.6.1. Değerlendirme gelir-gider bütçesini belirtme U.6.2. Değerlendirme için sağlanan finansman kaynaklarını belirtme U.6.3. Değerlendirme süreci içinde yapılan harcamaları belirtme U.6.4. Gelir-gider bütçesinin raporlarını gerektiğinde değerlendirme raporuna ekleme
“Doğruluk” Standartları	<ul style="list-style-type: none"> • Program paydaşlarının program ile ilgili görüşlerini belirleme • Değerlendirme modelini belirleme • Araştırma desenini belirleme • Desen ve modelin sınırlılıklarını ifade etme • Veri çözümleme işlemleri için istatistiksel tekniklere karar verme • İstatistiksel tekniklerin varsayımlarının nasıl karşılandığını açıklamak • Alternatif yorumların neden elendiği ifade etme • Şeffaf Raporlaştırma Standardının göstergeleri ile birlikte “Uygunluk” Standartları Alanına Eklenmesi Önerilmiştir. <ul style="list-style-type: none"> • Raporun paydaş taleplerinden bağımsız olması • Raporun değerlendiricinin nesnellliğini yansıtması • Raporun tarafsız bir değerlendirmeci tarafından incelenmesi 	<p>D.3.4. Belirlenen amaçların hangi gereksinimleri karşılayacağını gerekçeleriyle ifade etme</p> <p>D.4.1. Program değerlendirme araştırmasında temel alınan program değerlendirme modelini ve sınırlılıklarını belirleme</p> <p>D.4.2. Program değerlendirme araştırmasının desenini ve sınırlılıklarını belirleme</p> <p>D.7.1. Nicel verileri çözümlmek için uygun nicel veri çözümleme tekniklerini belirleme</p> <p>D.7.2. Nitel verileri çözümlmek için uygun nitel veri çözümleme tekniklerini belirleme</p>

Uzmanlardan gelen bu görüşler tez izleme jürisi ile birlikte değerlendirildikten sonra veri toplama aracı yeniden düzenlenmiştir. Kapsam geçerliği çalışmasından sonra araştırmacının çalışma evrenindeki uzman sayısının azlığı nedeniyle nitel çalışmalarda başvurulan güvenilirlik testine başvurulmuştur. Bunun için öncelikle araştırmacının oluşturmuş olduğu veri toplama aracı taslağında yer alan göstergeler standart başlıkları kaldırılarak kapsam geçerliği çalışmasının yapıldığı uzmanlardan farklı altı uzmana verilmiş ve uzmanlardan standart alanından bağımsız olarak verilen göstergeleri ayrı ayrı “Yararlık”, “Yürütülebilirlik”, “Uygunluk” ve “Doğruluk” standart alanlarındaki standartların altına yerleştirmeleri istenmiştir. Araştırmacı ile uzmanların ilgili göstergeleri standartların altına yerleştirmelerine ilişkin tutarlığın/güvenirliğin araştırıldığı bu aşamada, araştırmacı ile uzmanların her bir göstergeye ilişkin yaptıkları işaretlemeler birbirleriyle aynı ise bu durum bir “uzmanlar arası görüş birliği” olarak, işaretlemeler birbirinden farklıysa bir “görüş ayrılığı” olarak kabul edilmiştir. Araştırmacı ile uzmanlar arasındaki bu tutarlığın hesaplanması için öncelikle araştırmacı ile uzmanlar arasında görüş birliğine varılan göstergeler ve görüş ayrılığı olan göstergeler belirlenip Miles ve Huberman (1994)’ın

$$\text{Güvenirlik} = \frac{\text{Görüş Birliği}}{\text{Görüş Birliği} + \text{Görüş Ayrılığı}}$$

formülünden yararlanılarak veri toplama aracının güvenilirliğine ilişkin hem her standart alanına hem de veri toplama aracının geneline ilişkin bir katsayı belirlenmiştir.

Tablo 11.

Veri Toplama Aracının Standart Alanlarına İlişkin Miles ve Huberman Güvenirlik Katsayıları

Standart Alanları	Gösterge Sayısı	Görüş Birliği Sayısı	Görüş Ayrılığı Sayısı	Güvenirlik Katsayısı
Yararlık Standartları	24	118	26	0,82
Yürütülebilirlik Standartları	17	75	27	0,74
Uygunluk Standartları	33	173	25	0,87
Doğruluk Standartları	36	184	32	0,85
Genel	110	550	110	0,83

Tablo 11’de dört standart alanına ilişkin güvenilirlik katsayılarının .70’in üzerinde olduğu görülmekte, veri toplama aracının geneline yönelik güvenilirlik katsayısının ise .83

olduğu görülmektedir. Hesaplanan güvenilirlik puanları %70'in üzerinde olduğu için araştırmacının standart göstergelerine ilişkin görüşü ile uzmanların standart göstergelerine ilişkin görüşleri arasında bir tutarlılık/güvenirlik olduğu, böylelikle veri toplama aracının güvenilir olduğu sonucuna varılmıştır (Miles ve Huberman, 1994). Öte yandan standart alanları altındaki göstergelerden güvenilirlik katsayısı .70'in altında çıkan göstergeler ile ilgili olarak da uzmanlarla tekrar görüşülerek görüş birliğine varılmaya çalışılmıştır.

Tüm bu süreçlerden sonra gelen dönütler bir araya getirilmiş ve ifadeler gözden geçirilerek veri toplama aracında gerekli düzenlemeler yapıp alan uzmanlarına yönelik iki bölümden oluşan bir veri toplama aracı düzenlenmiştir. Bu bölümler şunlardır:

- Birinci Bölüm: Anketin bu bölümünde çalışma evrenine giren uzmanları tanımaya dönük sorular yer almaktadır. Bu kapsamda uzmanların adları ve görev yaptıkları üniversiteleri saptanmaya çalışılmıştır.
- İkinci Bölüm: Anketin bu bölümünde ise Türkiye için uygunluğu araştırılan program değerlendirme standartları yer almaktadır. Bu bölüm de kendi içinde şu alt bölümlerden oluşmaktadır:
 - ▮ Beş standart ve bu standartların altında yer alan 24 göstergeden oluşan “*Yararlık*” standartları.
 - ▮ Dört standart ve bu standartların altında yer alan 17 göstergeden oluşan “*Yürütülebilirlik*” standartları.
 - ▮ Altı standart ve bu standartların altında yer alan 33 göstergeden oluşan “*Uygunluk*” standartları.
 - ▮ Sekiz standart ve bu standartların altında yer alan 36 göstergeden oluşan “*Doğruluk*” standartları.

3.5. Veri Toplama Aracının Uygulanması

Uzman görüşüne dayalı olarak kapsam geçerliği ve tutarlık çalışmaları yapılan veri toplama aracının asıl uygulamasının web ortamı üzerinden yapılmasına karar verilmiştir. Hazırlanan veri toplama aracı formu, alan uzmanlarına uygulanmaya başlanmadan önce web ortamına aktarılmış, erişilebilirliği ve uygulanabilirliği gözden geçirilmiştir. Daha sonra doktorasını eğitim programları ve öğretim alanında yapmış ve

devlet üniversitelerine görev yapan uzmanların sayısı belirlenmiş ve belirlenen tüm öğretim elemanlarının e-posta adresleri çalıştıkları üniversitelerin web sayfaları, veritabanları ve kişisel web sayfaları ve telefon görüşmeleri yoluyla alınmıştır. Her uzmana, veri toplama aracını web üzerinden doldurabilmesi için bir e-davet gönderilmiştir. Veri toplama aracının web adresinde yer alan yönergelere uygun olarak uzmanlardan, unvan ve isimlerini belirterek veri toplama aracını doldurmaları istenmiştir. Veri toplama aracına erişim sağlayamayan uzmanlarla iletişime geçilmiş ve gerekli yardımlarda bulunulmuştur. Veri toplama aracını dolduran uzmanların tüm verileri Anadolu Üniversitesi tarafından sağlanan veritabanında kayıt altına alınmıştır. Veri toplama aracına birden fazla erişimin sağlanmasını engellemek amacıyla her kullanıcının anket formunu doldurduğu bilgisayarın IP numarası veri tabanına kayıt edilmiş ve çoklu kayıtlar, veri toplama aracını dolduran uzmanın adı ve IP numaraları karşılaştırılarak değerlendirme dışı bırakılmıştır. Ayrıca elde edilen IP numaraları ile adını belirtmeden veri toplama aracını dolduran uzmanların hangi üniversitelerden veri toplama araçlarını doldurdukları belirlenmiştir.

3.6. Verilerin Çözümlemesi

Web ortamına aktarılan “Türkiye İçin Program Değerlendirme Standartları Oluşturma Çalışması Anketi” ile toplanan ve Anadolu Üniversitesi tarafından sağlanan veritabanında kayıta altına alınan veriler, uzman adı ve IP numarası karşılaştırması yapıldıktan ve eksik formlar değerlendirme dışı bırakıldıktan sonra araştırmanın amaçları doğrultusunda gerekli istatistiksel çözümler için SPSS (The Statistical Packet for Social Sciences) paket programına aktarılmıştır. Araştırmada yapılan istatistiksel çözümlerlerde $\alpha=0.05$ anlamlılık düzeyi kullanılmıştır.

Araştırmanın amaçlarını gerçekleştirirken SPSS paket programından yararlanarak;

1. Program değerlendirme standartları taslağının “*Yararlılık*”, “*Yürütülebilirlik*”, “*Uygunluk*” ve “*Doğruluk*” standartlarına ilişkin uzman görüşlerinin çözümlenmesinde sayı, yüzde, aritmetik ortalama ve standart sapmalardan yararlanılmıştır.

2. Program değerlendirme standartları uygunluk düzeyinin standart alanlarına göre farklılık gösterip göstermediğini belirlemede her bir standart alanının aritmetik ortalaması ve standart sapmalarından yararlanılmıştır. Aritmetik ortalamalar arasındaki farkın anlamlı olup olmadığını test etmede tek yönlü varyans çözümlemesinden yararlanılmıştır. Tek yönlü varyans çözümlemesinin ortaya koyduğu farkın kaynağının belirlenmesinde ise Tukey HSD testi kullanılmıştır.

Program değerlendirme standartlarının uygunluğuna ilişkin uzman görüşlerinin belirlenmesi amacıyla beşli likert tipinde hazırlanan ve kullanılan veri toplama aracında standartlara verilecek puanlar da beşli olarak derecelendirilmiştir. Derecelendirme; veri toplama aracındaki her bir ifade için; “tamamen uygun”, “oldukça uygun”, “kısmen uygun”, “çok az uygun” ve “hiç uygun değil” şeklindedir. Veri toplama aracındaki her bir maddeye verilecek yanıt puanları da bu derecelendirmelere uygun olarak olumludan olumsuzaya doğru 5.00 ile 1.00 arasında değişmektedir. Ölçme aracında yer alan aralıkların eşit olduğu (4/5) varsayımından hareket ederek önce seçeneklere ait alt ve üst sınırlar belirlenmiştir. Bunun için, veri toplama aracı maddelerinden elde edilen beşli likert dağılımlarını değerlendirmek için aritmetik ortalamalara paralel olarak her bir madde için standart görüş aralıkları belirlemek amacıyla $(n-1)/n=(5-1)/5=0.80$ formülü kullanılmıştır. Daha sonra, program değerlendirme standartlarının uygunluğuna ilişkin görüşlerden elde edilen verilerin ortalamalara göre yorumlanması ve ortalamaların hangi düzeylere girdiğinin belirlenmesi amacıyla bu derecelendirmeler üç düzeyde toplanmıştır. Bu düzeyler, ölçme aracındaki her bir madde dikkate alınarak *tamamen uygun*, *oldukça uygun* “uygun”; *kısmen uygun* “kısmen uygun” ve *çok az uygun ve hiç uygun değil* “uygun değil” olarak belirlenmiştir. Ölçme aracındaki her bir maddeye ilişkin derecelendirme sınırları ve düzeyleri Tablo 12’de verilmiştir.

Tablo 12.

Seçeneklere İlişkin Sınırlar ve Gruplamalar

Seçenekler	Sınırlar	Düzyerler	Sınırlar
Tamamen Uygun (5)	4.20 – 5.00	Uygun	3.40 – 5.00
Oldukça Uygun (4)	3.40 – 4.19		
Kısmen Uygun (3)	2.60 – 3.39	Kısmen Uygun	2.60 – 3.39
Çok Az Uygun (2)	1.80 – 2.59	Uygun Değil	1.00 – 2.59
Hiç Uygun Değil (1)	1.00 – 1.79		

Sonuç olarak veri toplama aracındaki herhangi bir “i” maddesine ait aritmetik ortalama (\bar{X}_i) eğer;

- $1.00 \leq 2.59$ arasında ise maddenin uzmanlar tarafından Türkiye için program değerlendirme standardı ve göstergesi olarak *uygun bulunmadığı*,
- $2.60 \leq 3.39$ arasında ise maddenin uzmanlar tarafından Türkiye için program değerlendirme standardı ve göstergesi olarak *kısmen uygun bulunduğu*,
- $3.40 \leq 5.00$ arasında ise maddenin uzmanlar tarafından Türkiye için program değerlendirme standardı ve göstergesi olarak *uygun bulunduğu* şeklinde değerlendirilmiştir.

4. BULGULAR VE YORUMLARI

Türkiye’de devlet üniversitelerinin eğitim fakültelerinde görev yapan ve doktorasını eğitim programları ve öğretim alanında yapmış olan alan uzmanlarının görüşlerine göre, OKPDS temel alınarak oluşturulan program değerlendirme standartlarının ve göstergelerinin Türkiye için uygunluğunun belirlenmesinin amaçlandığı araştırmanın bu bölümünde, araştırmanın amacına dönük soruların yanıtlanması için toplanan verilerin istatistiksel tekniklerle çözümlenmesiyle elde edilen bulgulara ve yorumlarına yer verilmiştir. Bulgular ve yorumlarının verilmesinde, araştırmanın amacı doğrultusunda yanıtı aranan soruların sırası izlenmiştir.

4.1. Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarının Uygunluğuna İlişkin Görüşleri

Araştırmada yanıtı aranan ilk soru, alan uzmanlarının program değerlendirme standartları taslağında yer alan ve “Yararlık” standart alanını oluşturan alt standartların göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerinin ne olduğudur. Alan uzmanlarının görüşleri aşağıdaki beş alt standartın göstergeleri kapsamında incelenmiştir:

- *Paydaşların Tanımlanması*
- *Değerlendirici Yeterliği ve Güvenirliği*
- *Verilerin Kapsamı ve Seçimi /Yararlığı*
- *Raporun Kapsamlılığı ve Anlaşılabilirliği*
- *Sonuçların Yararlılığı/Etkisi*

Uzmanlarının görüşlerinin çözümlenmesinde sayı, yüzde ve aritmetik ortalamalar kullanılmış, verilerin yorumlanmasında ise yöntem bölümünde belirlenen sınırlar ve düzeylere göre aritmetik ortalamalar temel alınmıştır.

4.1.1. Alan Uzmanlarının *Paydaşların Tanımlanması* Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “Yararlık” standartlarını oluşturan beş alt standarttan ilki “Paydaşların Tanımlanması” standarttır. Alan uzmanlarının “Paydaşların Tanımlanması” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 13’te verilmiştir.

Tablo 13.

Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarından “Paydaşların Tanımlanması” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Program değerlendirme araştırmasını destekleyenleri tanımlama	70	51,5	34	25	17	12,5	7	5,1	8	5,9	4.11	.82
2. Program değerlendirme araştırmasını planlayacak, uygulayacak ve değerlendirecek değerlendirme uzman/ uzmanlarını tanımlama	97	71,3	25	18,4	3	2,2	2	1,5	9	6,6	4.46	.81
3. Program değerlendirme araştırmasının araçlarının uygulanacağı grup ya da bireyleri tanımlama	106	77,9	16	11,8	3	2,2	2	1,5	9	6,6	4.52	.91
4. Program değerlendirme araştırmasının sonuçlarından doğrudan ya da dolaylı etkilenecek grup ya da bireyleri belirtme	89	65,4	27	19,9	7	5,1	5	3,7	8	5,9	4.35	.85
5. Program değerlendirme araştırmasında sonuçların uygulanmasından sorumlu grup ya da bireyleri belirtme	34	25,0	6	4,4	5	3,7	34	25	57	41,9	2.45	1.64
6. Program değerlendirme araştırmasında sonuçların yayınlanmasından sorumlu grup ya da bireyleri belirtme	26	19,1	11	8,1	15	11,0	27	19,9	57	41,9	2.42	1.55
	3.72											

Tablo 13’te görüldüğü gibi, program değerlendirmede “Yararlık” standartlarından ilki olan “Paydaşların Tanımlanması” standartının altı göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Program değerlendirme araştırmasını destekleyenleri tanımlama”dır. Alan uzmanlarının %76,5’i bu göstergeyi “uygun” bulurken, uzmanların %12,5’i bu göstergenin “kısmen uygun”, %11’i de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.11’dir. Bu ortalama, “Program değerlendirme araştırmasını destekleyenleri tanımlama” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “*Paydaşların Tanımlanması*” standartıyla ilgili ikinci gösterge “Program değerlendirme araştırmasını planlayacak, uygulayacak ve değerlendirecek değerlendirme uzman/uzmanlarını tanımlama”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %89,7 “uygun”, %2,2 “kısmen uygun” ve %8,1 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.46’dır. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Program değerlendirme araştırmasını planlayacak, uygulayacak ve değerlendirecek değerlendirme uzman/uzmanlarını tanımlama” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Paydaşların Tanımlanması*” standartının üçüncü göstergesi, “Program değerlendirme araştırmasının araçlarının uygulanacağı grup ya da bireyleri tanımlama”dır. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %89,7 “uygun”, % 2,2 “kısmen uygun” %8,1 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.52’dir. Bu ortalama, bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“*Paydaşların Tanımlanması*” standartının dördüncü göstergesi, “Program değerlendirme araştırmasının sonuçlarından doğrudan ya da dolaylı etkilenecek grup ya da bireyleri belirtme”dir. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %85,3’ünün göstergeyi “uygun” bulduğu, %5,1’lik bir kesiminin “kısmen uygun” bulduğu ve %9,6’sının da uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.35’tir. Bu ortalamaya göre, bu göstergenin de uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Paydaşların Tanımlanması*” standartının beşinci göstergesi, “Program değerlendirme araştırmasında sonuçların uygulanmasından sorumlu grup ya da bireyleri belirtme”dir. Bu göstergeye ilişkin yanıtlar incelendiğinde 40 uzmanın (%29,4) göstergenin “uygun” olduğu yönünde görüş bildirdiği, buna karşılık 91 uzmanın (%66,9) uygun olmadığı yönünde görüş bildirdiği görülmektedir. Bu göstergeyi “kısmen uygun” bulan

uzmanların sayısı beştir (%3,7). Uzmanların bu göstergeye ilişkin yanıtlarının aritmetik ortalaması ise 2.45'tir. Bu değerlere göre, "Program değerlendirme araştırmasında sonuçların uygulanmasından sorumlu grup ya da bireyleri belirtme" göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiştir.

"*Paydaşların Tanımlanması*" standartının son göstergesi, "Program değerlendirme araştırmasında sonuçların yayınlanmasından sorumlu grup ya da bireyleri belirtme"dir. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %27,2'si bu göstergesi "uygun" bulurken, %61,8'i uygun bulmamaktadır. Uzmanların % 11'i bu göstergenin "kısmen uygun" olduğu görüşündedir. Uzmanların görüşlerinin aritmetik ortalaması ise 2.42'dir. Bu değerler, bu göstergenin de uzmanlar tarafından Türkiye için uygun bulunmadığını göstermektedir.

Program değerlendirmede "*Yararlık*" standartının ilk alt standardı olan "*Paydaşların Tanımlanması*" standartının göstergelerinin Türkiye için uygunluğuna ilişkin olarak uzmanlar dört göstergesi uygun bulmuşlar, iki göstergesi ise uygun bulmamışlardır. "*Paydaşların Tanımlanması*" standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 3.72 olarak bulunmuştur. Bu değer, alan uzmanlarının "*Paydaşların Tanımlanması*" standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standardı beşinci ve altıncı göstergeler dışında diğer göstergeleriyle birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.1.2. Alan Uzmanlarının Değerlendirici Yeterliği ve Güvenirliği Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede "*Yararlık*" standartlarını oluşturan beş alt standarttan ikincisi "*Değerlendirici Yeterliği ve Güvenirliği*" standartıdır. Alan uzmanlarının "*Değerlendirici Yeterliği ve Güvenirliği*" standartının altı göstergesinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 14'te verilmiştir.

Tablo 14.

*Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarından
“Değerlendirici Yeterliği ve Güvenirliği” Standartının Göstergelerinin Uygunluğuna
İlişkin Görüşleri*

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
	1. Program değerlendirme konusunda akademik bir süreçten geçmiş olma	107	78,7	16	11,8	2	1,5	1	,7	10		
2. Program değerlendirme araştırmasının planlanma, uygulama ve değerlendirme süreçlerinde gerekli bilgi, beceri ve deneyimlere sahip olma	106	77,9	16	11,8	2	1,5	1	,7	11	8,1	4.51	1,14
3. Bireysel sınırlılıkların farkına varma, bu sınırlılıkları aşan durumlarda uzman yardımı alma	9	6,6	24	17,6	18	13,2	23	16,9	62	45,6	2.23	1,36
4. Farklı gruplar ile çalışmak için uygun iletişim stratejilerini ve becerilerini kullanma	84	61,8	34	25,0	4	2,9	5	3,7	9	6,6	4.32	1,14
5. Program değerlendirme sürecinde gerekli durumlarda dış değerlendirmeci yardımına başvurma	81	59,6	35	25,7	7	5,1	4	2,9	9	6,6	4.29	1,13
6. Program değerlendirme sürecinde tüm ilgililer karşısında tarafsızlığını koruma	100	73,5	21	15,4	2	1,5	2	1,5	11	8,1	4.45	1,16
											4.05	

Tablo 14’te görüldüğü gibi, alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Program değerlendirme konusunda akademik bir süreçten geçmiş olma”dır. Alan uzmanlarının %90,5’i bu göstereyi “uygun” bulurken, uzmanların %1,5’i bu göstergenin “kısmen uygun”, %8,1’i de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.54’tür. Bu ortalama, “Program değerlendirme konusunda akademik bir süreçten geçmiş olma” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “Değerlendirici Yeterliği ve Güvenirliği” standartıyla ilgili ikinci gösterge “Program değerlendirme araştırmasını planlanma, uygulama ve değerlendirme süreçlerinde gerekli bilgi, beceri ve deneyimlere sahip olma”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %89,7 “uygun”, %1,5 “kısmen uygun” ve %8,8 de “uygun değil” şeklindedir.

Alan uzmanlarının bu göstergenin uygunluđuna ilişkin görüřlerinin aritmetik ortalaması ise 4.51'dir. Bu ortalamaya göre, alan uzmanlarının büyük çođunluđu, "Program deđerlendirme arařtırmasını planlanma, uygulama ve deđerlendirme süreçlerinde gerekli bilgi, beceri ve deneyimlere sahip olma" göstergesinin Türkiye için "uygun" olduđu görüřündedir.

"Deđerlendirici Yeterliđi ve Güvenirliđi" standartının üçüncü göstergesi, "Bireysel sınırlılıkların farkına varma, bu sınırlılıkları aşan durumlarda uzman yardımı alma"dır. Bu göstergeye ilişkin yanıtlar incelendiđinde 33 uzmanın (%24,2) göstergenin "uygun" olduđu yönünde görüř bildirdiđi, buna karřılık 85 uzmanın (%62,5) ise uygun olmadıđı yönünde görüř bildirdiđi görülmektedir. Bu göstergesi "kısmen uygun" bulan uzmanların sayısı 18'dir (%13,2). Uzmanların bu göstergeye ilişkin yanıtlarının aritmetik ortalaması 2.23'dir. Bu deđerlere göre, "Bireysel sınırlılıkların farkına varma, bu sınırlılıkları aşan durumlarda uzman yardımı alma" göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiřtir.

"Deđerlendirici Yeterliđi ve Güvenirliđi" standartının dördüncü göstergesi, "Farklı gruplar ile çalışmak için uygun iletiřim stratejilerini ve becerilerini kullanma"dır. Alan uzmanlarının yanıtları incelendiđinde, uzmanların %86,8'inin göstergesi "uygun" bulduđu, %2,9'luk bir kesiminin "kısmen uygun" bulduđu ve %10,3'ünün de uygun bulmadıđı görülmektedir. Bu göstergenin uygunluđuna ilişkin uzman görüřlerinin aritmetik ortalaması 4.32'dir. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için "uygun" görüldüđu söylenebilir.

"Deđerlendirici Yeterliđi ve Güvenirliđi" standartının beřinci göstergesi, "Program deđerlendirme sürecinde gerekli durumlarda dıř deđerlendirmeci yardımına başvurma"dır. Bu göstergenin uygunluđuna ilişkin arařtırma anketini yanıtlayan uzmanların görüřlerinin dađılımı; %85,3 "uygun", % 5,1 "kısmen uygun" %9,5 "uygun deđil" şeklindedir. Alan uzmanlarının yanıtlarının bu dađılımı, uzmanların büyük çođunluđunun bu göstergesi uygun bulduđunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüřlerini yansıtan yanıtların aritmetik ortalaması da 4.29'dur. Bu ortalama, bu göstergenin Türkiye için uygun bir gösterge olduđunu göstermektedir.

“Değerlendirici Yeterliği ve Güvenirliği” standartının son göstergesi, “Program değerlendirme sürecinde tüm ilgililer karşısında tarafsızlığını koruma”dır. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %88,9’u bu göstergeyi “uygun” bulurken, %9,6’sı uygun bulmamaktadır. Uzmanların %1,5’i de bu göstergenin “kısmen uygun” olduğu görüşündedir. Uzmanların görüşlerinin aritmetik ortalaması 4.45’tir. Bu değerler, bu göstergenin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Program değerlendirmede “Yararlık” standartının ikinci alt standardı olan “Değerlendirici Yeterliği ve Güvenirliği” standartının göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın beş göstergesini uygun bulmuşlar, bir göstergeyi ise uygun bulmamışlardır. “Değerlendirici Yeterliği ve Güvenirliği” standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 4.05 olarak bulunmuştur. Bu değer, alan uzmanlarının “Değerlendirici Yeterliği ve Güvenirliği” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı üçüncü gösterge dışında diğer göstergeleriyle birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.1.3. Alan Uzmanlarının Verilerin Kapsamı ve Seçimi /Yararlığı Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “Yararlık” standartlarını oluşturan üçüncü standart “Verilerin Kapsamı ve Seçimi /Yararlığı”dır. Alan uzmanlarının “Verilerin Kapsamı ve Seçimi /Yararlığı” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 15’te verilmiştir.

Tablo 15.

*Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarından
“Verilerin Kapsamı ve Seçimi /Yararlılığı” Standartının Göstergelerinin Uygunluğuna
İlişkin Görüşleri*

(n=136)

Göstergeler	Tamamen Uygun		Orduka Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Program değerlendirme araştırmasının planlanmasında araştırma sorularının yanıtlanması için gerekli/yararlı bilgileri belirleme/seçme	94	69,1	24	17,6	6	4,4	2	1,5	10	7,4	4.40	1,14
2. Toplanan/toplanacak verilerden paydaşların gereksinimlerini karşılayacak olanlarına öncelik verme	74	54,4	37	27,2	13	9,6	4	2,9	8	5,9	4.21	1,12
											4.31	

Tablo 15’te görüldüğü gibi, alan uzmanlarından program değerlendirmede “Yararlık” standartlarından biri olan “Verilerin Kapsamı ve Seçimi /Yararlılığı” standartıyla ilgili görüş belirtmeleri istenen göstergelerden ilki “Program değerlendirme araştırmasının planlanmasında araştırma sorularının yanıtlanması için gerekli/yararlı bilgileri belirleme/seçme”dir. Alan uzmanlarının %86,7’si bu göstergeyi “uygun” bulurken, uzmanların %4,4’ü bu göstergenin “kısmen uygun”, %8,9’u da uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.40’tır. Bu ortalama, “Program değerlendirme araştırmasının planlanmasında araştırma sorularının yanıtlanması için gerekli/yararlı bilgileri belirleme/seçme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

“Verilerin Kapsamı ve Seçimi /Yararlılığı” standartıyla ilgili alan uzmanlarından görüş bildirmeleri istenen ikinci gösterge “Toplanan/toplanacak verilerden paydaşların gereksinimlerini karşılayacak olanlarına öncelik verme” dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %81,6 “uygun”, %9,6 “kısmen uygun” ve %8,8 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.21’dir. Bu ortalama göre, alan uzmanlarının büyük çoğunluğu, “Toplanan/toplanacak verilerden paydaşların

gereksinimlerini karşılayacak olanlarına öncelik verme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

Alan uzmanlarının program değerlendirmede “*Yararlık*” standart alanının üçüncü standartı olan “*Verilerin Kapsamı ve Seçimi /Yararlığı*” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan yanıtların aritmetik ortalamalarının ortalaması 4.31 olarak bulunmuştur. Bu değer, alan uzmanlarının “*Verilerin Kapsamı ve Seçimi /Yararlığı*” standartının göstergeleri ile birlikte Türkiye için program değerlendirme standartı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.1.4. Alan Uzmanlarının *Raporun Kapsamlılığı ve Anlaşılabilirliği* Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “*Yararlık*” standartlarını oluşturan dördüncü standart “*Raporun Kapsamlılığı ve Anlaşılabilirliği*” standartıdır. Alan uzmanlarının “*Raporun Kapsamlılığı ve Anlaşılabilirliği*” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 16’da verilmiştir.

Tablo 16.

*Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarından
“Raporun Kapsamlılığı ve Anlaşılabilirliği” Standartının Göstergelerinin Uygunluğuna
İlişkin Görüşleri*

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
	1. Raporunda program değerlendirme araştırmasının amacını, içeriğini, araştırma sorularını, süreçte kullanılan yöntemleri ve değerlendirme bulgularını açık ve kapsamlı olarak betimleme	105	77,2	13	9,6	6	4,4	4	2,9	8		
2. Raporu hedeflenen tüm kitlelerin özelliklerine göre özet rapor, ana rapor, teknik rapor ve sözel sunumlar halinde hazırlama	84	61,8	28	20,6	14	10,3	3	2,2	7	5,1	4.32	1,09
3. Raporunda sonuçların tüm ilgililer tarafından anlaşılabilirliğini sağlama için örnekler kullanma	85	62,5	35	25,7	5	3,7	2	1,5	9	6,6	4.36	1,09
4. Raporunda ilgililerin anlayabileceği nitelikte bir dil kullanma/ teknik dili az kullanma	84	61,8	27	19,9	9	6,6	4	2,9	12	8,8	4.23	1,25
5. Raporunda kullanılan önemli kavramları tanımlayan bir sözlük hazırlama	71	52,2	39	28,7	14	10,3	4	2,9	8	5,9	4.18	1,12
6. Raporun dili, biçimi, zamanlaması gibi konularda paydaşların durumlarını/konularını dikkate alma	62	45,6	54	39,7	7	5,1	4	2,9	9	6,6	4.15	1,10
7. Raporun gözden geçirilmesi için bağımsız değerlendirmeci görüşüne sunma	89	65,4	30	22,1	6	4,4	1	,7	10	7,4	4.38	1,12
											4.30	

Tablo 16’da görüldüğü gibi, program değerlendirme standartları altında yer alan “Yararlık” standartlarından biri olan “Raporun Kapsamlılığı ve Anlaşılabilirliği” standartının yedi göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Raporunda program değerlendirme araştırmasının amacını, içeriğini, araştırma sorularını, süreçte kullanılan yöntemleri ve değerlendirme bulgularını açık ve kapsamlı olarak betimleme”dir. Alan uzmanlarının %86,8’i bu göstergeyi “uygun” bulurken, uzmanların %4,4’ü bu göstergenin “kısmen uygun”, %8,8’i de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.49’dur. Bu ortalama, “Raporunda program değerlendirme araştırmasının amacını, içeriğini, araştırma sorularını, süreçte kullanılan yöntemleri ve

değerlendirme bulgularını açık ve kapsamlı olarak betimleme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından “*Raporun Kapsamlılığı ve Anlaşılabilirliği*” standartıyla ilgili görüş bildirmeleri istenen ikinci gösterge “Raporu hedeflenen tüm kitlelerin özelliklerine göre özet rapor, ana rapor, teknik rapor ve sözel sunumlar halinde hazırlama”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %82,4 “uygun”, %10,3 “kısmen uygun” ve %7,3 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.32’dir. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Raporu hedeflenen tüm kitlelerin özelliklerine göre özet rapor, ana rapor, teknik rapor ve sözel sunumlar halinde hazırlama” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Raporun Kapsamlılığı ve Anlaşılabilirliği*” standartının üçüncü göstergesi, “Raporda sonuçların tüm ilgililer tarafından anlaşılabilirliğini sağlama için örnekler kullanma”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtları incelendiğinde, uzmanların %88,2’sinin göstergeyi “uygun” bulduğu, %3,7’lik bir kesiminin “kısmen uygun” bulduğu ve %8,1’inin de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.36’dır. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Raporun Kapsamlılığı ve Anlaşılabilirliği*” standartının dördüncü göstergesi, “Raporda ilgililerin anlayabileceği nitelikte bir dil kullanma/ teknik dili az kullanma”dır. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %81,7 “uygun”, % 6,6 “kısmen uygun” %11,7 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.23’dir. Bu ortalama, bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“*Raporun Kapsamlılığı ve Anlaşılabilirliği*” standartının beşinci göstergesi, “Raporda kullanılan önemli kavramları tanımlayan bir sözlük hazırlama”dır. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %80,9’u bu göstergeyi “uygun”

bulurken, %10,3'ü uygun bulmamaktadır. Uzmanların % 8,8'i de bu göstergenin "kısmen uygun" olduğu görüşündedir. Uzmanların görüşlerinin aritmetik ortalaması ise 4.18'dir. Bu değerler, bu göstergenin de uzmanlar tarafından uygun bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen "*Raporun Kapsamlılığı ve Anlaşılabilirliği*" standartıyla ilgili altıncı gösterge "Raporun dili, biçimi, zamanlaması gibi konularda paydaşların durumlarını/konumlarını dikkate alma"dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %85,3 "uygun", %5,1 "kısmen uygun" ve %9,5 de "uygun değil" şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.15'tir. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, "Raporun dili, biçimi, zamanlaması gibi konularda paydaşların durumlarını/konumlarını dikkate alma" göstergesinin Türkiye için "uygun" olduğu görüşündedir.

"*Raporun Kapsamlılığı ve Anlaşılabilirliği*" standartının son göstergesi, "Raporun gözden geçirilmesi için bağımsız değerlendirmeci görüşüne sunma"dır. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %87,5'inin göstergelyi "uygun" bulduğu, %4,4'lük bir kesiminin "kısmen uygun" bulduğu ve %8,1'inin de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.38'dir. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için "uygun" görüldüğü söylenebilir.

"*Yararlılık*" standartının dördüncü standardı olan "*Raporun Kapsamlılığı ve Anlaşılabilirliği*" standartının göstergelerinin Türkiye için uygunluğuna ilişkin olarak uzmanlar yedi göstergelyi de uygun bulmuşlardır. "*Raporun Kapsamlılığı ve Anlaşılabilirliği*" standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 4.30 değeri bulunmuştur. Bu değer, alan uzmanlarının "*Raporun Kapsamlılığı ve Anlaşılabilirliği*" standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı tüm göstergeleri ile birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.1.5. Alan Uzmanlarının *Sonuçların Yararlılığı/Etkisi* Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “*Yararlık*” standartlarını oluşturan son standart “*Sonuçların Yararlılığı/Etkisi*” standarttır. Bu standartın göstergelerinin uygunluğuna ilişkin alan uzmanlarının görüşlerini yansıtan sayısal veriler Tablo 17’de verilmiştir.

Tablo 17.

Alan Uzmanlarının Program Değerlendirmede “Yararlık” Standartlarından “Sonuçların Yararlılığı/Etkisi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Program değerlendirme sonuçlarının nasıl, nerede ve hangi amaçla kullanılacağını belirleme	101	74,3	20	14,7	3	2,2	1	,7	11	8,1	4,46	1,15
2. Program değerlendirme sonuçlarının paydaşlar yönünden olası yararlarını/etkilerini belirleme	96	70,6	24	17,6	4	2,9	1	,7	11	8,1	4,42	1,15
3. Program değerlendirme sonuçlarının paydaşlarda yeni bakış açısı/anlayış geliştirmesi	74	54,4	39	28,7	11	8,1	2	1,5	10	7,4	4,21	1,14
											4,37	

Tablo 17’de görüldüğü gibi, program değerlendirme standartlarından altında yer alan “*Yararlık*” standartlarının altında yer alan “*Sonuçların Yararlılığı/Etkisi*” standartının üç göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge, “Program değerlendirme sonuçlarının nasıl, nerede ve hangi amaçla kullanılacağını belirleme”dir. Alan uzmanlarının %89’u bu göstergelyi “uygun” bulurken, uzmanların %2,2’si bu göstergenin “kısmen uygun”, %8,8’i de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.46’dır. Bu ortalama, “Program değerlendirme sonuçlarının nasıl, nerede ve hangi amaçla kullanılacağını belirleme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “*Sonuçların Yararlılığı/Etkisi*” standartıyla ilgili ikinci gösterge “Program değerlendirme sonuçlarının paydaşlar

yönünden olası yararlarını/etkilerini belirleme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %88,2 “uygun”, %2,9 “kısmen uygun” ve %8,8 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.42’dir. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Program değerlendirme sonuçlarının paydaşlar yönünden olası yararlarını/etkilerini belirleme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Sonuçların Yararlılığı/Etkisi*” standartının son göstergesi, “Program değerlendirme sonuçlarının paydaşlarda yeni bakış açısı/anlayış geliştirmesi”dir. Alan uzmanlarının bu göstergeye ilişkin görüşleri incelendiğinde, uzmanların %83,1’inin göstergesi “uygun” bulduğu, %8,1’lik bir kesiminin “kısmen uygun” bulduğu ve %8,9’unun da uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.21’tir. Bu ortalamaya göre, bu göstergenin de uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Yararlık*” standartının beşinci alt standardı olan “*Sonuçların Yararlılığı/Etkisi*” standartının göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın üç göstergesini de uygun bulmuşlardır. “*Sonuçların Yararlılığı/Etkisi*” standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması 4.37 olarak bulunmuştur. Bu değer, alan uzmanlarının “*Sonuçların Yararlılığı/Etkisi*” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı göstergeleri ile birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

Program değerlendirmede “*Yararlık*” standart alanını oluşturan beş standartın altında yer alan standart göstergelerinin uygunluğuna ilişkin alan uzmanlarının görüşleriyle ilgili Tablo 13,14,15,16 ve 17’de verilen beş standartın göstergelerine ilişkin sayısal veriler bir bütün olarak değerlendirildiğinde, Tablo 18’de da görüldüğü gibi, uzmanlar “*Paydaşların Tanımlanması*” standartının altı göstergesinden dördünü, “*Değerlendirici Yeterliği ve Güvenirliği*” standartının altı göstergesinden beşini, “*Verilerin Kapsamı ve Seçimi /Yararlılığı*”, “*Raporun Kapsamlılığı ve Anlaşılabilirliği*” ve “*Sonuçların*

Yararlılığı/Etkisi” standartlarının tüm göstergelerini uygun bulmuşlardır. Buna karşılık uzmanlar *“Paydaşların Tanımlanması”* standartının iki ve *“Değerlendirici Yeterliği ve Güvenirliği”* standartının bir göstergesini uygun bulmamışlardır. Öte yandan, uzmanların standart göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalamaları da standartlara göre 3.72 (en düşük) ile 4.37 (en yüksek) arasında değişmektedir. Bu değerler, standartların uzmanlarca uygun bulunduğunu göstermektedir. Sonuç olarak alan uzmanları, beş *“Yararlık”* standartını ve bu standartlara ilişkin 21 göstergeyi Türkiye için “uygun” program değerlendirme standartları olarak kabul etmişlerdir.

Tablo 18.

Program Değerlendirmede “Yararlık” Standartlarının Aritmetik Ortalamaları

Standart	GS	\bar{X}	Uygun Bulunmayan Göstergeler	\bar{X}
1. Paydaşların Tanımlanması	6	3.72	1.5 Program değerlendirme araştırmasında sonuçların uygulanmasından sorumlu grup ya da bireyleri belirtme	2.45
			1.6 Program değerlendirme araştırmasında sonuçların yayınlanmasından sorumlu grup ya da bireyleri belirtme	2.42
2. Değerlendirici Yeterliği ve Güvenirliği	6	4.05	2.3 Bireysel sınırlılıkların farkına varma, bu sınırlılıkları aşan durumlarda uzman yardımı alma	2.23
3. Verilerin Kapsamı ve Seçimi /Yararlığı	2	4.31	Uygun bulunmayan gösterge yoktur	
4. Raporun Kapsamlılığı ve Anlaşılabilirliği	7	4.30		
5. Sonuçların Yararlılığı/Etkisi	3	4.37		

GS: Gösterge Sayısı

4.2. Alan Uzmanlarının Program Değerlendirmede “Yürütülebilirlik” Standartlarının Uygunluğuna İlişkin Görüşleri

Araştırmada yanıtı aranan ikinci alt soru, alan uzmanlarının program değerlendirme standartları taslağında yer alan ve *“Yürütülebilirlik”* standart alanını oluşturan alt standartların göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerinin ne olduğudur. Alan uzmanlarının görüşleri aşağıdaki dört alt standartın göstergeleri kapsamında incelenmiştir:

- *Uygun Yöntemler*
- *Grup ve Kurum Desteği*

- *Maliyet-Yarar Çözümlemesi*
- *Değerlendirme Araştırmasının Yönetimi*

Uzmanlarının görüşlerinin çözümlenmesinde sayı, yüzde ve aritmetik ortalamalar kullanılmış, verilerin yorumlanmasında ise yöntem bölümünde belirlenen sınırlar ve düzeylere göre aritmetik ortalamalar temel alınmıştır.

4.2.1. Alan Uzmanlarının Uygun Yöntemler Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “Yürütülebilirlik” standartlarını oluşturan dört standartın ilki “Uygun Yöntemler” standarttır. Alan uzmanlarının “Uygun Yöntemler” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 19’da verilmiştir.

Tablo 19.

Alan Uzmanlarının Program Değerlendirmede “Yürütülebilirlik” Standartlarından “Uygun Yöntemler” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Program değerlendirme araştırmasında seçilen yöntemin maliyet yönünden uygunluğunu inceleme	84	60,3	38	27,9	9	6,6	3	2,2	4	2,9	4.40	.92
2. Program değerlendirme araştırmasında seçilen yöntemin zaman/süre yönünden uygunluğunu inceleme	91	66,9	34	25,0	4	2,9	-	-	7	5,1	4.48	.96
3. Program değerlendirme araştırmasında seçilen yöntemin etik ilkeler yönünden uygunluğunu inceleme	102	75,0	19	14,0	9	6,6	-	-	6	4,4	4.55	.95
4. Program değerlendirme araştırmasında seçilen yöntemin güçlü ve zayıf yönlerini ortaya koyma	100	73,5	21	15,4	9	6,6	-	-	6	4,4	4.54	.95
											4.49	

Tablo 19’da görüldüğü gibi, program değerlendirme standartlarından “Yürütülebilirlik” standartlarının altında yer alan standartlardan biri olan “Uygun Yöntemler” standartının dört göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Program değerlendirme araştırmasında seçilen yöntemin maliyet yönünden

uygunluğunu inceleme”dir. Alan uzmanlarının %88,2’si bu göstergeyi “uygun” bulurken, uzmanların %6,6’sı bu göstergenin “kısmen uygun”, %5,1’i de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.40’dır. Bu ortalama, “Program değerlendirme araştırmasında seçilen yöntemin maliyet yönünden uygunluğunu inceleme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “*Uygun Yöntemler*” standartıyla ilgili ikinci gösterge “Program değerlendirme araştırmasında seçilen yöntemin zaman/süre yönünden uygunluğunu inceleme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %91,9 “uygun”, %2,9 “kısmen uygun” ve %5,1 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.48’dır. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Program değerlendirme araştırmasında seçilen yöntemin zaman/süre yönünden uygunluğunu inceleme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Uygun Yöntemler*” standartının üçüncü göstergesi, “Program değerlendirme araştırmasında seçilen yöntemin etik ilkeler yönünden uygunluğunu inceleme”dir. Araştırma anketini yanıtlayan uzmanların %89’u bu göstergenin “uygun” olduğu, %6,6’sı “kısmen uygun” olduğu ve %4,4’ü de uygun olmadığı yönünde görüş bildirmişlerdir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.55’dır. Bu ortalama, uzmanlara göre bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“*Uygun Yöntemler*” standartının dördüncü göstergesi, “Program değerlendirme araştırmasında seçilen yöntemin güçlü ve zayıf yönlerini ortaya koyma”dır. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %88,9’unun göstergeyi “uygun” bulduğu, %6,6’lık bir kesiminin “kısmen uygun” bulduğu ve %4,4’ünün de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzmanların yanıtlarının

aritmetik ortalaması 4.54'tür. Bu ortalamaya göre, bu göstergenin de uzmanlar tarafından Türkiye için "uygun" görüldüğü söylenebilir.

Araştırmaya katılan uzmanların büyük çoğunluğu "Yürütülebilirlik" standartının ilk alt standardı olan "Uygun Yöntemler" standartının dört göstergesini de uygun bulmuşlardır. Standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 4.49 ortalaması elde edilmiştir. Bu değer, alan uzmanlarının "Uygun Yöntemler" standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı göstergeleri ile birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.2.2. Alan Uzmanlarının Grup ve Kurum Desteği/Yararlığı Standartının Göstergelerine İlişkin Görüşleri

"Yürütülebilirlik" standartlarının ikincisi "Grup ve Kurum Desteği/Yararlığı" standardıdır. Alan uzmanlarının "Grup ve Kurum Desteği/Yararlığı" standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 20'de verilmiştir.

Tablo 20.

Alan Uzmanlarının Program Değerlendirmede "Yürütülebilirlik" Standartlarından "Grup Ve Kurum Desteği/Yararlığı" Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri
(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
	1. Program değerlendirme araştırması sürecine doğrudan ya da dolaylı katılacak kişi, kurum ve kuruluşların desteğini alma	88	64,7	30	22,1	9	6,6	3	2,2	6		
2. Program değerlendirme araştırmasının sonuçlarını kullanacak kişi, kurum ve kuruluşların desteğini alma	80	58,8	36	26,5	11	8,1	3	2,2	6	4,4	4.33	.81
3. Program değerlendirme sürecinden doğrudan ya da dolaylı etkilenebilecek kişi, kurum ve kuruluşların desteğini alma	35	25,7	8	5,9	4	2,9	22	16,2	67	49,3	2.42	1.70
4. Veri toplama kaynaklarının (kişi, grup, kurum ve kuruluşlar) desteğini alma	20	14,7	8	5,9	10	7,4	35	25,7	63	46,3	2.17	1.44
											3.33	

Tablo 20’de görüldüğü gibi, program değerlendirmede “Yürütülebilirlik” standartlarından biri olan “Grup ve Kurum Desteği/Yararlılığı” standartının dört göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Program değerlendirme araştırması sürecine doğrudan ya da dolaylı katılacak kişi, kurum ve kuruluşların desteğini alma”dır. Alan uzmanlarının %86,8’i bu göstergeyi “uygun” bulurken, %6,6’sı “kısmen uygun” bulmakta, %6,6’sı da bu göstergeyi uygun bulmamaktadır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.40’tır. Bu ortalama, “Program değerlendirme araştırması sürecine doğrudan ya da dolaylı katılacak kişi, kurum ve kuruluşların desteğini alma” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “Grup ve Kurum Desteği/Yararlılığı” standardıyla ilgili ikinci gösterge “Program değerlendirme araştırmasının sonuçlarını kullanacak kişi, kurum ve kuruluşların desteğini alma”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %85,3 “uygun”, %8,1 “kısmen uygun” ve %6,6 “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.33’dır. Bu ortalama göre, alan uzmanlarının büyük çoğunluğu, “Program değerlendirme araştırmasının sonuçlarını kullanacak kişi, kurum ve kuruluşların desteğini alma” göstergesinin Türkiye için “uygun” bir gösterge olduğu görüşündedir.

“Grup ve Kurum Desteği/Yararlılığı” standartının üçüncü göstergesi, “Program değerlendirme sürecinden doğrudan ya da dolaylı etkilenebilecek kişi, kurum ve kuruluşların desteğini alma”dır. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %31,6 “uygun”, % 2,9 “kısmen uygun” %65,5 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulmadığını göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 2.43’dir. Bu ortalama, bu göstergenin Türkiye için uygun bir gösterge olmadığını göstermektedir.

“Grup ve Kurum Desteği/Yararlılığı” standartının dördüncü göstergesi, “Veri toplama kaynaklarının (kişi, grup, kurum ve kuruluşlar) desteğini alma”dır. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %20,6’sının göstergesi “uygun” bulduğu, %7,4’lük bir kesiminin “kısmen uygun” bulduğu ve %72’sinin de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması ise 2.17’dir. Bu ortalamaya göre, bu göstergenin de uzmanlar tarafından Türkiye için uygun görülmediği söylenebilir.

“Yürütülebilirlik” standartının ikinci alt standardı olan “Grup ve Kurum Desteği/Yararlılığı” standartının göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın iki göstergesini uygun bulmuşlar, iki göstergesini ise uygun bulmamışlardır. Standartın uygunluk düzeyinin belirlenmesi için bu dört göstergenin aritmetik ortalamalarının ortalaması hesaplanmış ve 3.33 değeri bulunmuştur. Bu değer, “Grup ve Kurum Desteği/Yararlılığı” standartının uzmanlarca üçüncü ve dördüncü göstergeleri dışında diğer iki göstergesiyle birlikte Türkiye için “kısmen uygun” görüldüğünü göstermektedir.

4.2.3. Alan Uzmanlarının Maliyet-Yarar Çözümlemesi Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “Yürütülebilirlik” alanının üçüncü standardı “Maliyet-Yarar Çözümlemesi” standardıdır. Alan uzmanlarının “Maliyet-Yarar Çözümlemesi” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 21’de verilmiştir.

Tablo 21.

*Alan Uzmanlarının Program Değerlendirmede “Yürütülebilirlik” Standartlarından
“Maliyet-Yarar Çözümlemesi” Standartının Göstergelerinin Uygunluğuna İlişkin
Görüşleri*

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Program değerlendirme araştırmasının öngörülen süresini belirleme	95	69,9	28	20,6	7	5,1	-	-	6	4,4	4.51	,94
2. Program değerlendirme araştırmasının öngörülen maliyetini (insan kaynakları, araç-gereçler, eğitim, seyahat vb. yönünden) belirleme	94	69,1	27	19,9	9	6,6	-	-	6	4,4	4.49	,95
3. Program değerlendirme araştırmasının yarar-maliyet ilişkisini ortaya koyma	26	19,1	17	12,5	9	6,6	44	32,4	40	29,4	2.59	1,49
4. Çalışma bütçesini ayrıntılı olarak hazırlama	83	61,0	36	26,5	10	7,4	1	,7	6	4,4	4.39	,98
											3.99	

Tablo 21’de görüldüğü gibi, program değerlendirme standartları altında yer alan “Yürütülebilirlik” standartlarından biri olan “Maliyet-Yarar Çözümlemesi” standartının dört göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Program değerlendirme araştırmasının öngörülen süresini belirleme”dir. Alan uzmanlarının %90,5’i bu göstergeyi “uygun” bulurken, uzmanların %5,1’i bu göstergenin “kısmen uygun”, %4,4’i de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.51’dir. Bu ortalama, “Program değerlendirme araştırmasının öngörülen süresini belirleme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “Maliyet-Yarar Çözümlemesi” standartıyla ilgili ikinci gösterge, “Program değerlendirme araştırmasının öngörülen maliyetini (insan kaynakları, araç-gereçler, eğitim, seyahat vb. yönünden) belirleme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %89 “uygun”, %6,6 “kısmen uygun” ve %4,4 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.49’dur. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Program

değerlendirme araştırmasının öngörülen maliyetini (insan kaynakları, araç-gereçler, eğitim, seyahat vb. yönünden) belirleme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Maliyet-Yarar Çözümlemesi*” standartının üçüncü göstergesi, “Program değerlendirme araştırmasının yarar-maliyet ilişkisini ortaya koyma”dır. Bu göstergeye ilişkin yanıtlar incelendiğinde 43 uzmanın (%31,6) göstergenin “uygun” olduğu yönünde görüş bildirdiği, buna karşılık 84 (%61,8) uzmanın ise uygun olmadığı yönünde görüş bildirdiği görülmektedir. Bu göstergesi “kısmen uygun” bulan uzmanların sayısı dokuzdur (%6,6). Uzmanların bu göstergeye ilişkin yanıtlarının aritmetik ortalaması 2.59’dur. Bu değerlere göre, “Program değerlendirme araştırmasının yarar-maliyet ilişkisini ortaya koyma” göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiştir.

“*Maliyet-Yarar Çözümlemesi*” standartının dördüncü göstergesi, “Çalışma bütçesini ayrıntılı olarak hazırlama”dır. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %87,5’ünün göstergesi “uygun” bulduğu, %7,4’lük bir kesiminin “kısmen uygun” bulduğu ve %5,1’inin de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.39’tir. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Yürütülebilirlik*” standartının üçüncü alt standardı olan “*Maliyet-Yarar Çözümlemesi*” standartının göstergelerinin Türkiye için uygunluğuna ilişkin olarak görüşlerine başvuru alan uzmanları üç göstergesi uygun bulmuşlar, bir göstergesi ise uygun bulmamışlardır. “*Maliyet-Yarar Çözümlemesi*” standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 3.99 olarak bulunmuştur. Bu değer, alan uzmanlarının “*Maliyet-Yarar Çözümlemesi*” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standardı üçüncü gösterge dışında diğer göstergeleriyle birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.2.4. Alan Uzmanlarının Değerlendirme Araştırmasının Yönetimi Standartının Göstergelerine İlişkin Görüşleri

“Yürütülebilirlik” standartlarının sonuncusu “Değerlendirme Araştırmasının Yönetimi” standartıdır. Alan uzmanlarının “Değerlendirme Araştırmasının Yönetimi” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 22’de verilmiştir.

Tablo 22.

Alan Uzmanlarının Program Değerlendirmede “Yürütülebilirlik” Standartlarından “Değerlendirme Araştırmasının Yönetimi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Değerlendirilmesi amaçlanan programın değerlendirilebilirliğini ortaya koyma	89	65,4	25	18,4	11	8,1	3	2,2	8	5,9	4.35	1,11
2. Değerlendirme araştırmasının önerisini hazırlama	89	65,4	30	22,1	9	6,6	2	1,5	6	4,4	4.43	1,00
3. Zaman, süre, işlem ve kaynaklar yönünden değerlendirme çalışmasının yönetim planını hazırlama	96	70,6	23	16,9	8	5,9	2	1,5	7	5,1	4.46	1,04
4. Değerlendirme araştırmasının yürütme stratejilerini belirleme	91	66,9	31	22,8	6	4,4	2	1,5	6	4,4	4.46	,98
5. Değerlendirme çalışmasının sonuçlarının paydaşlarla paylaşılmasını ve raporlaştırılmasını sağlama	26	19,1	8	5,9	6	4,4	34	25,0	62	45,6	2.28	1,54
											4.00	

Tablo 22’de görüldüğü gibi, program değerlendirmede “Yürütülebilirlik” standartlarından biri olan “Değerlendirme Araştırmasının Yönetimi” standartının beş göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Değerlendirilmesi amaçlanan programın değerlendirilebilirliğini ortaya koyma”dır. Bu gösterge için alan uzmanlarının %83,8’i “uygun”, %8,1’i “kısmen uygun”, %8,1’i de “uygun değil” görüşünü belirtmişlerdir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.35’dir. Bu ortalama, “Değerlendirilmesi amaçlanan programın değerlendirilebilirliğini ortaya koyma” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “*Değerlendirme Araştırmasının Yönetimi*” standartıyla ilgili ikinci gösterge “Değerlendirme araştırmasının önerisini hazırlama”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %87,5 “uygun”, %6,6 “kısmen uygun” ve %5,9 “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.43’dir. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Değerlendirme araştırmasının önerisini hazırlama” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Değerlendirme Araştırmasının Yönetimi*” standartının üçüncü göstergesi, “Program değerlendirme araştırmasının araçlarının uygulanacağı grup ya da bireyleri tanımlama”dır. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %87,5 “uygun”, % 5,9 “kısmen uygun” %6,6 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergelyi uygun bulduğunu göstermektedir. Nitekim bu göstergelye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.46’dir. Bu ortalama, bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“*Değerlendirme Araştırmasının Yönetimi*” standartının dördüncü göstergesi, “Değerlendirme araştırmasının yürütme stratejilerini belirleme”dir. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %89,7’sinin göstergelyi “uygun” bulduğunu, %4,4’ünün “kısmen uygun” bulduğunu ve %5,9’unun da uygun bulmadığını görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.46’dir. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Değerlendirme Araştırmasının Yönetimi*” standartının beşinci göstergesi, “Değerlendirme çalışmasının sonuçlarının paydaşlarla paylaşılmasını ve raporlaştırılmasını sağlama”dır. Bu göstergelye ilişkin yanıtlar incelendiğinde, 34 uzmanın (%25) göstergenin “uygun” olduğu yönünde görüş bildirdiğini, buna karşılık 96 (%70,6) uzmanın ise uygun olmadığı yönünde görüş bildirdiğini görülmektedir. Bu göstergelyi “kısmen uygun” bulan uzmanların sayısı altıdır (%4,4). Uzmanların bu göstergelye ilişkin yanıtlarının aritmetik ortalaması 2.28’dir. Bu değerlere göre,

“Değerlendirme çalışmasının sonuçlarının paydaşlarla paylaşılması ve raporlaştırılmasını sağlama” göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiştir.

“Yürütülebilirlik” standartının dördüncü standardı olan “Değerlendirme Araştırmasının Yönetimi” standartının göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın dört göstergesini uygun bulmuşlar, bir göstergesini ise uygun bulmamışlardır. “Değerlendirme Araştırmasının Yönetimi” standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 4.00 olarak bulunmuştur. Bu değer, alan uzmanlarının “Değerlendirme Araştırmasının Yönetimi” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı beşinci gösterge dışında diğer göstergeleriyle birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

Program değerlendirmede “Yürütülebilirlik” standart alanını oluşturan dört standartın altında yer alan standart göstergelerinin uygunluğuna ilişkin alan uzmanlarının görüşleriyle ilgili Tablo 19,20, 21 ve 22’de verilen dört standartın göstergelerine ilişkin sayısal veriler bir bütün olarak değerlendirildiğinde, Tablo 23’te de görüldüğü gibi, uzmanlar “Uygun Yöntemler” standartının tüm göstergelerini, “Grup ve Kurum Desteği” standartının dört göstergesinden ikisini, “Maliyet-Yarar Çözümlemesi” standartının dört göstergesinden üçünü ve “Değerlendirme Araştırmasının Yönetimi” standartının beş göstergesinden dördünü uygun bulmuşlardır. Buna karşılık uzmanlar “Grup ve Kurum Desteği” standartının iki göstergesini, “Maliyet-Yarar Çözümlemesi” ve “Değerlendirme Araştırmasının Yönetimi” standartlarının birer göstergesini uygun bulmamışlardır. Öte yandan, uzmanların standart göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalamaları da standartlara göre 3.33 (en düşük) ile 4.49 (en yüksek) arasında değişmektedir. Bu değerlerden 3.33 aritmetik ortalamaya sahip “Grup ve Kurum Desteği” standardı alan uzmanları tarafından Türkiye için “kısmen uygun” diğer standartlar ise “uygun” bulunmuştur. Sonuç olarak alan uzmanları, dört “Yürütülebilirlik” standartını ve bu standartlara ilişkin 13 göstergesi Türkiye için uygun program değerlendirme standartları olarak kabul etmişlerdir.

Tablo 23.

Program Değerlendirmede “Yürütülebilirlik” Standartlarının Aritmetik Ortalamaları

Standart	GS	\bar{X}	Uygun Bulunmayan Göstergeler	\bar{X}
• Uygun Yöntemler	4	4.49	-	
• Grup ve Kurum Desteği	4	3.33	2.3 Program değerlendirme sürecinden doğrudan ya da dolaylı etkilenebilecek kişi, kurum ve kuruluşların desteğini alma	2.43
			2.4 Veri toplama kaynaklarının (kişi, grup, kurum ve kuruluşlar) desteğini alma	2.17
• Maliyet-Yarar Çözümlemesi	4	3.99	3.3 Program değerlendirme araştırmasının yarar-maliyet ilişkisini ortaya koyma	2.59
• Değerlendirme Araştırmasının Yönetimi	5	4.00	4.5 Değerlendirme çalışmasının sonuçlarının paydaşlarla paylaşılmasını ve raporlaştırılmasını sağlama	2.28

GS: Gösterge Sayısı

4.3. Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarının Uygunluğuna İlişkin Görüşleri

Araştırmada yanıtı aranan üçüncü soru, alan uzmanlarının program değerlendirme standartları taslağında yer alan ve “Uygunluk” standart alanını oluşturan alt standartların göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerinin ne olduğudur. Alan uzmanlarının görüşleri bu alanın aşağıdaki altı standartının göstergelerine göre incelenmiştir:

- *Resmi Sözleşme*
- *Etik Kurallara Uygunluk*
- *Adil Değerlendirme*
- *Şeffaf Raporlaştırma*
- *Sonuçların Açıklanması*
- *Finansal Uygunluk*

Uzmanlarının görüşlerinin çözümlenmesinde sayı, yüzde ve aritmetik ortalamalar kullanılmış, verilerin yorumlanmasında ise yöntem bölümünde belirlenen sınırlar ve düzeylere göre aritmetik ortalamalar temel alınmıştır.

4.3.1. Alan Uzmanlarının *Resmi Sözleşme* Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “Uygunluk” standartlarını oluşturan altı alt standarttan ilki “Resmi Sözleşme” standarttır. Alan uzmanlarının program değerlendirmede “Uygunluk” standartlarından “Resmi Sözleşme” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 24’te verilmiştir.

Tablo 24.

Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Resmi Sözleşme” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Sözleşmede paydaş bilgilerine yer verme	72	52,9	38	27,9	21	15,4	1	0,7	4	2,9	4,27	.95
2. Sözleşmede program değerlendirme araştırmasının gerekçesini, amacını ve yöntemini belirtme	88	64,7	21	15,4	16	11,8	9	6,6	2	1,5	4,35	1.02
3. Sözleşmede program değerlendirme araştırmasının zaman çizelgesine yer verme	83	61	31	22,8	15	11,0	5	3,7	2	1,5	4,38	.93
4. Sözleşmede tarafların (destekleyici ve kullanıcıların) sağlayacağı hizmetleri (personel, bilgi ve araç-gereçler) belirtme	76	55,9	24	17,6	16	11,8	19	14,0	1	0,7	4,14	1.13
5. Değerlendirme sürecinde kullanılacak hizmet (personel, danışmanlar, seyahat, kaynaklar, posta ve telefon, verilerin çözümlenmesi, baskı ve yayınlama) harcamalarını ayrıntılı olarak belirtme	73	53,7	39	28,7	17	12,5	3	2,2	4	2,9	4,28	.97
6. Süreç içinde hizmetlerin ve harcamaların ayrıntılı kayıtlarını tutma	20	14,7	19	14,0	13	9,6	18	13,2	66	48,5	2,33	1.54
											3,96	

Tablo 24’te görüldüğü gibi, program değerlendirmede “Uygunluk” standartlarından “Resmi Sözleşme” standartının altı göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Sözleşmede paydaş bilgilerine yer verme”dir. Alan uzmanlarının %80,8’i bu göstergeyi “uygun” bulurken, uzmanların %15,4’ü bu göstergenin “kısmen uygun”, %3,6’sı da uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.27’dir. Bu ortalama, “Sözleşmede paydaş bilgilerine yer verme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “*Resmi Sözleşme*” standartıyla ilgili ikinci gösterge “Sözleşmede program değerlendirme araştırmasının gerekçesini, amacını ve yöntemini belirtme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %80,1 “uygun”, %11,8 “kısmen uygun” ve %8,1 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergeyle ilgili yanıtlarının aritmetik ortalaması 4.35’tir. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Sözleşmede program değerlendirme araştırmasının gerekçesini, amacını ve yöntemini belirtme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Resmi Sözleşme*” standartının üçüncü göstergesi, “Sözleşmede program değerlendirme araştırmasının zaman çizelgesine yer verme”dir. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %83,8 “uygun”, % 11 “kısmen uygun” %5,2 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.38’dir. Bu ortalama, bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“*Resmi Sözleşme*” standartının dördüncü göstergesi, “Sözleşmede tarafların (destekleyici ve kullanıcıların) sağlayacağı hizmetleri (personel, bilgi ve araç-gereçler) belirtme”dir. Alan uzmanlarının bu göstergeyle ilgili yanıtları incelendiğinde, uzmanların %73,5’i için bu gösterge “uygun”, %11,8’i için “kısmen uygun” ve %14,7’si için de uygun değildir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.14’tür. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Resmi Sözleşme*” standartının beşinci göstergesi, “Değerlendirme sürecinde kullanılacak hizmet (personel, danışmanlar, seyahat, kaynaklar, posta ve telefon, verilerin çözümlenmesi, baskı ve yayınlama) harcamalarını ayrıntılı olarak belirtme”dir. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %82,4’ü bu göstergeyi “uygun” bulurken, %12,5’i uygun bulmamaktadır. Uzmanların % 5,1’i ise bu göstergenin “kısmen uygun” olduğu görüşündedir. Uzmanların görüşlerinin aritmetik

ortalaması 4.28'dir. Bu değerler, bu göstergenin de uzmanlar tarafından "uygun" bulunduğunu göstermektedir.

"Resmi Sözleşme" standartının altıncı göstergesi, "Süreç içinde hizmetlerin ve harcamaların ayrıntılı kayıtlarını tutma"dır. Bu göstergeye ilişkin yanıtlar incelendiğinde 39 uzmanın (%28,7) göstergenin uygun olduğu yönünde görüş bildirdiği, buna karşılık 84 (%61,7) uzmanın ise uygun olmadığı yönünde görüş bildirdiği görülmektedir. Bu göstergesi "kısmen uygun" bulan uzmanların sayısı 13'tür (%9,6). Uzmanların bu göstergeye ilişkin yanıtlarının aritmetik ortalaması 2.33'tir. Bu değerlere göre, "Süreç içinde hizmetlerin ve harcamaların ayrıntılı kayıtlarını tutma" göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiştir.

"Uygunluk" standartının ilk alt standardı olan "Resmi Sözleşme" standartının göstergelerinin Türkiye için uygunluğuna ilişkin olarak uzmanlar beş göstergesi uygun bulmuşlar, bir göstergesi ise uygun bulmamışlardır. "Resmi Sözleşme" standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 3.96 olarak bulunmuştur. Bu değer, alan uzmanlarının "Resmi Sözleşme" standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı altıncı gösterge dışında diğer göstergeleriyle birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.3.2. Alan Uzmanlarının Etik Kurallara Uygunluk Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede "Uygunluk" standartlarını oluşturan ikinci alt standart "Etik Kurallara Uygunluk" standartıdır. Alan uzmanlarının program değerlendirmede "Uygunluk" standartlarından "Etik Kurallara Uygunluk" standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 25'te verilmiştir.

Tablo 25.

Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Etik Kurallara Uygunluk” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Değerlendirme uygulamaları için resmi ve özel izin alma	96	70,6	18	13,2	17	12,5	3	2,2	2	1,5	4.49	.90
2. Katılımcılara değerlendirme sürecinde sahip oldukları temel hakları bildirme	94	69,1	20	14,7	11	8,1	9	6,6	2	1,5	4.43	.99
3. Katılımcıların kişisel özelliklerine, kültürel ve sosyal değerlerine saygı gösterme	101	74,3	22	16,2	9	6,6	1	0,7	3	2,2	4.60	.83
4. Veri sağlayıcıların ve katılımcıların gizliliğini (kimlik vb) güvence altına alma	101	74,3	18	13,2	12	8,8	1	0,7	4	2,9	4.55	.91
5. Araştırma sürecinde eğitim hakkına özen gösterme (özellikle deneysel araştırmalarda)	86	63,2	26	19,1	16	11,8	1	0,7	7	5,1	4.35	1.06
6. Katılımcı bilgilerini değerlendirme amacı dışında kurum, grup ya da birey hizmetine sunmama güvencesi verme	97	71,3	21	15,4	12	8,8	2	1,5	4	2,9	4.51	.93
7. Verileri yalnızca bilimsel süreç ve sonuçlar için kullanma	106	77,9	11	8,1	9	6,6	3	2,2	7	5,1	4.51	1.07
8. Finansal kaynakların kullanımı sürecine ilişkin bilgi ve belgeleri paydaşlarla paylaşma	71	52,2	35	25,7	23	16,9	3	2,2	4	2,9	4.22	1.00
											4,46	

Tablo 25’te görüldüğü gibi, program değerlendirmede “Uygunluk” standartlarından biri olan “Etik Kurallara Uygunluk” standartının sekiz göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Değerlendirme uygulamaları için resmi ve özel izin alma”dır. Alan uzmanlarının %83,8’i bu göstergeyi “uygun” bulurken, %12,5’i bu göstergenin “kısmen uygun”, %3,7’si de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.49’dur. Bu ortalama, “Değerlendirme uygulamaları için resmi ve özel izin alma” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “Etik Kurallara Uygunluk” standartıyla ilgili ikinci gösterge “Katılımcılara değerlendirme sürecinde sahip oldukları temel hakları bildirme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının

dağılımı; %83,8 “uygun”, %8,1 “kısmen uygun” ve %8,1 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.43’tür. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Katılımcılara değerlendirme sürecinde sahip oldukları temel hakları bildirme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Etik Kurallara Uygunluk*” standartının üçüncü göstergesi, “Katılımcıların kişisel özelliklerine, kültürel ve sosyal değerlerine saygı gösterme”dir. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %90,5 “uygun”, % 6,6 “kısmen uygun” ve %2,9 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.60’tır. Bu ortalama, bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“*Etik Kurallara Uygunluk*” standartının dördüncü göstergesi, “Veri sağlayıcıların ve katılımcıların gizliliğini (kimlik vb.) güvence altına alma”dır. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %87,5’inin göstergeyi “uygun” bulduğunu, %8,8’lik bir kesiminin “kısmen uygun” bulduğunu ve %3,6’sının da uygun bulmadığını görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.55’tir. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Etik Kurallara Uygunluk*” standartının beşinci göstergesi, “Araştırma sürecinde eğitim hakkına özen gösterme (özellikle deneysel araştırmalarda)”dır. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %82,3’ü bu göstergeyi “uygun” bulurken, %11,8’i uygun bulmamaktadır. Uzmanların % 5,8’i de bu göstergenin “kısmen uygun” olduğu görüşündedir. Uzmanların görüşlerinin aritmetik ortalaması 4.35’tir. Bu değerler, bu göstergenin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

“*Etik Kurallara Uygunluk*” standartının altıncı göstergesi,, “Katılımcı bilgilerini değerlendirme amacı dışında kurum, grup ya da birey hizmetine sunmama güvencesi

verme”dir. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %86,7’sinin göstergesi “uygun” bulunduğu, %8,8’lik bir kesiminin “kısmen uygun” bulunduğu ve %4,4’ünün de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.51’tir. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

Alan uzmanlarından görüş bildirmeleri istenen “*Etik Kurallara Uygunluk*” standartıyla ilgili yedinci gösterge “Verileri yalnızca bilimsel süreç ve sonuçlar için kullanma”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %86 “uygun”, %6,6 “kısmen uygun” ve %7,3 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.51’dir. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Verileri yalnızca bilimsel süreç ve sonuçlar için kullanma” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Etik Kurallara Uygunluk*” standartının son göstergesi, “Finansal kaynakların kullanımı sürecine ilişkin bilgi ve belgeleri paydaşlarla paylaşma”dır. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %77,9’u bu göstergesi “uygun” bulurken, %16,9’u uygun bulmamaktadır. Uzmanların %5,1’i bu göstergenin “kısmen uygun” olduğu görüşündedir. Uzmanların görüşlerinin aritmetik ortalaması 4.22’dir. Bu değerler, bu göstergenin de uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

“*Uygunluk*” standartının ikinci alt standardı olan “*Etik Kurallara Uygunluk*” standartının göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın sekiz göstergesini de uygun bulmuşlardır. Standartın uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 4.46 olarak bulunmuştur. Bu değer, alan uzmanlarının “*Etik Kurallara Uygunluk*” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı tüm göstergeleri ile birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.3.3. Alan Uzmanlarının Adil Değerlendirme Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “Uygunluk” standartlarını oluşturan üçüncü alt standart “Adil Değerlendirme” standarttır. Alan uzmanlarının “Adil Değerlendirme” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 26’da verilmiştir.

Tablo 26.

Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Adil Değerlendirme” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Orduka Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Programın güçlü yanlarını açık bir biçimde belirtme	102	75,0	20	14,7	11	8,1	1	0,7	2	1,5	4.61	.79
2. Programın zayıf yanlarını açık bir biçimde belirtme	104	76,5	19	14,0	10	7,4	1	0,7	2	1,5	4.63	.78
3. Paydaşların beklenti ve isteklerinden etkilenmemesi	22	16,2	12	8,8	12	8,8	20	14,7	70	51,5	2.24	1.54
4. Uygulamanın paydaş taleplerinden bağımsız yürütülmesi	79	58,1	27	19,9	21	15,4	5	3,7	4	2,9	4.26	1.04
5. Program değerlendirmecinin uygulama sürecinde etik ilkelere bağlı kalması	111	81,6	13	9,6	8	5,9	1	0,7	3	2,2	4.68	.81
6. Uygulama sürecinde değerlendirme uzmanının politik, ideolojik ve kültürel anlayışını paydaşlara yansıtmaması	104	76,5	16	11,8	11	8,1	1	0,7	4	2,9	4.58	.90
											4.25	

Tablo 26’da görüldüğü gibi, program değerlendirme standartları altında yer alan “Uygunluk” standartlarından biri olan “Adil Değerlendirme” standartının altı göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Programın güçlü yanlarını açık bir biçimde belirtme”dir. Alan uzmanlarının %89,7’si bu göstergenin “uygun” olduğu, %8,1’i “kısmen uygun”, %2,2’si de uygun olmadığı yönünde görüş bildirmişlerdir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.61’dir. Bu ortalama, “Programın güçlü yanlarını açık bir biçimde belirtme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “*Adil Değerlendirme*” standartıyla ilgili ikinci gösterge “Programın zayıf yanlarını açık bir biçimde belirtme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %90,5 “uygun”, %7,4 “kısmen uygun” ve %2,2 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.63’tür. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Programın zayıf yanlarını açık bir biçimde belirtme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Adil Değerlendirme*” standartının üçüncü göstergesi, “Paydaşların beklenti ve isteklerinden etkilenmemesi”dir. Bu göstergeye ilişkin yanıtlar incelendiğinde 34 uzmanın (%25) göstergenin “uygun” olduğu yönünde görüş bildirdiği, buna karşılık 90 (%66,2) uzmanın ise uygun olmadığı yönünde görüş bildirdiği görülmektedir. Bu göstergeyi “kısmen uygun” bulan uzmanların sayısı 12’dir (%8,8). Uzmanların bu göstergeye ilişkin yanıtlarının aritmetik ortalaması 2.24’tür. Bu değerlere göre, “Paydaşların beklenti ve isteklerinden etkilenmemesi” göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiştir.

“*Adil Değerlendirme*” standartının dördüncü göstergesi, “Uygulamanın paydaş taleplerinden bağımsız yürütülmesi”dir. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %78’inin göstergeyi “uygun” bulduğu, %15,4’lük bir kesiminin “kısmen uygun” bulduğu ve %6,6’sı da uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.26’dır. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Adil Değerlendirme*” standartının beşinci göstergesi, “Program değerlendirmecinin uygulama sürecinde etik ilkelere bağlı kalması”dır. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %91,2’si bu göstergeyi “uygun” bulurken, %2,9’i uygun bulmamaktadır. Uzmanların %5,9’i bu göstergenin “kısmen uygun” olduğu görüşündedir. Uzmanların görüşlerinin aritmetik ortalaması 4.68’dir. Bu değerler, bu göstergenin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

“*Adil Değerlendirme*” standartının altıncı göstergesi,, “Uygulama sürecinde değerlendirme uzmanının politik, ideolojik ve kültürel anlayışını paydaşlara

yansıtması”dır. Bu göstergenin uygunluđuna ilişkin arařtırma anketini yanıtlayan uzmanların grřlerinin dađılımı; %88,3 “uygun”, %8,1 “kısmen uygun” %3,6 "uygun deđil" řeklinde-dir. Alan uzmanlarının yanıtlarının bu dađılımı, uzmanların byk ođunluđunun bu gstergeyi uygun bulduđunu gstermektedir. Nitekim bu gstergeye ilişkin uzmanların grřlerini yansıtan yanıtların aritmetik ortalaması da 4.58’dir. Bu ortalama, bu gstergenin Trkiye iin “uygun” bir gsterge olduđunu gstermektedir.

“*Uygunluk*” standartının nc alt standardı olan “*Adil Deđerlendirme*” standartının gstergelerinin Trkiye iin uygunluđuna ilişkin olarak uzmanlar beř gstergeyi uygun bulmuřlar, bir gstergeyi ise uygun bulmamıřlardır. Standartın uzmanlar tarafından uygun bulunup bulunmadıđının saptanması iin de uzmanların bu standartın gstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmıř ve 4.25 olarak bulunmuřtur. Bu deđer, alan uzmanlarının “*Adil Deđerlendirme*” standartının uygunluđuna ilişkin olarak olumlu grř belirttikleri ve bu standartı nc gsterge dıřında diđer gstergeleriyle birlikte Trkiye iin program deđerlendirme standartı ve gstergeleri olarak kabul ettiklerini gstermektedir.

4.3.4. Alan Uzmanlarının *řeffaf Raporlařtırma* Standartının Gstergelerine İliřkin Grřleri

Alan uzmanlarının program deđerlendirmede “*Uygunluk*” standartlarından “*řeffaf Raporlařtırma*” standartının gstergelerinin uygunluđuna ilişkin grřlerini yansıtan sayısal veriler Tablo 27’de verilmiřtir.

Tablo 27.

Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Şeffaf Raporlaştırma” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Program değerlendirme raporunun paydaş taleplerinden bağımsız olması	83	61,0	26	19,1	19	14,0	6	4,4	2	1,5	4,34	.97
2. Program değerlendirme raporunun değerlendiricinin nesnellliğini yansıtması	100	73,5	14	10,3	14	10,3	2	1,5	6	4,4	4,47	.92
3. Program değerlendirme raporunun tarafsız bir değerlendirmeci tarafından incelenmesi	99	72,8	16	11,8	9	6,6	1	0,7	11	8,1	4,40	.84
											4,40	

Tablo 27’de görüldüğü gibi, program değerlendirmede “Uygunluk” standartlarından biri olan “Şeffaf Raporlaştırma” standartının üç göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Program değerlendirme raporunun paydaş taleplerinden bağımsız olması”dır. Alan uzmanlarının %80,1’i bu göstergeyi “uygun” bulurken, uzmanların %14’ü bu göstergenin “kısmen uygun”, %5,9’u da uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.34’tür. Bu ortalama, “Program değerlendirme raporunun paydaş taleplerinden bağımsız olması” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “Şeffaf Raporlaştırma” standartıyla ilgili ikinci gösterge “Program değerlendirme raporunun değerlendiricinin nesnellliğini yansıtması”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %83,8 “uygun”, %10,3 “kısmen uygun” ve %5,9 “uygun değil” şeklindedir. Uzmanların bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.47’dir. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Program değerlendirme raporunun değerlendiricinin nesnellliğini yansıtması” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Şeffaf Raporlaştırma*” standartının üçüncü göstergesi, “Program değerlendirme raporunun tarafsız bir değerlendirmeci tarafından incelenmesi”dir. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %84,6’sının göstergelyi “uygun” bulduđu, %6,6’sının “kısmen uygun” bulduđu ve %8,8’inin de uygun bulmadıđı görölmektedir. Bu göstergenin uygunluđuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.40’tır. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” göröldüđu söylenebilir.

“*Uygunluk*” standartının dördüncü standartı olan “*Şeffaf Raporlaştırma*” standartının göstergelerinin Türkiye için uygunluđuna ilişkin görüşlerine başvurulanan uzmanlar, bu standartın üç göstergelyi de uygun bulmuşlardır. “*Şeffaf Raporlaştırma*” standartının uzmanlar tarafından uygun bulunup bulunmadıđının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 4.40 olarak bulunmuştur. Bu deđer, alan uzmanlarının “*Şeffaf Raporlaştırma*” standartının uygunluđuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı tüm göstergeleri birlikte Türkiye için program değerlendirme standartı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.3.5. Alan Uzmanlarının *Sonuçların Açıklanması* Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “*Uygunluk*” standartlarının beşincisi “*Sonuçların Açıklanması*” standartıdır. Alan uzmanlarının “*Sonuçların Açıklanması*” standartının göstergelerinin uygunluđuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 28’de verilmiştir.

Tablo 28.

*Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından
“Sonuçların Açıklanması” Standartının Göstergelerinin Uygunluğuna İlişkin
Görüşleri*

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Çözümlemeler sonucu ulaşılan tüm bulguları eksiksiz olarak yazılı ifade etme	100	73,5	20	14,7	12	8,8	2	1,5	2	1,5	4.57	.83
2. Ulaşılan sonuçların ve sınırlılıkların dayanaklarını belirtme	21	15,4	19	14,0	26	19,1	2	1,5	68	50,0	2.43	1.57
3. Sonuçları alan uzmanları ile tartışma	100	73,5	22	16,2	11	8,1	1	0,7	2	1,5	4.60	.79
4. Sonuçları paydaşlarla tartışma	83	61,0	36	26,5	14	10,3	1	0,7	2	1,5	4.45	.82
5. Sonuçları karar vericilerle tartışma	93	68,4	28	20,6	12	8,8	1	0,7	2	1,5	4.54	.81
6. Sonuçları ilgili kesimlere ulaştırma	94	69,1	25	18,4	12	8,8	1	0,7	4	2,9	4.50	.91
											4.18	

Tablo 28’de görüldüğü gibi, program değerlendirme standartları altında yer alan “Uygunluk” standartlarından biri olan “Sonuçların Açıklanması” standartının altı göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Çözümlemeler sonucu ulaşılan tüm bulguları eksiksiz olarak yazılı ifade etme”dir. Alan uzmanlarının %88,2’si bu göstergeyi “uygun”, %8,8’i de “kısmen uygun” bulurken, %3’ü uygun bulmamıştır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.57’dir. Bu ortalama, “Çözümlemeler sonucu ulaşılan tüm bulguları eksiksiz olarak yazılı ifade etme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

“Sonuçların Açıklanması” standartının ikinci göstergesi, “Ulaşılan sonuçların ve sınırlılıkların dayanaklarını belirtme”dir. Bu göstergeye ilişkin yanıtlar incelendiğinde, 40 uzmanın (%29,4) göstergenin “uygun” olduğu yönünde görüş bildirdiği, buna karşılık 70 (%51,5) uzmanın ise uygun olmadığı yönünde görüş bildirdiği görülmektedir. Bu göstergeyi kısmen uygun bulan uzmanların sayısı 26’dır (%19,1). Uzmanların bu göstergeye ilişkin yanıtlarının aritmetik ortalaması 2.43’tür. Bu

değerlere göre, “Ulaşılan sonuçların ve sınırlılıkların dayanaklarını belirtme” göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiştir.

Alan uzmanlarından görüş bildirmeleri istenen “*Sonuçların Açıklanması*” standartıyla ilgili üçüncü gösterge “Sonuçları alan uzmanları ile tartışma”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %89,7 “uygun”, %8,1 “kısmen uygun” ve %2,2 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması 4.60’tır. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Sonuçları alan uzmanları ile tartışma” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Sonuçların Açıklanması*” standartının dördüncü göstergesi, “Sonuçları paydaşlarla tartışma”dır. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %87,5 “uygun”, % 10,3 “kısmen uygun” %2,2 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.45’tir. Bu ortalama, bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“*Sonuçların Açıklanması*” standartının beşinci göstergesi, “Sonuçları karar vericilerle tartışma”dır. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %89’unun göstergeyi “uygun” bulduğu, %8,8’lik bir kesiminin “kısmen uygun” bulduğu ve %2,2’inin de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.54’tür. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Sonuçların Açıklanması*” standartının altıncı ve son göstergesi “Sonuçları ilgili kesimlere ulaştırma”dır. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %87,5’i bu göstergeyi “uygun” bulurken, %3,6’sı uygun bulmamaktadır. Uzmanların % 8,8’i bu göstergenin “kısmen uygun” olduğu görüşündedir. Uzmanların görüşlerinin aritmetik ortalaması ise 4.50’dir. Bu değerler, bu göstergenin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

“Uygunluk” standartının beşinci alt standardı olan “Sonuçların Açıklanması” standartının göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın beş göstergesini uygun bulmuşlar, bir göstergesini ise uygun bulmamışlardır. Standartın uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 4.18 olarak bulunmuştur. Bu değer, alan uzmanlarının “Sonuçların Açıklanması” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı ikinci gösterge dışında diğer göstergeleriyle birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.3.6. Alan Uzmanlarının Finansal Uygunluk Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “Uygunluk” standartlarının sonuncusu “Finansal Uygunluk” standardıdır. Alan uzmanlarının program değerlendirmede “Uygunluk” standartlarından “Finansal Uygunluk” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 29’da verilmiştir.

Tablo 29.

Alan Uzmanlarının Program Değerlendirmede “Uygunluk” Standartlarından “Finansal Uygunluk” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Değerlendirme gelir-gider bütçesini belirtme	76	55,9	31	22,8	18	13,2	2	1,5	9	6,6	4.20	1.15
2. Değerlendirme için sağlanan finansman kaynaklarını belirtme	77	56,6	34	25,0	7	5,1	5	3,7	13	9,6	4.15	1.27
3. Değerlendirme süreci içinde yapılan harcamaları belirtme	66	48,5	38	27,9	16	11,8	4	2,9	12	8,8	4.04	1.23
4. Gelir-gider bütçesinin raporlarını gerektiğinde değerlendirme raporuna ekleme	76	55,9	25	18,4	3	2,2	3	2,2	29	21,3	3.85	1.61
											4.06	

Tablo 29’da görüldüğü gibi, program değerlendirme standartları altında yer alan “Uygunluk” standartlarının sonuncusu olan “Finansal Uygunluk” standartının dört

göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Değerlendirme gelir-gider bütçesini belirtme”dir. Alan uzmanlarının %78,7’si bu göstergeyi “uygun” bulurken, uzmanların %13,2’si bu göstergenin “kısmen uygun”, %8,1’i de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.20’dir. Bu ortalama, “Değerlendirme gelir-gider bütçesini belirtme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından “*Finansal Uygunluk*” standartıyla ilgili görüş bildirmeleri istenen ikinci gösterge “Değerlendirme için sağlanan finansman kaynaklarını belirtme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %81,6 “uygun”, %5,1 “kısmen uygun” ve %13,3de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.15’tir. Bu ortalama göre, alan uzmanlarının büyük çoğunluğu, “Değerlendirme için sağlanan finansman kaynaklarını belirtme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Finansal Uygunluk*” standartının üçüncü göstergesi, “Değerlendirme süreci içinde yapılan harcamaları belirtme”dir. Bu göstergenin uygunluğuna ilişkin olarak araştırma anketini yanıtlayan uzmanların %76,42’ü “uygun”, % 11,8’i “kısmen uygun” ve %11,7’si "uygun değil" yanıtını vermiştir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim alan uzmanlarının bu göstergeye ilişkin yanıtlarının aritmetik ortalaması da 4.04’tür. Bu ortalama, bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“*Finansal Uygunluk*” standartının dördüncü göstergesi, “Gelir-gider bütçesinin raporlarını gerektiğinde değerlendirme raporuna ekleme”dir. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %74,3’ünün göstergeyi “uygun” bulduğu, %2,2’lik bir kesiminin “kısmen uygun” bulduğu ve %23,5’inin de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 3.85’tir. Bu ortalama göre, bu göstergenin de uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“Uygunluk” standartının beşinci ve son alt standartı olan “Finansal Uygunluk” standartının göstergelerinin Türkiye için uygunluğuna ilişkin olarak uzmanlar dört göstergeyi de uygun bulmuşlardır. “Finansal Uygunluk” standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 4.06 olarak bulunmuştur. Bu değer, alan uzmanlarının “Finansal Uygunluk” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı göstergeleri ile birlikte Türkiye için program değerlendirme standartı ve göstergeleri olarak kabul ettiklerini göstermektedir.

Program değerlendirmede “Uygunluk” standart alanını oluşturan altı standartın altında yer alan standart göstergelerinin uygunluğuna ilişkin alan uzmanlarının görüşleriyle ilgili Tablo 24, 25, 26, 27, 28 ve 29’da verilen altı standartın göstergelerine ilişkin sayısal veriler bir bütün olarak değerlendirildiğinde, Tablo 30’da da görüldüğü gibi, uzmanlar “Resmi Sözleşme” standartının altı göstergesinden beşini, “Etik Kurallara Uygunluk” standartının sekiz göstergesinin tümünü, “Adil Değerlendirme” standartının altı göstergesinden beşini, “Şeffaf Raporlaştırma” standartının tüm göstergelerini, “Sonuçların Açıklanması” standartının altı göstergesinden beşini ve “Finansal Uygunluk” standartının dört göstergesinin tamamını uygun bulmuşlardır. Buna karşılık uzmanlar “Etik Kurallara Uygunluk”, “Adil Değerlendirme” ve “Sonuçların Açıklanması” standartlarının birer göstergesini uygun bulmamışlardır. Öte yandan, uzmanların standart göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalamaları da standartlara göre 3.96 (en düşük) ile 4.46 (en yüksek) arasında değişmektedir. Bu değerler, standartların uzmanlarca uygun bulunduğunu göstermektedir. Sonuç olarak alan uzmanları, altı “Uygunluk” standartını ve bu standartlara ilişkin 30 göstergeyi Türkiye için uygun program değerlendirme standartları olarak kabul etmişlerdir.

Tablo 30.

Program Değerlendirmede “Uygunluk” Standartlarının Aritmetik Ortalamaları

Standart	GS	\bar{X}	Uygun Bulunmayan Göstergeler	\bar{X}
1. Resmi Sözleşme	6	3.96	1.6 Süreç içinde hizmetlerin ve harcamaların ayrıntılı kayıtlarını tutma	2.33
2. Etik Kurallara Uygunluk	8	4.46		
3. Adil Değerlendirme	6	4.25	3.3 Paydaşların beklenti ve isteklerinden etkilenmemesi	2.24
4. Şeffaf Raporlaştırma	3	4.40		
5. Sonuçların Açıklanması	6	4.18	5.2 Ulaşılan sonuçların ve sınırlılıkların dayanaklarını belirtme	2.43
6. Finansal Uygunluk	4	4.06		

GS: Gösterge Sayısı

4.4. Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarının Uygunluğuna İlişkin Görüşleri

Araştırmada yanıtı aranan dördüncü soru, alan uzmanlarının program değerlendirme standartları taslağındaki “Doğruluk” standart alanını oluşturan alt standartların göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerinin ne olduğudur. Alan uzmanlarının görüşleri aşağıdaki sekiz alt standartın göstergeleri kapsamında incelenmiştir:

- *Değerlendirilecek Programın Tanımlanması*
- *Durum Çözümlemesi*
- *Değerlendirme Araştırmasının Amacının Belirlenmesi*
- *Değerlendirme Araştırmasının Yönteminin Belirlenmesi*
- *Veri Kaynaklarının Tanımlanması*
- *Verilerin Geçerliliği ve Güvenirliği*
- *Verilerin Çözümlemesi*
- *Sonuçların Gerekçelenendirilmesi*

Uzmanlarının görüşlerinin çözümlenmesinde sayı, yüzde ve aritmetik ortalamalar kullanılmış, verilerin yorumlanmasında ise yöntem bölümünde belirlenen sınırlar ve düzeylere göre aritmetik ortalamalar temel alınmıştır.

4.4.1. Alan Uzmanlarının Değerlendirilecek Programın Tanımlanması Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “Doğruluk” standartlarını oluşturan sekiz alt standarttan ilki “Değerlendirilecek Programın Tanımlanması” standarttır. Alan uzmanlarının “Değerlendirilecek Programın Tanımlanması” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 31’de verilmiştir.

Tablo 31.

Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Değerlendirilecek Programın Tanımlanması” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Programla ilgili farklı kaynaklardan bilgi ve belgeleri toplama	117	86,0	15	11,0	1	,7	2	1,5	1	,7	4.80	.59
2. Programı açık, anlaşılır ve eksiksiz tanımlama	86	63,2	28	20,6	18	13,2	3	2,2	1	,7	4.43	.85
3. Programın amaçlarını belirleme	102	75,0	20	14,7	10	7,4	3	2,2	1	,7	4.61	.78
4. Programın var olan görevlerini belirleme	26	19,1	53	39,0	45	33,1	11	8,1	1	,7	3.68	.60
											4.38	

Tablo 31’de görüldüğü gibi, program değerlendirmede “Doğruluk” standartlarından “Değerlendirilecek Programın Tanımlanması” standartının dört göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Programla ilgili farklı kaynaklardan bilgi ve belgeleri toplama”dır. Alan uzmanlarının %97’si bu göstergeyi “uygun”, %0,7’si “kısmen uygun” bulurken, %2,2’si uygun bulmamaktadır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.80’dir. Bu ortalama, “Programla ilgili farklı kaynaklardan bilgi ve belgeleri toplama” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “Değerlendirilecek Programın Tanımlanması” standartıyla ilgili ikinci gösterge “Programı açık, anlaşılır ve eksiksiz tanımlama”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %83,8 “uygun”, %13,2 “kısmen uygun” ve %2,9 de “uygun değil”

şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.43'tür. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, "Programı açık, anlaşılır ve eksiksiz tanımlama" göstergesinin Türkiye için "uygun" olduğu görüşündedir.

"Değerlendirilecek Programın Tanımlanması" standartının üçüncü göstergesi, "Programın amaçlarını belirleme"dır. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %89,7 "uygun", %7,4 "kısmen uygun" %2,9 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.61'dir. Bu ortalama, bu göstergenin Türkiye için "uygun" bir gösterge olduğunu göstermektedir.

"Değerlendirilecek Programın Tanımlanması" standartının dördüncü göstergesi, "Programın var olan görevlerini belirleme"dir. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %58,1'inin göstergeyi "uygun" bulduğunu, %33,1'lik bir kesiminin "kısmen uygun" bulduğunu ve %8,8'inin de uygun bulmadığını görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 3.68'dir. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için "uygun" görüldüğü söylenebilir.

"Doğruluk" standartının ilk alt standardı olan *"Değerlendirilecek Programın Tanımlanması"* standartının göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın dört göstergesini de uygun bulmuşlardır. Standartın uzmanlar tarafından uygun bulunup bulunmadığının saptanması amacıyla uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 4.38 olarak bulunmuştur. Bu değer, alan uzmanlarının *"Değerlendirilecek Programın Tanımlanması"* standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı tüm göstergeleri ile birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.4.2. Alan Uzmanlarının *Durum Çözümlemesi* Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “*Doğruluk*” standartlarını oluşturan sekiz alt standarttan ikincisi “*Durum Çözümlemesi*” standarttır. Alan uzmanlarının “*Durum Çözümlemesi*” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 32’de verilmiştir.

Tablo 32.

Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Durum Çözümlemesi” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Kurum kayıtları, paydaş ve destekleyici görüşleri, raporlar vb. kaynaklardan yararlanarak programın durumunu betimleme	71	52,2	29	21,3	27	19,9	5	3,7	4	2,9	4.16	1.06
2. Programın felsefi ve kuramsal temelleri ile yasal, sosyal ve ekonomik boyutlarını belirleme	99	72,8	24	17,6	10	7,4	2	1,5	1	,7	4.60	.75
3. Program değerlendirmeyi etkileyebilecek olası etmenleri belirleme	68	50,0	41	30,1	16	11,8	7	5,1	4	2,9	4.19	1.03
											4.32	

Tablo 32’de görüldüğü gibi, program değerlendirme standartları altında yer alan “*Doğruluk*” standartlarından biri olan “*Durum Çözümlemesi*” standartının üç göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Kurum kayıtları, paydaş ve destekleyici görüşleri, raporlar vb. kaynaklardan yararlanarak programın durumunu betimleme”dir. Alan uzmanlarının %73,5’i bu göstergeyi “uygun” bulurken, uzmanların %19,9’u kısmen uygun olduğu, %6,6’i de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.16’dır. Bu ortalama, “Kurum kayıtları, paydaş ve destekleyici görüşleri, raporlar vb. kaynaklardan yararlanarak programın durumunu betimleme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “*Durum Çözümlemesi*” standartıyla ilgili ikinci gösterge “Programın felsefi ve kuramsal temelleri ile yasal, sosyal ve

ekonomik boyutlarını belirleme”dir. Alan uzmanlarının bu göstergenin uygunluđuna ilişkin yanıtlarının dađılımlı; %90,4 “uygun”, %7,4 “kısmen uygun” ve %2,2 de “uygun deđil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluđuna ilişkin yanıtlarının aritmetik ortalaması ise 4.60’dır. Bu ortalamaya gre, alan uzmanlarının byk ođunluđu, “Programın felsefi ve kuramsal temelleri ile yasal, sosyal ve ekonomik boyutlarını belirleme” göstergesinin Trkiye iin “uygun” olduđu grşndedir.

“Durum zmlemesi” standartının nc göstergesi, “Program deđerlendirmeyi etkileyebilecek olası etmenleri belirleme”dir. Bu göstergenin uygunluđuna ilişkin arařtırma anketini yanıtlayan uzmanların grşlerinin dađılımlı; %80,1 “uygun”, % 11,8 “kısmen uygun” %8 "uygun deđil" şeklindedir. Alan uzmanlarının yanıtlarının bu dađılımlı, uzmanların byk ođunluđunun bu gstergiyi uygun bulduđunu gstermektedir. Nitekim bu gstergeye ilişkin uzmanların grşlerini yansıtan yanıtların aritmetik ortalaması da 4.19’dır. Bu ortalama, bu gstergenin Trkiye iin “uygun” bir gsterge olduđunu gstermektedir.

“Dođruluk” standartının ilk alt standardı olan “Durum zmlemesi” standartının gstergelerinin Trkiye iin uygunluđuna ilişkin olarak uzmanlar drt gstergiyi de uygun bulmuřlardır. “Durum zmlemesi” standartının uzmanlar tarafından uygun bulunup bulunmadıđının saptanması iin de uzmanların bu standartın gstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması hesaplanmış ve 4.32 olarak bulunmuřtur. Bu deđer, alan uzmanlarının “Durum zmlemesi” standartının uygunluđuna ilişkin olarak olumlu grş belirttikleri ve bu standardı tm gstergeleri ile birlikte Trkiye iin program deđerlendirme standardı ve gstergeleri olarak kabul ettiklerini gstermektedir.

4.4.3. Alan Uzmanlarının Deđerlendirme Arařtırmasının Amacının Belirlenmesi Standartının Gstergelerine İliřkin Grřleri

Program deđerlendirmede “Dođruluk” standartlarını oluřturan nc standart “Deđerlendirme Arařtırmasının Amacının Belirlenmesi” standartıdır. Alan uzmanlarının “Deđerlendirme Arařtırmasının Amacının Belirlenmesi” standartının gstergelerinin uygunluđuna ilişkin grşlerini yansıtan sayısal veriler Tablo 33’te verilmiřtir.

Tablo 33.

*Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından
“Değerlendirme Araştırmasının Amacının Belirlenmesi” Standartının Göstergelerinin
Uygunluğuna İlişkin Görüşleri*

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Paydaş ve destekleyici katılımıyla değerlendirme amaçlarını belirleme	94	69,1	36	26,5	3	2,2	2	1,5	1	,7	4.62	.68
2. Paydaş ve destekleyici katılımıyla değerlendirme sorularını belirleme	89	65,4	36	26,5	7	5,1	3	2,2	1	,7	4.54	.76
3. Program değerlendirmenin kim için ve neden yapıldığını tanımlama	105	77,2	26	19,1	3	2,2	1	,7	1	,7	4.71	.62
4. Belirlenen amaçların hangi gereksinimleri karşılayacağını gerekçeleriyle ifade etme	51	37,5	49	36,0	20	14,7	9	6,6	7	5,1	3.94	1.12
											4.45	

Tablo 33’te görüldüğü gibi, program değerlendirme standartları altında yer alan “Doğruluk” standartlarından biri olan “Değerlendirme Araştırmasının Amacının Belirlenmesi” standartının dört göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Paydaş ve destekleyici katılımıyla değerlendirme amaçlarını belirleme”dir. Alan uzmanlarının %95,6’sı bu göstergeyi “uygun” bulurken, uzmanların %2,2’si “kısmen uygun” bulmuş ve %2,2’si de uygun bulmamıştır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.62’dir. Bu ortalama, “Paydaş ve destekleyici katılımıyla değerlendirme amaçlarını belirleme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “Değerlendirme Araştırmasının Amacının Belirlenmesi” standartıyla ilgili ikinci gösterge “Paydaş ve destekleyici katılımıyla değerlendirme sorularını belirleme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %91,9 “uygun”, %5,1 “kısmen uygun” ve %2,9de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.54’tür. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Paydaş ve destekleyici katılımıyla değerlendirme sorularını belirleme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“Değerlendirme Araştırmasının Amacının Belirlenmesi” standartının üçüncü göstergesi, “Program değerlendirmenin kim için ve neden yapıldığını tanımlama”dır. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %96,3 “uygun”, % 2,2 “kısmen uygun” %1,4 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.71’dir. Bu ortalama, bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“Değerlendirme Araştırmasının Amacının Belirlenmesi” standartının dördüncü göstergesi, “Belirlenen amaçların hangi gereksinimleri karşılayacağını gerekçeleriyle ifade etme”dir. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %73,5’inin göstergeyi “uygun” bulduğu, %14,7’lik bir kesiminin “kısmen uygun” bulduğu ve %11,7’sinin de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.94’tür. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“Doğruluk” standartının üçüncü alt standardı olan “Değerlendirme Araştırmasının Amacının Belirlenmesi” standartının göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın dört göstergesini de uygun bulmuşlardır. Standartın uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalamasına bakılmış ve bu ortalama 4.45 olarak hesaplanmıştır. Bu değer, alan uzmanlarının “Değerlendirme Araştırmasının Amacının Belirlenmesi” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standardı tüm göstergeleri ile birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.4.4. Alan Uzmanlarının Değerlendirme Araştırmasının Yönteminin Belirlenmesi Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “Doğruluk” standartlarını oluşturan dördüncü standart “Değerlendirme Araştırmasının Yönteminin Belirlenmesi” standartıdır. Alan uzmanlarının “Değerlendirme Araştırmasının Yönteminin Belirlenmesi” standartının

göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 34’te verilmiştir.

Tablo 34.

*Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından
“Değerlendirme Araştırmasının Yönteminin Belirlenmesi” Standartının
Göstergelerinin Uygunluğuna İlişkin Görüşleri*

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Program değerlendirme araştırmasında temel alınan program değerlendirme modelini ve sınırlılıklarını belirleme	98	72,1	22	16,2	9	6,6	6	4,4	1	,7	4.54	.86
2. Program değerlendirme araştırmasının desenini ve sınırlılıklarını belirleme	86	63,2	33	24,3	15	11,0	1	,7	1	,7	4.49	.78
3. Program değerlendirme araştırmasının evren ve örneklemini belirleme	100	73,5	20	14,7	14	10,3	1	,7	1	,7	4.60	.76
4. Program değerlendirme araştırmasının veri toplama ve çözümleme yöntem/tekniklerini belirleme	96	70,6	22	16,2	16	11,8	1	,7	1	,7	4.55	.79
											4.54	

Tablo 34’te görüldüğü gibi, program değerlendirmede “Doğruluk” standartlarından “Değerlendirme Araştırmasının Yönteminin Belirlenmesi” standartının dört göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge, “Program değerlendirme araştırmasında temel alınan program değerlendirme modelini ve sınırlılıklarını belirleme”dir. Alan uzmanlarının %88,3’ü bu göstergenin “uygun” olduğu, %6,6’sı bu göstergenin “kısmen uygun” olduğu, %5,1’i de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.54’tür. Bu ortalama, “Program değerlendirme araştırmasında temel alınan program değerlendirme modelini ve sınırlılıklarını belirleme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından “Değerlendirme Araştırmasının Yönteminin Belirlenmesi” standartıyla ilgili görüş bildirmeleri istenen ikinci gösterge, “Program değerlendirme araştırmasının desenini ve sınırlılıklarını belirleme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %87,5 “uygun”, %11 “kısmen

uygun” ve %1,4 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.49’dur. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Program değerlendirme araştırmasının desenini ve sınırlılıklarını belirleme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Değerlendirme Araştırmasının Yönteminin Belirlenmesi*” standartının üçüncü göstergesi, “Program değerlendirme araştırmasının evren ve örneklemini belirleme”dir. Araştırma anketini yanıtlayan alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin dağılımı; %88,2 “uygun”, % 10,3 “kısmen uygun” %1,4 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.60’tır. Bu ortalama, bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“*Değerlendirme Araştırmasının Yönteminin Belirlenmesi*” standartının dördüncü göstergesi, “Program değerlendirme araştırmasının veri toplama ve çözümleme yöntem/tekniklerini belirleme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarına göre, uzmanların %86,8’inin bu göstergeyi “uygun” bulduğu, %11,8’lik bir kesiminin “kısmen uygun” bulduğu ve %1,4’ünün de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.55’tir. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Doğruluk*” standartının dördüncü standardı olan “*Değerlendirme Araştırmasının Yönteminin Belirlenmesi*” standartının göstergelerinin Türkiye için uygunluğuna ilişkin olarak uzmanlar dört göstergeyi de uygun bulmuşlardır. Standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için hesaplanan standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması 4.54’tür. Bu değer, alan uzmanlarının “*Değerlendirme Araştırmasının Yönteminin Belirlenmesi*” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standardı tüm göstergeleri ile birlikte Türkiye için program değerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.4.5. Alan Uzmanlarının *Veri Kaynaklarının Tanımlanması* Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “*Doğruluk*” standartlarını oluşturan beşinci standart “*Veri Kaynaklarının Tanımlanması*” standarttır. Alan uzmanlarının “*Veri Kaynaklarının Tanımlanması*” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 35’te verilmiştir.

Tablo 35.

Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Veri Kaynaklarının Tanımlanması” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Program değerlendirme ile ilgili geçerli ve güvenilir bilgileri toplama	23	16,9	46	33,8	53	39,0	4	2,9	10	7,4	3.50	1.05
2. Program değerlendirme araştırmasında kullanılacak veri kaynaklarını (kurumlar, bireyler, tv-video, araştırmalar vb.) belirleme	121	89,0	12	8,8	1	,7	1	,7	1	,7	4.85	.53
3. Veri kaynaklarının seçiminde kullanılan yöntem ve ölçütleri belirleme	107	78,7	19	14,0	3	2,2	2	1,5	5	3,7	4.63	.90
4. Veri kaynaklarını ayrıntılı inceleme (evreni, örnekleme, araştırma deseni, veri toplama aracı vb. yönlerden)	117	86,0	4	2,9	6	4,4	7	5,1	2	1,5	4.67	.89
											4.41	

Tablo 35’te görüldüğü gibi, program değerlendirmede “*Doğruluk*” standartları kapsamında yer alan “*Veri Kaynaklarının Tanımlanması*” standartının beş göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Program değerlendirme ile ilgili geçerli ve güvenilir bilgileri toplama”dır. Alan uzmanlarının %50,7’si bu göstergeyi “uygun” bulurken, uzmanların %39’u bu göstergenin “kısmen uygun”, %10,3’ü de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 3.50’dir. Bu ortalama, “Program değerlendirme ile ilgili geçerli ve güvenilir bilgileri toplama” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “*Veri Kaynaklarının Tanımlanması*” standartıyla ilgili ikinci gösterge “Program değerlendirme araştırmasında kullanılacak veri kaynaklarını (kurumlar, bireyler, tv-video, araştırmalar vb.) belirleme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %97,8 “uygun”, %0,7 “kısmen uygun” ve %1,4 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının aritmetik ortalaması ise 4.85’tir. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Program değerlendirme araştırmasında kullanılacak veri kaynaklarını (kurumlar, bireyler, tv-video, araştırmalar vb.) belirleme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“*Veri Kaynaklarının Tanımlanması*” standartının üçüncü göstergesi, “Veri kaynaklarının seçiminde kullanılan yöntem ve ölçütleri belirleme”dir. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %92,7 “uygun”, % 2,2 “kısmen uygun” %5,2 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.63’tür. Bu ortalama, bu göstergenin Türkiye için “uygun” bir gösterge olduğunu göstermektedir.

“*Veri Kaynaklarının Tanımlanması*” standartının dördüncü göstergesi, “Veri kaynaklarını ayrıntılı inceleme (evreni, örnekleme, araştırma deseni, veri toplama aracı vb. yönlerden)”dir. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %88,9’unun göstergeyi “uygun” bulduğu, %4,4’lük bir kesiminin “kısmen uygun” bulduğu ve %6,6’sının ise uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.67’dir. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

“*Doğruluk*” standartının beşinci standardı olan “*Veri Kaynaklarının Tanımlanması*” standartının göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın dört göstergesini de “uygun” bulmuşlardır. “*Veri Kaynaklarının Tanımlanması*” standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalamasına bakılmış ve bu ortalama 4.41 olarak

hesaplanmıştır. Bu değer, alan uzmanlarının “*Veri Kaynaklarının Tanımlanması*” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı tüm göstergeleri ile birlikte Türkiye için program değerlendirme standartı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.4.6. Alan Uzmanlarının *Verilerin Geçerliliği ve Güvenirliği* Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “*Doğruluk*” standartlarını oluşturan altıncı standart “*Verilerin Geçerliliği ve Güvenirliği*” standartıdır. Alan uzmanlarının “*Verilerin Geçerliliği ve Güvenirliği*” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 36’de verilmiştir.

Tablo 36.

Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından “Verilerin Geçerliliği ve Güvenirliği” Standartının Göstergelerinin Uygunluğuna İlişkin Görüşleri

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Program değerlendirme araştırmasının amaçlarına uygun nitel ve nicel araştırma süreçleri belirleme	9	6,6	22	16,2	21	15,4	30	22,1	54	39,7	2.28	1.31
2. Program değerlendirme araştırmasının amaçları ile veri toplama yöntemlerinin uygunluğunu sağlama	105	77,2	12	8,8	15	11,0	2	1,5	2	1,5	4.59	.86
3. Veri toplama araçlarının geçerlik ve güvenilirlik çalışmaları için pilot uygulamalar yapma	108	79,4	13	9,6	3	2,2	7	5,1	5	3,7	4.56	1.02
4. Veri toplama araçları ile toplanacak verilerin kapsamını tanımlama	2	1,5	8	5,9	44	32,4	49	36,0	33	24,2	2.24	.94
5. Bulguların geçerlik ve güvenirliliğini etkileyebilecek katılımcı özelliklerini ifade etme (okuma yetersizliği, dil beceri düzeyi, fiziksel engeller vb.)	115	84,6	14	10,3	1	,7	4	2,9	2	1,5	4.74	.75
											3.68	

Tablo 36’da görüldüğü gibi, program değerlendirmede “*Doğruluk*” standartlarından “*Verilerin Geçerliliği ve Güvenirliği*” standartının beş göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Program değerlendirme araştırmasının amaçlarına uygun nitel ve nicel araştırma süreçleri belirleme”dir. Alan

uzmanlarının bu göstergeye ilişkin yanıtlar incelendiğinde 31 uzmanın (%22,8) göstergenin “uygun” olduğu yönünde görüş bildirdiği, buna karşılık 84 uzmanın (%61,8) ise uygun olmadığı yönünde görüş bildirdiği görülmektedir. Bu göstergesi “kısmen uygun” bulan uzmanların sayısı 21’dir (%15,4). Uzmanların bu göstergeye ilişkin yanıtlarının aritmetik ortalaması 2.28’dir. Bu değerlere göre, “Program değerlendirme araştırmasının amaçlarına uygun nitel ve nicel araştırma süreçleri belirleme” göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiştir.

Alan uzmanlarından görüş bildirmeleri istenen “*Verilerin Geçerliliği ve Güvenirliği*” standartıyla ilgili ikinci gösterge “Program değerlendirme araştırmasının amaçları ile veri toplama yöntemlerinin uygunluğunu sağlama”dır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %86 “uygun”, %11 “kısmen uygun” ve %3 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.59’dur. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Program değerlendirme araştırmasının amaçları ile veri toplama yöntemlerinin uygunluğunu sağlama” göstergesinin Türkiye için uygun olduğu görüşündedir.

“*Verilerin Geçerliliği ve Güvenirliği*” standartının üçüncü göstergesi, “Veri toplama araçlarının geçerlik ve güvenilirlik çalışmaları için pilot uygulamalar yapma”dır. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %89 “uygun”, % 2,2 “kısmen uygun” %8,8 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergesi uygun bulunduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.56’dır. Bu ortalama, bu göstergenin Türkiye için uygun bir gösterge olduğunu göstermektedir.

“*Verilerin Geçerliliği ve Güvenirliği*” standartının dördüncü göstergesi, “Veri toplama araçları ile toplanacak verilerin kapsamını tanımlama”dır. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %7,4’ü bu göstergesi “uygun” bulurken, %60,2’si uygun bulmamaktadır. Uzmanların % 32,4’ü bu göstergenin “kısmen uygun” olduğu görüşündedir. Uzmanların görüşlerinin aritmetik ortalaması 2.24’tür. Bu

değerler, bu göstergenin de uzmanlar tarafından Türkiye için uygun bulunmadığını göstermektedir.

“*Verilerin Geçerliliği ve Güvenirliđi*” standartının beşinci göstergesi, “Bulguların geçerlik ve güvenilirliğini etkileyebilecek katılımcı özelliklerini ifade etme (okuma yetersizliđi, dil beceri düzeyi, fiziksel engeller vb.)”dir. Bu göstergenin uygunluđuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %94,9 “uygun”, % 0,7 “kısmen uygun” %4,4 "uygun deđil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çođunluđunun bu göstergeyi uygun bulduđunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.74’tür. Bu ortalama, bu göstergenin Türkiye için uygun bir gösterge olduđunu göstermektedir.

“*Dođruluk*” standartının altıncı standardı olan “*Verilerin Geçerliliđi ve Güvenirliđi*” standartının göstergelerinin Türkiye için uygunluđuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın üç göstergesini uygun bulmuşlar, iki göstergesini ise uygun bulmamışlardır. Standartın uzmanlar tarafından uygun bulunup bulunmadığının saptanmasında ölçüt olarak alınan bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalaması 4.68’dir. Bu deđer, alan uzmanlarının “*Verilerin Geçerliliđi ve Güvenirliđi*” standartının uygunluđuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı birinci ve dördüncü göstergeleri dışında diđer göstergeleriyle birlikte Türkiye için program deđerlendirme standardı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.4.7. Alan Uzmanlarının *Verilerin Çözümlemesi* Standartının Göstergelerine İlişkin Görüşleri

“*Dođruluk*” standartlarının yedinci standardı “*Verilerin Çözümlemesi*” standardıdır. Alan uzmanlarının “*Verilerin Çözümlemesi*” standartının göstergelerinin uygunluđuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 37’de verilmiştir.

Tablo 37.

*Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından
“Verilerin Çözümlemesi” Standartının Göstergelerinin Uygunluğuna İlişkin
Görüşleri*

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Nicel verileri çözümlenmek için uygun istatistiksel teknikleri belirleme	107	78,7	16	11,8	11	8,1	1	,7	1	,7	4.67	.72
2. Nitel verilerin çözümlenmesi için uygun nitel veri çözümleme tekniklerini belirleme	96	70,6	25	18,4	11	8,1	2	1,5	2	1,5	4.55	.82
3. Veri girişlerinin uzmanlarca yapılmasını sağlama	7	5,1	7	5,1	21	15,4	35	25,7	66	48,5	1.93	1.15
4. Verileri doğru kodlama ve puanlama	90	66,2	32	23,5	8	5,9	2	1,5	4	2,9	4.49	.90
5. Tablo, grafikler vb. görsellerden yararlanarak çözümlenmelerin sunumunu görselleştirme	100	73,5	19	14,0	14	10,3	1	,7	2	1,5	4.57	.82
6. Çözümlenmeler sonucu ortaya çıkan veri tablolarını gözden geçirme ve eksiklikleri giderme	24	17,6	4	2,9	18	13,2	31	22,8	59	43,4	2.29	1.49
7. Olanaklar ölçüsünde verilerin doğrulanması ve veri tablolarının gözden geçirilmesi için farklı uzman görüşlerine başvurma	16	11,8	24	17,6	27	19,9	18	13,2	51	37,5	2.53	1.43
											3.57	

Tablo 37’de görüldüğü gibi, program değerlendirmede “Doğruluk” standartlarından biri olan “Verilerin Çözümlemesi” standartının yedi göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Nicel verileri çözümlenmek için uygun istatistiksel teknikleri belirleme”dir. Alan uzmanlarının %90,5’i bu göstergelyi “uygun” bulurken, uzmanların %8,1’i bu göstergenin “kısmen uygun”, %1,4’ü de uygun olmadığı görüşündedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.67’dir. Bu ortalama, “Nicel verileri çözümlenmek için uygun istatistiksel teknikleri belirleme” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “Verilerin Çözümlemesi” standartıyla ilgili ikinci gösterge “Nitel verilerin çözümlenmesi için uygun nitel veri çözümleme tekniklerini belirleme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin

yanıtlarının dağılımı; %89 “uygun”, %8,1 “kısmen uygun” ve %3 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.55’tir. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Nitel verilerin çözümlenmesi için uygun nitel veri çözümleme tekniklerini belirleme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“Verilerin Çözümlenmesi” standartının üçüncü göstergesi, “Veri girişlerinin uzmanlarca yapılmasını sağlama”dır. Bu göstergeye ilişkin yanıtlar incelendiğinde 14 uzmanın (%10,2) göstergenin “uygun” olduğu yönünde görüş bildirdiği, buna karşılık 101 uzmanın (%74,2) ise uygun olmadığı yönünde görüş bildirdiği görülmektedir. Bu göstergesi “kısmen uygun” bulan uzmanların sayısı 21’dir (%15,4). Uzmanların bu göstergeye ilişkin yanıtlarının aritmetik ortalaması 1.93’tür. Bu değerlere göre, “Veri girişlerinin uzmanlarca yapılmasını sağlama” göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiştir.

“Verilerin Çözümlenmesi” standartının dördüncü göstergesi, “Verileri doğru kodlama ve puanlama”dır. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %89,7’sinin göstergesi “uygun” bulunduğu, %5,9’luk bir kesiminin “kısmen uygun” bulunduğu ve %4,4’ünün de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.49’dur. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için “uygun” görüldüğü söylenebilir.

Alan uzmanlarından görüş bildirmeleri istenen “Verilerin Çözümlenmesi” standartıyla ilgili beşinci gösterge “Tablo, grafikler vb. görsellerden yararlanarak çözümlemelerin sunumunu görselleştirme”dir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin yanıtlarının dağılımı; %87,5 “uygun”, %10,3 “kısmen uygun” ve %2,2 de “uygun değil” şeklindedir. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin görüşlerinin aritmetik ortalaması ise 4.57’dir. Bu ortalamaya göre, alan uzmanlarının büyük çoğunluğu, “Tablo, grafikler vb. görsellerden yararlanarak çözümlemelerin sunumunu görselleştirme” göstergesinin Türkiye için “uygun” olduğu görüşündedir.

“Verilerin Çözümlenmesi” standartının altıncı göstergesi, “Çözümlemeler sonucu ortaya çıkan veri tablolarını gözden geçirme ve eksiklikleri giderme”dir. Bu göstergeye

ilişkin yanıtlar incelendiğinde 28 uzmanın (%20,5) göstergenin “uygun” olduğu yönünde görüş bildirdiği, buna karşılık 90 uzmanın (%66,2) ise uygun olmadığı yönünde görüş bildirdiği görülmektedir. Bu göstergesi kısmen uygun bulan uzmanların sayısı 18’dir (%13,2). Uzmanların bu göstergesi ilişkin yanıtlarının aritmetik ortalaması ise 2.29’dur. Bu değerlere göre, “Çözümlemeler sonucu ortaya çıkan veri tablolarını gözden geçirme ve eksiklikleri giderme” göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiştir.

“*Verilerin Çözümlemesi*” standartının son göstergesi, “Olanaklar ölçüsünde verilerin doğrulanması ve veri tablolarının gözden geçirilmesi için farklı uzman görüşlerine başvurma”dır. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %29,4’ü bu göstergesi “uygun” bulurken, %50,7’si uygun bulmamaktadır. Uzmanların %19,9’u bu göstergenin “kısmen uygun” olduğu görüşündedir. Uzmanların görüşlerinin aritmetik ortalaması ise 2.53’tür. Bu değerler, bu göstergenin de uzmanlar tarafından Türkiye için uygun bulunmadığını göstermektedir.

“*Doğruluk*” standartının yedinci standartı olan “*Verilerin Çözümlemesi*” standartının göstergelerinin Türkiye için uygunluğuna ilişkin olarak uzmanlar dört göstergesi uygun bulmuşlar, üç göstergesi ise uygun bulmamışlardır. Standartın uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalamasına bakılmış ve bu ortalama 3.57 olarak hesaplanmıştır. Bu değer, alan uzmanlarının “*Verilerin Çözümlemesi*” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı üçüncü, altıncı ve yedinci göstergeler dışında diğer göstergeleriyle birlikte Türkiye için program değerlendirme standartı ve göstergeleri olarak kabul ettiklerini göstermektedir.

4.4.8. Alan Uzmanlarının *Sonuçların Gerekçelenirilmesi* Standartının Göstergelerine İlişkin Görüşleri

Program değerlendirmede “*Doğruluk*” standartlarının sonuncusu “*Sonuçların Gerekçelenirilmesi*” standartıdır. Alan uzmanlarının “*Sonuçların Gerekçelenirilmesi*” standartının göstergelerinin uygunluğuna ilişkin görüşlerini yansıtan sayısal veriler Tablo 38’de verilmiştir.

Tablo 38.

*Alan Uzmanlarının Program Değerlendirmede “Doğruluk” Standartlarından
“Sonuçların Gereçekleştirilmesi” Standartının Göstergelerinin Uygunluğuna İlişkin
Görüşleri*

(n=136)

Göstergeler	Tamamen Uygun		Oldukça Uygun		Kısmen Uygun		Çok Az Uygun		Hiç Uygun Değil		\bar{X}	SS
	S	%	S	%	S	%	S	%	S	%		
1. Program değerlendirme araştırmasının sorularının yanıtlarına ilişkin bulgular sunma	103	75,7	24	17,6	2	1,5	6	4,4	1	,7	4,63	.79
2. Ulaşılan sonuçlara ilişkin kanıtlar ortaya koyma	17	12,5	15	11,0	22	16,2	44	32,4	38	27,9	2,48	1.34
3. Sonuçlar doğrultusunda önerilerde bulunma	112	82,4	18	13,2	4	2,9	1	,7	1	,7	4,76	.61
4. Bulguların alternatif yorumlamalarına yer verme	36	26,5	15	11,0	29	21,3	55	40,4	1	,7	3,22	1.25
5. Zaman, ortam, birey ve değerlendirme amaçları ve süreçleri yönünden sonuçların sınırlılıklarını ifade etme	106	77,9	19	14,0	4	2,9	4	2,9	3	2,2	4,63	.86
											3,94	

Tablo 38’de görüldüğü gibi, program değerlendirmede “Doğruluk” standartlarından biri olan “Sonuçların Gereçekleştirilmesi” standartının beş göstergesi vardır. Alan uzmanlarından görüş belirtmeleri istenen ilk gösterge “Program değerlendirme araştırmasının sorularının yanıtlarına ilişkin bulgular sunma”dır. Alan uzmanlarının %93,3’ü bu göstergeyi “uygun” bulurken, uzmanların %1,5’i bu göstergenin “kısmen uygun” bulmuş ve %5,1’i de uygun bulmamıştır. Alan uzmanlarının bu göstergenin uygunluğuna ilişkin verdikleri yanıtların aritmetik ortalaması 4.63’tür. Bu ortalama, “Program değerlendirme araştırmasını destekleyenleri tanımlama” göstergesinin uzmanlar tarafından “uygun” bulunduğunu göstermektedir.

Alan uzmanlarından görüş bildirmeleri istenen “Sonuçların Gereçekleştirilmesi” standartıyla ilgili ikinci gösterge “Ulaşılan sonuçlara ilişkin kanıtlar ortaya koyma”dır. Bu göstergeye ilişkin yanıtlar, 32 uzmanın (%23,5) göstergenin “uygun” olduğu yönünde görüş bildirdiğini, buna karşılık 82 uzmanın (%60,3) ise uygun olmadığı yönünde görüş bildirdiğini göstermektedir. Bu göstergeyi “kısmen uygun” bulan uzmanların sayısı 22’dir (%16,2). Uzmanların bu göstergeye ilişkin yanıtlarının

aritmetik ortalaması 2.48'dir. Bu değerlere göre, "Ulaşılan sonuçlara ilişkin kanıtlar ortaya koyma" göstergesi uzmanlar tarafından Türkiye için uygun bir gösterge olarak kabul edilmemiştir.

"*Sonuçların Gerekçelendirilmesi*" standartının üçüncü göstergesi, "Sonuçlar doğrultusunda önerilerde bulunma"dır. Bu göstergenin uygunluğuna ilişkin araştırma anketini yanıtlayan uzmanların görüşlerinin dağılımı; %95,6 "uygun", % 2,9 "kısmen uygun" %1,4 "uygun değil" şeklindedir. Alan uzmanlarının yanıtlarının bu dağılımı, uzmanların büyük çoğunluğunun bu göstergeyi uygun bulduğunu göstermektedir. Nitekim bu göstergeye ilişkin uzmanların görüşlerini yansıtan yanıtların aritmetik ortalaması da 4.76'dır. Bu ortalama, bu göstergenin Türkiye için "uygun" bir gösterge olduğunu göstermektedir.

"*Sonuçların Gerekçelendirilmesi*" standartının dördüncü göstergesi, "Bulguların alternatif yorumlamalarına yer verme"dır. Bu göstergenin uygunluğuna ilişkin uzman yanıtlarına göre, uzmanların %37,5'si bu göstergeyi "uygun" bulurken, %41,1'i uygun bulmamaktadır. Uzmanların % 21,3'ü bu göstergenin "kısmen uygun" olduğu görüşündedir. Uzmanların görüşlerinin aritmetik ortalaması 3.22'dir. Bu değerler, bu göstergenin de uzmanlar tarafından "kısmen uygun" bulunduğunu göstermektedir.

"*Sonuçların Gerekçelendirilmesi*" standartının beşinci göstergesi, "Zaman, ortam, birey ve değerlendirme amaçları ve süreçleri yönünden sonuçların sınırlılıklarını ifade etme"dır. Alan uzmanlarının yanıtları incelendiğinde, uzmanların %91,9'unun göstergeyi "uygun" bulduğu, %2,9'luk bir kesiminin "kısmen uygun" bulduğu ve %5,1'inin de uygun bulmadığı görülmektedir. Bu göstergenin uygunluğuna ilişkin uzman görüşlerinin aritmetik ortalaması 4.63'tür. Bu ortalamaya göre, bu göstergenin uzmanlar tarafından Türkiye için "uygun" görüldüğü söylenebilir.

"*Doğruluk*" standartının son alt standardı olan "*Sonuçların Gerekçelendirilmesi*" standartının göstergelerinin Türkiye için uygunluğuna ilişkin görüşlerine başvuru alan uzmanlar, bu standartın üç göstergesini uygun, bir göstergesini "kısmen uygun" bulmuşlar ve bir göstergesini de uygun bulmamışlardır. Standartının uzmanlar tarafından uygun bulunup bulunmadığının saptanması için de uzmanların bu standartın

göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalamasına bakılmış ve bu ortalama 3.72 olarak hesaplanmıştır. Bu değer, alan uzmanlarının “*Sonuçların Gereçlendirilmesi*” standartının uygunluğuna ilişkin olarak olumlu görüş belirttikleri ve bu standartı ikinci gösterge dışında diğer göstergeleriyle birlikte Türkiye için program değerlendirme standartı ve göstergeleri olarak kabul ettiklerini göstermektedir.

Program değerlendirmede “*Doğruluk*” standart alanını oluşturan sekiz standartın altında yer alan standart göstergelerinin uygunluğuna ilişkin alan uzmanlarının görüşleriyle ilgili Tablo 31,32, 33, 34, 35, 36, 37 ve 38’da verilen altı standartın göstergelerine ilişkin sayısal veriler bir bütün olarak değerlendirildiğinde, Tablo 39’da da görüldüğü gibi, uzmanlar “*Değerlendirilecek Programın Tanımlanması*”, “*Durum Çözümlemesi*”, “*Değerlendirme Araştırmasının Amacının Belirlenmesi*”, “*Değerlendirme Araştırmasının Yönteminin Belirlenmesi*”, ve “*Veri Kaynaklarının Tanımlanması*” standartlarının göstergelerinin tamamını, “*Verilerin Geçerliliği ve Güvenirliği*” standartının beş göstergesinden üçünü, “*Verilerin Çözümlemesi*” standartının yedi göstergesinden dördünü ve “*Sonuçların Gereçlendirilmesi*” standartının beş göstergesinden üçünü uygun, birini kısmen uygun bulmuşlardır. Buna karşılık uzmanlar “*Verilerin Geçerliliği ve Güvenirliği*” standartının iki göstergesini, “*Verilerin Çözümlemesi*” standartının üç göstergesini ve “*Sonuçların Gereçlendirilmesi*” standartının bir göstergesini uygun bulmamışlardır. Öte yandan, uzmanların standart göstergelerine verdikleri yanıtların aritmetik ortalamalarının ortalamaları da standartlara göre 3.57 (en düşük) ile 4.54 (en yüksek) arasında değişmektedir. Bu değerler, standartların uzmanlarca uygun bulunduğunu göstermektedir. Sonuç olarak alan uzmanları, sekiz “*Doğruluk*” standartını ve bu standartlara ilişkin 30 göstergeyi Türkiye için uygun program değerlendirme standartları olarak kabul etmişlerdir.

Tablo 39.

Program Değerlendirmede “Doğruluk” Standartlarının Aritmetik Ortalamaları

Standart	GS	\bar{X}	Uygun Bulunmayan Göstergeler	\bar{X}
1.Değerlendirilecek Programın Tanımlanması	4	4.38	Uygun bulunmayan gösterge yok	
2.Durum Çözümlemesi	3	4.32		
3.Değerlendirme Araştırmasının Amacının Belirlenmesi	4	4.45		
4.Değerlendirme Araştırmasının Yönteminin Belirlenmesi	4	4.54		
5.Verit Kaynaklarının Tanımlanması	4	4.41		
6.Verilerin Geçerliliği ve Güvenirliđi	5	3.68	6.1Program değerlendirme araştırmasının amaçlarına uygun nitel ve nicel araştırma süreçleri belirleme	2.28
			6.4 Veri toplama araçları ile toplanacak verilerin kapsamını tanımlama	2.24
			7.3 Veri girişlerinin uzmanlarca yapılmasını sağlama	1.93
			7.6 Çözümlemeler sonucu ortaya çıkan veri tablolarını gözden geçirme ve eksiklikleri giderme	2.29
			7.7 Olanaklar ölçüsünde verilerin doğrulanması ve veri tablolarının gözden geçirilmesi için farklı uzman görüşlerine başvurma	2.53
8.Sonuçların Gerekçelenirilmesi Standartları	5	3.94	8.2 Ulaşılan sonuçlara ilişkin kanıtlar ortaya koyma	2.48

GS: Gösterge Sayısı

4.5. Program Değerlendirme Standartlarının Uygunluk Düzeyinin Standart Alanlarına Göre İncelenmesi

Araştırmada yanıtı aranan beşinci soru, program değerlendirme standartlarının uygunluk düzeyinin standart alanlarına göre farklılık gösterip göstermediğidir. Standart alanlarının uygunluk düzeyini belirlemek için her standart alanının altında yer alan standartların göstergelerinin tümünün aritmetik ortalamalarının ortalaması alınmıştır. Standart alanının uygunluk düzeyinin yalnızca gösterge ortalamalarına göre belirlenmesinin temel nedeni, her standart altında eşit gösterge olmaması ve uzmanlardan göstergelere ilişkin olarak görüş alınmasıdır. Standart alanlarının uygunluk düzeyine ilişkin ortalamalar Tablo 40’ta verilmiştir.

Tablo 40.

Program Değerlendirme Standartlarının Standart Alanlarına Göre Uygunluk Düzeyi

Standart Alanı	Standart Sayısı	Gösterge Sayısı	Çıkan Gösterge Sayısı	\bar{X}	ss
Yararlık Standartları	5	24	3	4.10	.93
Yürütülebilirlik Standartları	4	17	4	3.96	.71
Uygunluk Standartları	6	33	3	4.21	.46
Doğruluk Standartları	8	36	6	4.09	.39
Toplam	23	110	16		

Tablo 40'ta görüldüğü gibi, alan uzmanlarının görüşüne sunulan 23 program değerlendirme standartlarının tümü uzmanlar tarafından Türkiye için uygun bulunmuştur. Bu standartlara ilişkin olarak alan uzmanlarının görüşüne sunulan 110 göstergeden 94'ü Türkiye için uygun, 16'sı ise uygun bulunmamıştır. Standart alanlarının uygunluk düzeyine ilişkin ortalamalar incelendiğinde, alan uzmanlarının görüşlerine göre, program değerlendirme standartlarının uygunluk düzeyi en yüksek olan standart alanı 4.21 aritmetik ortalama ile “Uygunluk” standartları olurken; bunu 4.10 aritmetik ortalama ile “Yararlık” standartları, 4.09 aritmetik ortalama ile “Doğruluk” standartları ve 3.96 aritmetik ortalama ile “Yürütülebilirlik” standartları izlemektedir. Görüldüğü gibi standart alanları arasında ortalamalara göre bir farklılık bulunmaktadır. Bu farklılığın istatistiksel olarak anlamlı olup olmadığını sınamak amacıyla tek yönlü varyans çözümlemesi uygulanmıştır (ANOVA). Elde edilen çözümleme sonuçları Tablo 41'de gösterilmiştir.

Tablo 41.

Alan Uzmanlarının Program Değerlendirme Standartlarının Uygunluk Düzeyine İlişkin Görüş Puanlarına Uygulanan Tek Yönlü Varyans Çözümlemesi Sonuçları

Varyans kaynağı	KT	SD	KO	F	P
Gruplararası	4,389	3	1.463	3.3933	.018*
Gruplar içi	232,871	540	.431		
Toplam	237,261	543			

$\alpha = 0,05$

Tablo 41'de görüldüğü gibi, program değerlendirme standartlarının uygunluk düzeyinin standart alanlarına göre farklılık gösterip göstermediğine ilişkin tek yönlü varyans

çözümlemesi (ANOVA) sonucu elde edilen *P* değeri, $\alpha=0.05$ anlamlılık düzeyinden küçük olduğu için gruplar arasındaki fark anlamlı bulunmuştur. Bu farklılığın hangi alan ya da alanlararası farklılıktan kaynaklandığını saptamak üzere Tukey HSD testi uygulanmıştır. Tukey HSD testi sonuçları Tablo 42’de verilmiştir.

Tablo 42.

Alan Uzmanlarının Program Değerlendirme Standartlarının Uygunluk Düzeyine İlişkin Görüş Puanlarına Uygulanan Tukey HSD Testi Sonuçları

Standart Alanları	Yararlık Standartları	Yürütülebilirlik Standartları	Uygunluk Standartları	Doğruluk Standartları
Yararlık Standartları	---	.265	.525	.999
Yürütülebilirlik Standartları	.265	---	.008*	.336
Uygunluk Standartları	.525	.008*	---	.438
Doğruluk Standartları	.999	.336	.438	---

Tablo 42 incelendiğinde, program değerlendirme standart alanlarının uygunluk düzeyleri arasındaki farkın “Yürütülebilirlik” standartları ile “Uygunluk” standartları arasındaki farklılıktan kaynaklandığı görülmektedir.

Sonuç olarak, alan uzmanlarının program değerlendirme standart alanlarının uygunluk düzeyine ilişkin görüşlerinin arasındaki farkın istatistiksel olarak anlamlı olduğu ve bu farkın “Yürütülebilirlik” standartları ile “Uygunluk” standartları arasındaki farktan kaynaklandığı görülmüştür. Bu farkın nedeni uzmanların program değerlendirmede etik ilkeleri ön plana çıkaran “Uygunluk” standartlarının bir programın değerlendirilebilirliğini veya yapılabilirliğini ön plana çıkaran “Yürütülebilirlik” standartlarından daha uygun veya önemli görmeleri olarak açıklanabilir.

4.6. Alan Uzmanlarının Program Değerlendirme Standartları ve Göstergelerine İlişkin Önerileri

Araştırma kapsamında, eğitim fakültelerinde görev yapan alan uzmanlarının ankette yer alan göstergelerine yönelik olarak varsa önerileri sorulmuştur. Her standart alanının

sonunda boş bırakılan bölmeye standartlar ve göstergeleri ile ilgili önerilerde bulunan alan uzmanlarının bu görüşleri standart alanlarına göre ayrı tablolarda verilmiştir.

4.6.1. Alan Uzmanlarının “Yararlık” Standartlarına İlişkin Önerileri

Araştırma kapsamında uzmanlardan değerlendirdikleri “Yararlık” standartları ve göstergelerine ilişkin olarak varsa önerilerini belirtmeleri istenmiştir. “Yararlık” standartlarına ilişkin olarak toplam 21 uzmandan öneri gelmiştir. Uzmanlardan gelen bu önerilerden aynı göstergeye gelen öneriler bir araya getirilmiştir. Madde başına öneride bulunan uzman sayısının araştırmaya katılan uzman sayısına göre oldukça düşük olmasından dolayı standartlarda ve göstergelerde bir değişikliğe gidilmemiştir. “Yararlık” standartlarına ve göstergelerine ilişkin olarak uzmanlardan gelen öneriler Tablo 43’te verilmiştir.

Tablo 43.

Alan Uzmanlarının “Yararlık” Standartlarının Göstergelerine İlişkin Yorumları ve Önerileri

Göstergeler	Gelen Yorumlar ve Öneriler
1.2 Program değerlendirme araştırmasını planlayacak, uygulayacak ve değerlendirecek değerlendirme uzman/Alan uzmanlarının tanımlama	<ul style="list-style-type: none"> Göstergede birkaç yargı aynı anda sorgulanmaktadır. Bu yargıların ayrılması gereklidir. Gösterge; “değerlendirme uzmanının tanımlanması” olarak değiştirilebilir.
1.4 Program değerlendirme araştırmasını sonuçlarından doğrudan ya da dolaylı etkilenecek grup ya da bireyleri belirtme	<ul style="list-style-type: none"> Birey ya da grupların belirlenmesi bir baskıya yol açabilir.
2.1 Program değerlendirme konusunda akademik bir süreçten geçmiş olma	<ul style="list-style-type: none"> “süreçten” sözcüğü yerine “eğitimden” sözcüğü kullanılmalıdır.
2.2 Program değerlendirme araştırmasını planlanma, uygulama ve değerlendirme süreçlerinde gerekli bilgi, beceri ve deneyimlere sahip olma	<ul style="list-style-type: none"> Gösterge birkaç yargı bildirmektedir. Bu nedenle yargılar ayrılmalıdır. Gösterge “program değerlendirme sürecine ilişkin bilgi, beceri ve deneyime sahip olma” olarak değiştirilebilir.
2.3 Bireysel sınırlılıkların farkına varma, bu sınırlılıkları aşan durumlarda uzman yardımı alma	<ul style="list-style-type: none"> Değerlendiricinin bireysel sınırlılıklarının farkında olması pek olanaklı değildir. Bireysel sınırlılıkların çerçevesi net değildir.
2.4 Farklı gruplar ile çalışmak için uygun iletişim stratejilerini ve becerilerini kullanma	<ul style="list-style-type: none"> “iletişim stratejileri ve becerileri” ifadesi iki farklı yargı bildirmektedir. Ayrılmalıdır.
3.1 Program değerlendirme araştırmasının planlanmasında araştırma sorularının yanıtlanması için gerekli/yararlı bilgileri belirleme/seçme	<ul style="list-style-type: none"> Gösterge daha kısa ifadelendirilebilir. Örneğin “değerlendirme sorularını yanıtlayıcı yararlı bilgileri seçme” gibi.
4.1 Raporda program değerlendirme araştırmasının amacını, içeriğini, araştırma sorularını, süreçte kullanılan yöntemleri ve değerlendirme bulgularını açık ve kapsamlı olarak betimleme	<ul style="list-style-type: none"> Virgül ile ifade edilen yargıları ayrı göstergeler halinde yazılmalıdır.
4.4 Raporda ilgililerin anlayabileceği nitelikte bir dil kullanma/ teknik dili az kullanma	<ul style="list-style-type: none"> Herkesin anlamasını sağlamak uğruna teknik dilden çok fazla taviz verilmemelidir. Teknik kelimeleri açıklayan bir sözlük verildiği sürece teknik kelimeleri azaltmaya değil de arı bir Türkçe kullanımına gidilmesi yeterli olacaktır.
5.2 Program değerlendirme sonuçlarının paydaşlar yönünden olası yararlarını/etkilerini belirleme	<ul style="list-style-type: none"> Yarar ve etki kavramları eş kavramlar değildir. Yarar sözcüğünün kullanılması daha uygundur.
5.3 Program değerlendirme sonuçlarının paydaşlarda yeni bakış açısı/anlayış geliştirmesi	<ul style="list-style-type: none"> Bakış açısı ve anlayış aynı kavramlar değildir. Bu göstergenin uygulanması olanaklı değildir. Çünkü anlayışın sorgulanması olası değildir.

Tablo 43'e göre, alan uzmanlarının “*Yararlık*” standartlarına ve göstergelerine yönelik önerileri incelendiğinde, alan uzmanlarının genellikle gösterge ifadelerinin “ve” bağlacıyla belirtilerin yargılarının iki/ya da daha fazla göstergeyle belirtilmesi gerektiğini vurguladıkları görülmektedir (gösterge 1.2, 2.2, 2.4, 4.1, 5.2). Bunun yanı sıra uzmanlar gösterge ifadelerinde yer alan bazı sözcüklerin anlam ve kullanım yönünden hatalı olduğunu vurgulamakta ve bu sözcüklerin yerine öneriler getirmektedirler. Örneğin, gösterge 2.1'de “süreçten” sözcüğü yerine “eğitimden” sözcüğünün kullanılması önerisi getirilmiştir. Bu önerilerle birlikte uzmanlar bazı göstergelerin ifadelendirilmesinin de değiştirilmesini önermektedirler. Bu öneriler şunlardır: Gösterge 1.2 için “Değerlendirme uzmanının tanımlanması” ve gösterge 3.1 için “Değerlendirme sorularını yanıtlayıcı yararlı bilgileri seçme” önerileri getirilmiştir. Yine uzmanlar gösterge 2.3'ün yapı olarak uygun olmadığını belirtmişlerdir. Nitekim bu göstergeye ilişkin uzman görüşleri incelendiğinde bu göstergenin uzmanlar tarafından Türkiye için uygun bulunmadığı da ortaya çıkmıştır.

4.6.2. Alan Uzmanlarının “*Yürütülebilirlik*” Standartlarına İlişkin Önerileri

Araştırma kapsamında uzmanlardan değerlendirdikleri “*Yürütülebilirlik*” standartları ve göstergelerine ilişkin olarak varsa önerilerini belirtmeleri istenmiştir. “*Yürütülebilirlik*” standartlarına ilişkin olarak 13 uzmandan öneri gelmiştir. Yine bir önceki standart alanında olduğu gibi, sınırlı sayıda uzmandan yalnızca birkaç madde ile ilgili öneri geldiğinden gelen öneriler doğrultusunda standartlarda ve göstergelerde bir değişikliğe gidilmemiştir. “*Yürütülebilirlik*” standartlarına ve göstergelerine ilişkin olarak uzmanlardan gelen öneriler Tablo 44'te verilmiştir.

Tablo 44.

*Alan Uzmanlarının “Yürütülebilirlik” Standartlarının Göstergelerine İlişkin
Yorumları ve Önerileri*

Göstergeler	Gelen Yorumlar ve Öneriler
1.1 Program değerlendirme araştırmasında seçilen yöntemin maliyet yönünden uygunluğunu inceleme	<ul style="list-style-type: none"> • İfade “değerlendirme amacına uygun yöntem belirleme” olarak değiştirilebilir
2.1 Program değerlendirme araştırması sürecine doğrudan ya da dolaylı katılacak kişi, kurum ve kuruluşların desteğini alma	<ul style="list-style-type: none"> • “katılacak kişi, kurum ve kuruluşların desteğini alma” ifadesi sonraki standartlarla bir karmaşa yaratıyor bu ifade yerine “katılacakların desteğini alma” ifadesi kullanılabilir.
2.2 Program değerlendirme araştırmasının sonuçlarını kullanacak kişi, kurum ve kuruluşların desteğini alma	<ul style="list-style-type: none"> • “kullanacakların desteğini alma” ifadesi kullanılabilir.
2.3 Program değerlendirme sürecinden doğrudan ya da dolaylı etkilenebilecek kişi, kurum ve kuruluşların desteğini alma	<ul style="list-style-type: none"> • “etkilenebileceklerin desteğini alma” ifadesi kullanılabilir. • Süreçten etkilenebilecek bireylerin belirtilmesi bir baskıya neden olabilir
4.1 Değerlendirilmesi amaçlanan programın değerlendirilebilirliğini ortaya koyma	<ul style="list-style-type: none"> • Bu gösterge “İlgili programın değerlendirilebilirliğini standartlarla ortaya koyma” olarak değiştirilebilir.
4.4 Değerlendirme araştırmasının yürütme stratejilerini belirleme	<ul style="list-style-type: none"> • “yürütme stratejileri” kavramı çok anlamlı değil, ne olduğu ayrıntılandırılmalıdır.

Tablo 44 incelendiğinde, uzmanların altı “Yürütülebilirlik” göstergesine ilişkin öneri ve yorumda buldukları görülmektedir. Önerilerin geneline bakıldığında iki göstergeye ilişkin olarak göstergelerin yeniden ifadelendirilmesi, diğer göstergelerde ise göstergelerin kısaltılması ve bazı ifadelerin değiştirilmesi önerilerinin getirildiği görülmektedir. Gösterge 1.1 için getirilen öneri “değerlendirme amacına uygun yöntem belirleme” önerisidir. Bu öneride gerçek göstergede vurgulanan maliyet yönünün kaldırılması önerilmektedir. İkinci gösterge önerisi ise, gösterge 4.1 için yapılmıştır. Önerilen gösterge “İlgili programın değerlendirilebilirliğini standartlarla ortaya koyma”dır. Bu öneriyle öneriyi yapan uzman programın değerlendirilebilirliği için standartların kullanılması gereğini ön plana çıkarmak istemektedir. Gösterge 2.1, 2.2, ve 2.3’te ise gösterge ifadesinin kısaltılması önerilmektedir. Bu göstergelerden gösterge 2.3’e ilişkin olarak öneride bulunan uzmanlar süreçten etkilenebilecek bireylerin belirtilmesinin bir baskıya neden olabileceğini belirterek standartın uygun olmadığını ifade etmişlerdir. Nitekim bu göstergeye ilişkin uzman görüşleri incelendiğinde uzmanların göstergeyi Türkiye için uygun bulmadıkları görülmektedir. Öneride bulunan uzmanlar gösterge 4.4 için ise “yürütme stratejileri” kavramının açık olmadığı ve yeniden tanımlanması gerektiği vurgulanmaktadır.

4.6.3. Alan Uzmanlarının “Uygunluk” Standartlarına İlişkin Önerileri

Araştırma kapsamında uzmanlardan değerlendirdikleri “Uygunluk” standartları ve göstergelerine ilişkin olarak varsa önerilerini belirtmeleri istenmiştir. Bu standartlara ilişkin olarak 16 uzman yedi göstergeye yönelik öneride bulunmuştur. Madde başına öneride bulunan uzman sayısının araştırmaya katılan uzman sayısına oranla oldukça düşük olmasından (madde başına 2-4 uzman önerisi) dolayı gelen öneriler doğrultusunda standartlarda ve göstergelerde bir değişikliğe gidilmemiştir. Uzmanlardan gelen bu önerilerden aynı göstergeye gelen öneriler bir araya getirilmiş ve Tablo 45’te verilmiştir.

Tablo 45.

Alan Uzmanlarının “Uygunluk” Standartlarının Göstergelerine İlişkin Yorumları ve Önerileri

Göstergeler	Gelen Yorumlar ve Öneriler
1.6 Süreç içinde hizmetlerin ve harcamaların ayrıntılı kayıtlarını tutma	<ul style="list-style-type: none">Yapı bir önceki göstergedeki yapıyla tümüyle bir uyuma içinde. Kaldırılmalıdır.
3.3 Paydaşların beklenti ve isteklerinden etkilenmemesi	<ul style="list-style-type: none">İfade “Program değerlendirme raporunun değerlendiricinin nesnellliğini yansıtması” ifadesiyle örtüşmektedir.Ayrıca bir sonraki gösterge ile binişik bir anlama sahip. Beklenti ve istek sözcükleri ile talep sözcüğü aynı anlamda diye düşünüyorum.
3.4 Uygulamanın paydaş taleplerinden bağımsız yürütülmesi	<ul style="list-style-type: none">Gösterge “Uygulamanın paydaşların yorumlarından bağımsız yürütülmesi” biçiminde ifade edilebilir.
3.5 Program değerlendirmecinin uygulama sürecinde etik ilkelere bağlı kalması	<ul style="list-style-type: none">Gösterge “<i>Etik Kurallara Uygunluk</i>” standardı ve göstergeleri ile bir binişiklik göstermektedir.
4.2 Program değerlendirme raporunun değerlendiricinin nesnellliğini yansıtması	<ul style="list-style-type: none">Değerlendiricinin nesnellığı değil, nesnel değerlendirme söz konusudur. Gösterge “Değerlendirmenin kişisel görüşler katılmaksızın nesnel bir anlayışla yürütülmesi” şeklinde düzeltilir.
5.1 Çözümler sonucu ulaşılan tüm bulguları eksiksiz olarak yazılı ifade etme	<ul style="list-style-type: none">Tüm bulguların paydaşlarla paylaşılmasını uygun bulmuyorum.<i>Adil Değerlendirme</i> standartındaki göstergelere ek olarak “Programın gelişmeye açık yönlerini belirtme” “Programın tehditleri ve fırsatlarının saptanması” göstergeleri eklenebilir.

Tablo 45’de görüldüğü gibi, uzmanlardan gelen öneriler incelendiğinde alan uzmanlarının iki göstergenin yeniden düzenlenmesine dönük öneri geliştirdiği görülmektedir. Uzmanlar gösterge 3.4’ün “Uygulamanın paydaşların yorumlarından bağımsız yürütülmesi” ve gösterge 4.2’nin ise “Değerlendirmenin kişisel görüşler katılmaksızın nesnel bir anlayışla yürütülmesi” olarak değiştirilmesi gerektiğini önermişlerdir. Bu önerilerin yanı sıra bir uzmandan “*Adil Değerlendirme*” standartındaki göstergelere “Programın gelişmeye açık yönlerini belirtme” ve “Programın tehditleri ve fırsatlarının saptanması” gibi iki göstergenin eklenmesi önerisi gelmiştir. Bu önerilerin dışında uzmanlar bazı gösterge ifadelerinin diğer göstergeler ile

binişik olduğu yönünde uyarılarda bulunmuşlardır. Örneğin gösterge 1.6'nın bir önceki gösterge ifadesiyle aynı içeriğe sahip olduğu, yine gösterge 3.3'ün ise bir sonraki gösterge ile binişik olduğu yorumları gelmiştir. Diğer taraftan bu göstergelere ilişkin uzman görüşleri incelediğinde her iki göstergenin de uzmanlar tarafından Türkiye için uygun program değerlendirme standartı göstergeleri olmadığı görülmektedir.

4.6.4. Alan Uzmanlarının “Doğruluk” Standartlarına İlişkin Önerileri

Araştırma kapsamında uzmanlardan değerlendirdikleri “Doğruluk” standartları ve göstergelerine ilişkin olarak varsa önerilerini belirtmeleri istenmiştir. Bu standartlara ilişkin olarak 19 uzman 12 göstergeye yönelik öneride bulunmuştur. Uzmanlardan gelen bu önerilerden aynı göstergeye gelen öneriler bir araya getirilmiş ve Tablo 46’da verilmiştir.

Tablo 46.

Alan Uzmanlarının “Doğruluk” Standartlarının Göstergelerine İlişkin Yorumları ve Önerileri

Göstergeler	Gelen Yorumlar ve Öneriler
1.1 Programla ilgili farklı kaynaklardan bilgi ve belgeleri toplama	<ul style="list-style-type: none"> Gösterge bu haliyle “Değerlendirilecek Programın Tanımlanması” standartı altına yer almamalıdır. Ya da göstergenin “Programla ilgili” ifadesi “Programın tanımlanmasında” ifadesiyle değiştirilmelidir.
1.4 Programın var olan görevlerini belirleme	<ul style="list-style-type: none"> Gösterge “Programın var olan işlevlerini belirleme” olarak yeniden düzenlenmelidir.
2.1 Kurum kayıtları, paydaş ve destekleyici görüşleri, raporlar vb. kaynaklardan yararlanarak programın durumunu betimleme	<ul style="list-style-type: none"> “...durumunu...” sözcüğü yerine “uygulandığı ortamı ayrıntılı olarak ifade etme” kullanılabilir. Durum sözcüğü çok anlam ifade etmiyor.
2.2 Programın felsefi ve kuramsal temelleri ile yasal, sosyal ve ekonomik boyutlarını belirleme	<ul style="list-style-type: none"> Değerlendirilecek program zaten felsefi ve kuramsal temelleri dikkate alınarak oluşturulduğu için, bu özelliklere uygun olup olmadığı yönünde bir gösterge oluşturulabilir ya da gösterge “Programın felsefi ve kuramsal temelleri ile yasal, sosyal ve ekonomik yönden uygunluğunu belirleme” olarak değiştirilebilir.
5.1 Program değerlendirme ile ilgili geçerli ve güvenilir bilgileri toplama	<ul style="list-style-type: none"> D6 bölümüyle ilgili olduğundan buradan çıkarılmalı Gösterge verilerin tanımlanması ile ilgili değil, toplanması ile ilgilidir.
6.1 Program değerlendirme araştırmasının amaçlarına uygun nitel ve nicel araştırma süreçleri belirleme	<ul style="list-style-type: none"> Verilerin geçerlik ve güvenilirliğinde bence ilk gösterge “veri toplama araçlarının geçerlik ve güvenilirliğinin sağlanması” olmalıdır. Bu gösterge standartın bütünlüğünü bozmaktadır.
7.2 Nitel verilerin çözümlenmesi için uygun nitel veri çözümleme tekniklerini belirleme	<ul style="list-style-type: none"> Burada tekniklerin ismi verilmelidir (tipolojik analiz, yorumcu analiz, polivokal analiz vb.)
7.3 Veri girişlerinin uzmanlarca yapılmasını sağlama	<ul style="list-style-type: none"> Veri girişleri dünyada her zaman araştırmacının kendisi tarafından yapılır. Göstergenin başına “değerlendirmenin bir proje olarak yürütüldüğü durumlarda” ibaresi eklenmelidir.
8.4 Bulguların alternatif yorumlamalarına yer verme	<ul style="list-style-type: none"> Alternatif yorumdan amaç nedir açık değil. Verilerin Geçerliği ve Güvenirliği standartının göstergeleri arasına “Deneme uygulamasının değerlendirilmesi” göstergesinin de eklenmesi

Tablo 46’da da görüldüğü gibi, uzmanlardan gelen öneriler incelendiğinde üç göstergeye ilişkin değişiklik önerisi geldiği, yeni bir göstergenin önerildiği görülmektedir. Uzmanlar gösterge 1.4’ün “Programın var olan işlevlerini belirleme” olarak, gösterge 2.2’nin “Programın felsefi ve kuramsal temelleri ile yasal, sosyal ve ekonomik yönden uygunluğunu belirleme” olarak ve gösterge 6.1’in “Veri toplama araçlarının geçerlik ve güvenilirliğinin sağlanması” olarak değiştirilmesi önerilerini getirmişlerdir. Nitekim uzmanların bu göstergenin uygunluğuna ilişkin görüşleri incelendiğinde bu göstergenin Türkiye için uygun bulunmaması bu öneriyi desteklemektedir. Ayrıca bir uzman “*Verilerin Geçerliliği ve Güvenirliği*” standartının göstergelerine ek olarak “Deneme uygulamasının (pilot) değerlendirilmesi” göstergesinin de eklenmesini önermektedir. Bu önerilerin yanı sıra uzmanlar gösterge 1.1’in ve gösterge 5.1’in buldukları standartın altından çıkarılmaları gerektiğini belirtmektedirler. Öte yandan uzmanlar bazı göstergelerin bazı ifadelerle desteklenmesi gerektiğini de belirtmişlerdir. Gösterge 7.2 ve 7.3 bu göstergeler arasındadır. Nitekim eksik olduğu vurgulanan bu göstergelerden gösterge 7.3 uzmanlar tarafından uygun program değerlendirme göstergesi olarak kabul edilmemiştir.

5. SONUÇLAR, TARTIŞMA VE ÖNERİLER

Bu bölümde, araştırmanın sorunu, amacı, yöntemi ve bulguları özetlenmiş, araştırmada ulaşılan sonuçlara, bu sonuçlara dayalı tartışmalara ve sorunun çözümüne ve ileriki araştırmalara dönük önerilere yer verilmiştir.

5.1. Sonuç

Günümüzde eğitim kurumlarının en önemli amaçlarından biri eğitimde niteliğin geliştirilmesidir. Eğitimin niteliğinin geliştirilmesi ise, hazırlanan eğitim programlarının başarısıyla doğrudan ilişkilidir. Çünkü eğitim programları, yetiştirilecek insan özelliklerinin belirleyicisidir. Bir ülkenin, bir kurumun eğitim programlarına bakılarak bu programların uygulanması sonucunda ortaya çıkacak eğitim çıktıları konusunda yordamlar yapılabilir. Bu nedenle uygulanan programların aksaklık ve eksiklikleri giderilmeli, toplumdaki ve bilim alanındaki değişmelere göre yeniden düzenlenmeli diğer bir ifadeyle program sürekli değerlendirilmelidir. Öte yandan eğitim programları gelişigüzel değerlendirilemez. Çünkü eğitim kurumlarında uygulanan eğitim programları, toplumun gereksinimleri ve bireylerin özelliklerini göz önünde bulunduran belli ölçütlere göre hazırlanır. Bu ölçütler, programları daha etkili, işlevsel hale getirecek, daha sağlıklı kararların alınabilmesini sağlayacak durumları içerir. Hazırlanan programların bu ölçütlere uygun olup olmadığı ve programları daha etkili hale getirecek doğru kararların alınabilmesi, bu kararların dayanaklarının bilimsel çalışmalarla araştırılmasına ve uygulamaların değerlendirilmesine bağlıdır. Uygulanan programlar değerlendirildiği sürece, eğitim kurumlarının işlevlerini yerine getirip getirmediği kontrol edilmiş ve eğitimin daha nitelikli duruma getirilmesi için programların geliştirilmesine olanak sağlanmış olur.

Program değerlendirmenin etkili olarak gerçekleştirilebilmesi için tıpkı program geliştirme sürecine benzer süreçlerden geçirilmesi; plânlanması, uygulanması ve değerlendirilmesi gerekmektedir. Çünkü değerlendirmenin tasarlanmasından sonuçlandırılmasına kadar olan süreçte birbiriyle ilişkili pek çok öge ve işlem yer almaktadır. Bu işlemlerden bazıları değerlendirme sürecinde alınması gereken kararları

yansıtırlar ve o nedenle de program değerlendirmenin temelini oluştururlar. Program değerlendirmenin bu sistematik içinde yürütülebilmesi için başta Amerika Birleşik Devletleri olmak üzere birçok gelişmiş ülke program değerlendirme standardı, yönergesi ya da ilkesi geliştirmiştir. Sağlam bir program değerlendirme uygulaması için geliştirilen bu standartlar değerlendirmenin, yararlık, uygunluk, yürütülebilirlik ve doğruluk gibi standartları içeren temel özellikleri dikkate alınarak geliştirilmiştir. Geliştirilen bu standartlar genel olarak öğretmenler, yöneticiler, okul müdürleri, değerlendirme eğiticileri, program geliştirme uzmanları, karar vericiler, personel yöneticileri, danışmanlar, toplum liderleri, iş ve eğitim birlikleri vb. tarafından kullanılmak üzere alana kazandırılmıştır.

Program değerlendirme standartlarının geliştirilmesine dönük ilk çalışmalar; 1960'lı yıllarda değerlendirmeciler tarafından yayınlanmayan ve bireysel olarak farklı değerlendirme standartlarını içeren kontrol listelerinin oluşturulması ile başlamıştır.. Daha sonraki yıllarda Amerika Birleşik Devletlerinde bir grup kuruluş “Eğitsel ve Psikolojik Test ve Yönergeler İçin Standartlar” geliştirmiştir. 1975 yılındaki bu çalışma program değerlendirme standartlarının geliştirilmesi için bir temel oluşturmuştur. Bu kuruluşların önderliğinde oluşturulan Ortak Komite tarafından 1981 yılında “Eğitsel Programların, Projelerin ve Gereçlerin Değerlendirmesi Standartları” yayınlanmıştır. 1980'lerde program değerlendirme uygulamalarındaki nicel artış, program değerlendirici olmak isteyen öğrencilerin ve öğretmenlerin eğitimi, değerlendirmeye ilişkin alanyazının genişlemesi, nitel değerlendirme, meta-değerlendirme, içerik, ölçüt ve yapı geçerlik kavramlarının sık vurgulanması vb. etkisiyle 1989 yılında kurul, bu standartların gözden geçirilmesine karar vermiş ve 1994 yılında standartları güncellemiştir. Bu tarihten sonra İsviçre, Almanya, Kanada, Avustralya gibi birçok ülke ya bu standartları kendi ülkelerinde kullanmak üzere kendi dillerine uyarlamış ya da bu standartlara dayalı olarak kendilerine özgü yeni standartlar geliştirmişlerdir.

Türkiye’de program değerlendirme standartlarının geliştirilmesine ilişkin herhangi bir çalışma bulunmamaktadır. Ancak son yıllarda eğitim programlarının değerlendirilmesine yönelik alanyazında oldukça fazla sayıda program değerlendirme çalışmasının yapıldığı görülmektedir. Ancak bu çalışmaların genellikle programın bir boyutuna odaklanmış olduğu, çoğunun nicel yöntemler ve genellikle yalnızca anketler

yoluyla yürütüldüğü, program değerlendirme sürecini tamamen yansıtmadığı, genellikle akademik kaygılarla derslerin etkililiğine yönelik ve okul, bölüm, fakülte veya il düzeyinde yürütülen çalışmalar niteliğinde olduğu ve bulgularının genellikle yetersiz olduğu görülmüştür. Bu sorunlar Türkiye’de program değerlendirme çalışmalarında izlenilmesi gereken değerlendirme ölçüt ve standartlarının eksikliğini de ortaya çıkarmaktadır. Bu eksikliğin giderilmesi için ilk olarak ABD’de Ortak Komite tarafından geliştirilmiş olan ve daha sonra İsviçre, Almanya, Avustralya, Kanada ve Kenya (Afrika ülkelerinde kullanılmak üzere) tarafından kendi ülkelerine uyarlanan program değerlendirme standartları temel alınarak Türkiye’deki program değerlendirme çalışmaları için kullanılacak program değerlendirme standartlarının oluşturulmasına gereksinim vardır. Bu araştırma bu gereksinimi karşılamaya dönük olarak hazırlanmıştır.

Bu araştırma, tarama modellerinden tekil ve ilişkisel tarama modelleri kullanılarak gerçekleştirilmiştir. Araştırmanın çalışma evrenini doktora eğitim programları ve öğretim alanında tamamlamış ve 2008–2009 öğretim yılında devlet üniversitelerine bağlı eğitim fakültelerinde görev yapan öğretim üyeleri oluşturmaktadır. Çalışma evrenindeki öğretim üyelerinin belirlenmesi için Yükseköğretim Kurulu’nun 2008–2009 öğretim yılı istatistiklerine dayalı olarak Türkiye’deki 94 devlet üniversitesi bünyesinde yer alan 64 eğitim fakültesinin web sayfaları ve öğretim elemanı veri tabanları incelenmiştir. Sınırlılıklar dahilinde yapılan inceleme sonucu doktora eğitim programları ve öğretim alanında yapmış ve halen devlet üniversitelerinin eğitim fakültelerinde görev yapmakta olan 158 öğretim üyesi araştırmanın çalışma evrenini oluşturmuştur. Araştırmanın çalışma evreni içinde yer alan 158 uzmandan 138’ine (%87) ulaşılabilmektedir.

Araştırmanın amacı kapsamında gerekli olan veriler, araştırmacı tarafından eğitim programları ve öğretim uzmanlarına uygulanmak üzere hazırlanan “Türkiye İçin Program Değerlendirme Standartları Oluşturma Çalışması Anketi” aracılığıyla web üzerinden toplanmıştır. Toplanan verilerin çözümlenmesinde göstergelerin ve standartların uygunluğunun belirlenmesinde sayı, yüzde ve aritmetik ortalamalar; standartların uygunluk düzeyinin standart alanlarına göre farklılık gösterip

göstermediğinin belirlenmesine ilişkin olarak da tek yönlü varyans analizi ve Tukey HSD testlerinden yararlanılmıştır.

Araştırmaya katılan alan uzmanlarının veri toplama aracında yer alan program değerlendirme standartlarının ve göstergelerinin Türkiye için uygunluk düzeyine ilişkin görüşlerinin çözümlenmesinden elde edilen sayısal verilere dayalı olarak bulgular, araştırma amaçlarına uygun başlıklar altında toplanarak yorumlanmıştır. Araştırmada bu bulgulara dayalı olarak şu sonuçlara ulaşılmıştır:

1. Program değerlendirmede “*Yararlık*” standartlarının uygunluk düzeyine ilişkin uzman görüşlerine dayalı bulgulardan ulaşılan sonuçlar:

- Program değerlendirmede “*Yararlık*” standartlarının uygunluk düzeyine ilişkin alan uzmanlarının görüşlerine göre “*Yararlık*” standartları altında yer alan “*Paydaşların Tanımlanması*”, “*Değerlendirici Yeterliği ve Güvenirliği*”, “*Verilerin Kapsamı ve Seçimi /Yararlığı*”, “*Raporun Kapsamlılığı ve Anlaşılabilirliği*” ve “*Sonuçların Yararlılığı/Etkisi*” standartları Türkiye için “uygun” bulunmuş, bu standartlar altında yer alan 24 göstergeden 21’i uzmanlar tarafından “uygun” bulunmuş, diğer üç gösterge ise uzmanlar tarafından uygun bulunmamıştır. Alana ilişkin genel ortalama (4.10) bu standart alanının beş standart ve 21 göstergesi ile birlikte uzmanlar tarafından Türkiye için uygun görüldüğünü ortaya koymaktadır. Uzmanlar tarafından uygun bulunmayarak program değerlendirme standartları ve göstergeleri listesinden çıkarılan göstergeler şunlardır:

- Program değerlendirme araştırmasında sonuçların uygulanmasından sorumlu grup ya da bireyleri belirtme
- Program değerlendirme araştırmasında sonuçların yayınlanmasından sorumlu grup ya da bireyleri belirtme
- Bireysel sınırlılıkların farkına varma, bu sınırlılıkları aşan durumlarda uzman yardımı alma

Alan uzmanlarının görüşlerine göre ortaya çıkan “*Yararlık*” standartları ve göstergeleri EK 1’de verilmiştir.

2. Program deęerlendirmede “Yürütülebilirlik” standartlarının uygunluk düzeyine ilişkin uzman görüşlerine dayalı bulgulardan ulaşılan sonuçlar:

- Program deęerlendirmede “Yürütülebilirlik” standartlarının uygunluk düzeyine ilişkin alan uzmanlarının görüşlerine göre “Yürütülebilirlik” standartları altında yer alan “Uygun Yöntemler”, “Maliyet-Yarar Çözümlemesi” ve “Deęerlendirme Araştırmasının Yönetimi” standartları Türkiye için “uygun” bulunmuş, “Grup ve Kurum Desteęi” standardı ise “kısmen uygun” bulunmuştur. Dört standart altında yer alan 17 göstergeden 13’ü uzmanlar tarafından “uygun” görülmüş, dört gösterge ise uygun görülmemiştir. Alana ilişkin genel ortalama (3.96) bu standart alanının dört standart ve 13 göstergesi ile birlikte uzmanlar tarafından Türkiye için uygun görüldüğünü ortaya koymaktadır. Uzmanlar tarafından uygun bulunmayarak program deęerlendirme standartları ve göstergeleri listesinden çıkarılan göstergeler şunlardır:

- Program deęerlendirme sürecinden doğrudan ya da dolaylı etkilenebilecek kişi, kurum ve kuruluşların desteęini alma
- Veri toplama kaynaklarının (kişi, grup, kurum ve kuruluşlar) desteęini alma
- Program deęerlendirme araştırmasının yarar-maliyet ilişkisini ortaya koyma
- Deęerlendirme çalışmasının sonuçlarının paydaşlarla paylaşılmasını ve raporlaştırılmasını sağlama

Alan uzmanlarının görüşlerine göre ortaya çıkan “Yürütülebilirlik” standartları ve göstergeleri EK 1’de verilmiştir.

3. Program deęerlendirmede “Uygunluk” standartlarının uygunluk düzeyine ilişkin uzman görüşlerine dayalı bulgulardan ulaşılan sonuçlar:

- Program deęerlendirmede “Uygunluk” standartlarının uygunluk düzeyine ilişkin alan uzmanlarının görüşlerine göre “Uygunluk” standartları altında yer alan “Resmi Sözleşme”, “Etik Kurallara Uygunluk”, “Adil Deęerlendirme”, “Şeffaf Raporlaştırma”, “Sonuçların Açıklanması” ve “Finansal Uygunluk” standartları Türkiye için “uygun” bulunmuş, bu standartlar altında yer alan 33 göstergeden 30’u uzmanlar tarafından “uygun” bulunmuş, dięer üç gösterge ise uzmanlar tarafından uygun

bulunmamıştır. Alana ilişkin genel ortalama (4.21) bu standart alanının altı standart ve 30 göstergesi ile birlikte uzmanlar tarafından Türkiye için uygun görüldüğünü ortaya koymaktadır. Uzmanlar tarafından uygun bulunmayarak program değerlendirme standartları ve göstergeleri listesinden çıkarılan göstergeler şunlardır:

- Süreç içinde hizmetlerin ve harcamaların ayrıntılı kayıtlarını tutma
- Paydaşların beklenti ve isteklerinden etkilenmemesi
- Ulaşılan sonuçların ve sınırlılıkların dayanaklarını belirtme

Alan uzmanlarının görüşlerine göre ortaya çıkan “*Uygunluk*” standartları ve göstergeleri EK 1’de verilmiştir.

4. Program değerlendirmede “*Doğruluk*” standartlarının uygunluk düzeyine ilişkin uzman görüşlerine dayalı bulgulardan ulaşılan sonuçlar:

- Program değerlendirmede “*Doğruluk*” standartlarının uygunluk düzeyine ilişkin alan uzmanlarının görüşlerine göre “*Doğruluk*” standartları altında yer alan “*Değerlendirilecek Programın Tanımlanması*”, “*Durum Çözümlemesi*”, “*Değerlendirme Araştırmasının Amacının Belirlenmesi*”, “*Değerlendirme Araştırmasının Yönteminin Belirlenmesi*”, “*Veri Kaynaklarının Tanımlanması*”, “*Verilerin Geçerliliği ve Güvenirliği*”, “*Verilerin Çözümlemesi*” ve “*Sonuçların Gerekçelenirilmesi*” standartları Türkiye için “uygun” bulunmuş, bu standartlar altında yer alan 36 göstergedan 30’u uzmanlar tarafından “uygun” bulunmuş, diğer altı gösterge ise uzmanlar tarafından uygun bulunmamıştır. Alana ilişkin genel ortalama (4.09) bu standart alanının sekiz standart ve 30 göstergesi ile birlikte uzmanlar tarafından Türkiye için uygun görüldüğünü ortaya koymaktadır. Uzmanlar tarafından uygun bulunmayarak program değerlendirme standartları ve göstergeleri listesinden çıkarılan göstergeler şunlardır:

- Program değerlendirme araştırmasının amaçlarına uygun nitel ve nicel araştırma süreçleri belirleme
- Veri toplama araçları ile toplanacak verilerin kapsamını tanımlama
- Veri girişlerinin uzmanlarca yapılmasını sağlama
- Çözümlemeler sonucu ortaya çıkan veri tablolarını gözden geçirme ve

eksiklikleri giderme

- Olanaklar ölçüsünde verilerin doğrulanması ve veri tablolarının gözden geçirilmesi için farklı uzman görüşlerine başvurma
- Ulaşılan sonuçlara ilişkin kanıtlar ortaya koyma

Alan uzmanlarının görüşlerine göre ortaya çıkan “Doğruluk” standartları ve göstergeleri EK 1’de verilmiştir.

5. Program değerlendirme standartları uygunluk düzeyinin standart alanlarına göre farklılık gösterip göstermediğinin belirlenmesine ilişkin bulgulardan ulaşılan sonuçlar:

- Program değerlendirme standartları dört standart alanından oluşmaktadır. Uzmanların görüşlerine göre program değerlendirme standart alanlarının uygunluk düzeyi istatistiksel olarak anlamlı farklılık göstermektedir. Bu fark, “Yararlık” standartları ile “Yürütülebilirlik” standartları; “Yararlık” standartları ile “Doğruluk” standartları; “Yürütülebilirlik” standartları ile “Uygunluk” standartları ve “Uygunluk” standartları ile “Doğruluk” standartları alanları arasındaki farklılıktan kaynaklanmaktadır.

6. Program değerlendirme standartları ve göstergelerine ilişkin uzmanların önerilerine dönük bulgulardan ulaşılan sonuçlar:

- Uzmanların “Yararlık” standartlarına ve göstergelerine yönelik önerileri incelendiğinde uzmanlar:
 - gösterge ifadelerinin “ve” bağlacıyla belirtilerin yargılarının iki/ya da daha fazla göstergeyle belirtilmesi gerektiğini (gösterge 1.2, 2.2, 2.4, 4.1, 5.2),
 - gösterge 1.2’in “Değerlendirme uzmanının tanımlanması” olarak değiştirilmesini,
 - gösterge 3.1’in “Değerlendirme sorularını yanıtlayıcı yararlı bilgileri seçme” olarak değiştirilmesini önermişlerdir.
- Uzmanların “Yürütülebilirlik” standartlarına ve göstergelerine yönelik önerileri incelendiğinde uzmanlar;

- gösterge 1.1'in "Değerlendirme amacına uygun yöntem belirleme" olarak değiştirilmesini,
 - gösterge 4.1'in "İlgili programın değerlendirilebilirliğini standartlarla ortaya koyma" olarak değiştirilmesini,
 - gösterge 2.1, 2.2, ve 2.3'te ifadelerin öznelinin kısaltılmasını önermişlerdir.
 - Uzmanların "*Uygunluk*" standartlarına ve göstergelerine yönelik önerileri incelendiğinde uzmanlar;
 - gösterge 3.4'ün "Uygulamanın paydaşların yorumlarından bağımsız yürütülmesi" olarak değiştirilmesini,
 - gösterge 4.2'nin "Değerlendirmenin kişisel görüşler katılmaksızın nesnel bir anlayışla yürütülmesi" olarak değiştirilmesini,
 - "*Adil Değerlendirme*" standardındaki göstergelere "Programın gelişmeye açık yönlerini belirtme" ve "Programın tehditleri ve fırsatlarının saptanması" gibi iki göstergenin eklenmesini önermişlerdir.
- "*Doğruluk*" standartlarına ilişkin olarak uzmanlardan gelen öneriler incelendiğinde üç göstergeye ilişkin değişiklik, bir göstergenin eklenmesi ve iki göstergenin çıkarılması önerisi görülmektedir. Uzmanlar;
 - gösterge 1.4'ün "Programın var olan işlevlerini belirleme" olarak değiştirilmesini,
 - gösterge 2.2'nin "Programın felsefi ve kuramsal temelleri ile yasal, sosyal ve ekonomik yönden uygunluğunu belirleme" olarak değiştirilmesini,
 - gösterge 6.1'in "Veri toplama araçlarının geçerlik ve güvenilirliğinin sağlanması" olarak değiştirilmesini,
 - "*Verilerin Geçerliği ve Güvenirliği*" standardı altına yeni bir gösterge olarak "Deneme uygulamasının (pilot) değerlendirilmesi" göstergesinin eklenmesini,
 - gösterge 1.1'in ("Programla ilgili farklı kaynaklardan bilgi ve belgeleri toplama") "*Doğruluk*" standart alanından çıkarılmasını,
 - gösterge 5.1'in ("Program değerlendirme ile ilgili geçerli ve güvenilir

bilgileri toplama”) “Doğruluk” standart alanından çıkarılmasını önermişlerdir.

5.2. Tartışma

Türkiye için program değerlendirme standartlarının ve göstergelerinin oluşturulmasının amaçlandığı bu araştırmada; standart alanlarının, bu alanlar içinde yer alan standartların ve standartlara ait göstergelerin tanımlanmasında çeşitli ülkelerde kullanılmakta olan program değerlendirme standartlarından yararlanılmıştır:

- ABD’de Ortak Komite tarafından 1981 yılında geliştirilen gözden geçirilip bazı maddeleri yenilenerek 1994 yılında yayınlanan, dört standart alanı ile 30 standarttan oluşan program değerlendirme standartları.
- Ortak Komite’nin geliştirdiği standartları kendi ülkesine uyarlayan İsviçre program değerlendirme standartları, Almanya program değerlendirme standartları ve Kenya program değerlendirme standartları. İsviçre program değerlendirme standartları dört temel alanda 26 standarttan; Almanya program değerlendirme standartları dört temel alanda 25 standarttan oluşmaktadır. Afrika ülkelerinde kullanılmak üzere uyarlanan Kenya program değerlendirme standartları da dört temel alanda 30 standarttan oluşmaktadır.
- ABD (Amerikan Değerlendirme Derneği), Fransa, Kanada ve Avustralya’da geliştirilen program değerlendirme standartlarıdır. Bu ülkelere ait program değerlendirme standartlarından özellikle standartların ifade edilmesinde ve göstergelerinin oluşturulmasında yararlanılmıştır.

Bu kaynaklarda yer alan standart alanları ve standartlar çalışmanın kuramsal bölümünde tartışılmış, ABD, İsviçre, Almanya ve Afrika program değerlendirme standartlarının karşılaştırmalı bir dökümü Tablo 10’da yapılarak örtüşen, ayrışan, benzeşen ve farklılaşan standart alanları ve standartlar belirlenmiştir. Bu döküme göre İsviçre, Almanya ve Afrika ülkelerinde kullanılan program değerlendirme standart alanlarının tümünün, standartların da büyük çoğunluğunun uyarlama çalışmalarında temel alınan Ortak Komite’nin program değerlendirme standart alanları ve standartları ile aynısı ya

da çok benzeri olduğundan, bu araştırmada da Ortak Komite'nin geliştirdiği program değerlendirme standartları aynen benimsenmiştir. Ancak standartların seçiminde, Türkiye'deki kültürel özelliklere ve değerlerle çatışmayan, Türkiye'deki araştırma misyonuna uygun olan maddelerin seçilmesine dikkat edilmiştir. Aynı şekilde standartların altında yer alan göstergelerin belirlenmesinde de kültürel değerlere uygunluğa önem verilmiştir. Çünkü Taut (2001) standartların temelinde yatan örtük kültürel değerlerin, standartların diğer kültürlerle aktarımında önemli olduğunu, bu değerlerin dikkate alınmaması halinde aktarımın güçleşebileceğini ifade etmektedir. Berry ve diğ (1997) de değer yargılarının ülkeden ülkeye farklılık gösterebileceğini, bunun temelinde ise kültür olduğunu belirtmiştir. Bu bağlamda standartların altına, standartın kapsam ve kastettiği, Türkiye'deki koşullar ve uygulamalar, anlaşılabilirlik ve uygulanabilirlik, vb. yönler dikkate alınarak göstergeler oluşturulmuştur. Bu çalışmalar sonucunda araştırmacı tarafından tez danışmanının rehberliğinde oluşturulan standart alanları, standartlar ve göstergeleri Tez İzleme Komitesi üyeleri tarafından anlaşılabilirlik, açıklık, uygunluk, uygulanabilirlik, işlevsellik vb. yönlerden incelenmiş, gerekli düzeltmeler, çıkarmalar ve eklemeler yapılmıştır. Tez İzleme Komitesi toplantılarında, araştırmacı tarafından oluşturulan program değerlendirme standart alanlarından *Yararlılık* standartları ve göstergeleri uygun bulunmuş; *Yürütülebilirlik*, *Uygunluk* ve *Doğruluk* standartlarında ise bazı değişiklik önerileri olmuştur.

Tez İzleme Komitesi toplantılarında *Yürütülebilirlik* standartları arasında yer alan *Politik Yeterlik* standartının değiştirilmesi önerilmiştir. Değerlendiricinin paydaşlar ve karar vericiler arasında denge sağlaması, her iki grubun beklenti ve taleplerini diplomatik yollarla karşılaşmasını içeren bu standart, komite tarafından "politik" sözcüğünün Türkiye kültüründe siyasi bir anlamı olmasından dolayı benimsenmemiştir. Nitekim Taut (2001) OKPDS'yi incelerken, bu standartın tümüyle kültürel öğeler taşıdığını, egalitaryan kültüre (siyasal ve sosyal eşitliğin ağır bastığı kültürler) uygun olduğunu, hiyerarşik kültüre sahip ülkelerde baskın (siyasi) erkin değerlendiriciyi ve değerlendirme sonuçlarını etkileyebileceğini ifade etmiştir. Rouge'da (2004) Taut'u destekleyerek, değerlendiricinin belirli kurumlar tarafından görevlendirildiği durumlarda kurumlar arası hiyerarşik yapıdan dolayı birtakım çıkar çatışmalarının yaşanabileceğini belirtmiştir. Buna karşın Beywl (2000) değerlendiricinin her kültürde

bu tür sorunlarla karşılaşabileceğini, bu tür sorunların üstesinden gelmesi için diplomatik yollar izlemesi gerektiğini savunmaktadır. Almanya da bu standarta ilişkin olarak benzer kaygıları yaşamış ve standartın adını uyarlama çalışmasında “*Diplomatik Yeterlik*” olarak değiştirmiştir. Sonuç olarak, komite *Politik Yeterlik* standartının *Grup ve Kurum Desteği* olarak değiştirilmesini ve bu standart altında yer alacak göstergelerin bu amaca dönük olarak oluşturulmasını önermiştir.

Tez İzleme Komitesi toplantılarında tartışılan diğer standart da araştırmacı tarafından *Uygunluk* standartları arasında yer verilen *Duyarlı ve Kapsamlı Yönlendirme, İnsan Hakları ve Saygı* ile *Birey Etkileşimleri* standartlarıdır. Komite üyeleri bu standartların özellikle içerik yönünden birbirini tamamlar nitelikte olduğunu düşünerek üç standartın tek bir standart altında birleşebileceği önerisi getirmiştir. Komite tarafından önerilen standart program değerlendirmede etik ilkeleri içeren *Etik Kurallara Uygunluk* standartıdır.

Tez İzleme Komitesi üyelerinin değişiklik önerdiği diğer standart alanı *Doğruluk* standartlarıdır. Komite üyeleri, araştırmacı tarafından OKPDS’den aktarılan *Geçerli Bilgi ve Güvenilir Bilgi* standartlarının *Verilerin Geçerliliği ve Güvenirliği* adı altında; *Nitel Bilginin Analizi ve Nicel Bilginin Analizi* standartlarının *Verilerin Çözümlemesi* adı altında ve son olarak *Savunulabilir Bilgi Kaynakları ve Sistemik Bilgiler* standartlarının ise *Veri Kaynaklarının Tanımlanması* adı altında birleştirilmelerini önermiştir.

Araştırmacı tarafından oluşturulan ve Tez İzleme Komitesi’nin önerileri ve düzeltmeleri sonucunda, kapsam geçerliği için uzman görüşüne sunulabilecek program değerlendirme standartlarını ve göstergelerini içeren bir veri toplama aracı taslağı ortaya çıkmıştır.

Kapsam geçerliği için görüşlerine başvuru alan uzmanlar (13 eğitim programları ve öğretim uzmanı) *Yararlık* standartlarından *Paydaşların Tanımlanması* standartı altında yer alan altı göstergeden üçünü uygun görmüşler, iki göstergenin yeniden tanımlanmasını, bir göstergenin de çıkarılmasını önermişlerdir. Uzmanların veri toplama aracından çıkarılmasını önerdiği gösterge altında yer aldığı standartlar aynı adı

çağrıştıran “Paydaşları Tanımlama” göstergesidir. *Paydaşların Tanımlanması* standardı altında yer alan ve değiştirilmesi önerilen göstergeler ise “Değerlendirmeyi destekleyenleri tanımlama”, “Değerlendirme sonuçlarının uygulanması ve yorumlanmasından sorumlu grup ya da bireyleri belirtme”dir. Bu göstergelerden “Değerlendirmeyi destekleyenleri tanımlama”, göstergesine ilişkin olarak uzmanlar yalnızca göstergenin ifade biçiminin değiştirilmesi gerektiğini belirtirken, “Değerlendirme sonuçlarının uygulanması ve yorumlanmasından sorumlu grup ya da bireyleri belirtme” göstergesinin iki ayrı gösterge olarak ifade edilmesini önermişlerdir.

Uzmanlar *Yararlık* alanında ikinci standardı olan *Değerlendirici Güvenirliği* standartına “Yeterliği” kavramının da eklenmesini önermişlerdir. Uzmanların bir program değerlendirmesi çalışmasında değerlendiricide güvenirliliğin yanında değerlendirme bilgi ve becerilerinin de aranması gerektiğini düşünmeleri bunda etkili olmuş olabilir. Bu bağlamda Taut (2001) “yeterlik” kavramının farklı kültürlere sahip ülkelerde değerlendiricinin güvenirliliği için bir ön koşul olarak görülemeyebileceğini ancak önemli olduğunu ifade etmektedir. Jang (2000) Kore’de değerlendiricinin seçiminin önemli olduğunu çünkü değerlendirici adının değerlendirme araştırmasının güvenirliliği için yeterli olduğunu ve değerlendiricinin adının değerlendiricinin yeterliklerini de ortaya koyduğunu belirtmektedir. (Beywl, 2000) ise, değerlendiricinin güvenilir olabilmesi için mesleki ve yöntemsel yeterliğe sahip olması gerektiğini belirtmektedir. Bu bağlamda uzmanlar *Değerlendirici Güvenirliği* standartına *Yeterliği* ifadesinin de eklenmesini önermişlerdir.

Uzmanların Türkiye kültürüne uygun olmadığını vurgulayarak taslak veri toplama aracından çıkarılmasını önerdikleri standart “*Değerlerin Belirlenmesi*”dir. Uzmanlar bu standartın uygulanabilirliğinin oldukça güç olduğunu değerlerin kişiden kişiye göre farklılık gösterebileceğini belirtmişlerdir. Ayrıca Türkiye’de program değerlendirme çalışmalarında araştırmacının değerlendirme bulgularını yorumlamada birtakım ölçütlerin olmayışı da bunda etkili olabilir. Taut (2001), değerlendirmenin temelinde yatan değer kavramının bireyden bireye, kültürden kültüre farklı olduğunu belirterek bu standartın diğer ülkelere aktarımında dikkatli olunması gerektiğini savunmuştur. Joint Committee on Standards for Educational Evaluation (1994) veri toplama ve çözümleme süreçlerinin belirli ölçütlere göre yapılmamasının değerlendiricinin bulguları

yorumlamada hangi deęerleri göz önüne alacağını da etkileyeceğini belirtmektedir. Sonuç olarak uzmanlar dięer ülkelerin standartları arasında bulunan “*Deęerlerin Belirlenmesi*” standartının Türkiye’de uygulanabilirliğinin olanaklı olmadığını vurgulamışlardır.

Yararlık standartlarına ilişkin olarak uzmanlar, bazı standartların da içerik yönünden birbirini kapsar nitelikte olduğunu ve birleştirilmesi gerektiğini belirtmişlerdir. Temel alındığı ülkelerde iki farklı standart olarak ifade edilen *Raporun Anlaşırılığı* ve *Rapor Zamanlaması* ve *Yayımlı* standartlarına ilişkin olarak uzmanlar her iki standartın *Raporun Kapsamlılığı* ve *Anlaşırılığı* adı altında tek bir standartla ifade edilmesi önerisinde bulunmuşlardır. Uzmanların görüşleri incelendiğinde *Rapor Zamanlaması* ve *Yayımlı* standartının Türkiye için uygulanabilir olmadığı yönünde öneride buldukları görülmektedir. Uzmanların bu görüşlerinin temel nedeni Türkiye’de program değerlendirme çalışmalarının çoğunlukla bireysel olarak yürütülmesi ve bu süreçte değerlendirmeye ilişkin bir takvimin oluşturulmaması olarak açıklanabilir. Nitekim Taut (2001) standartın altında yatan deęerin kültürelliğini ifade ederken “zamanlama” olgusunun kültürel olduğunu, küresel bir deęer olmadığını ve bu nedenle bazı kültürlere aktarımının zor olabileceğini ve standartlar arasından çıkarılması gerekebileceğini belirtmiştir. Patel ve Russon (1999) de bu savı destekleyerek (Afrika standartlarını örnek vererek) Afrikalıların zaman algılarıyla, batılıların zaman algılarının birbirinden çok farklı olduğunu ifade etmiştir.

Uzmanlar *Yararlık* standartlarına ilişkin son olarak, veri toplama aracında yer alan *Deęerlendirme Etkisi* standartıyla ilgili öneride bulunmuşlardır. Uzmanlar *Deęerlendirme Etkisi* standartının adının *Sonuçların Yararlılığı/Etkisi* olarak yeniden adlandırılması gerektiğini ve oluşturulan bu standartın altına üç yeni gösterge eklenmesini önermişlerdir. Uzmanlar oldukça genel olduğu ve standartın amacını yansıtmadığı düşüncesiyle çıkarılmasını önerdikleri eski göstergelerin yerine değerlendiriciye değerlendirme sonuçlarının kullanılması ile ilgili sorumluluk yükleyen yeni göstergeler önermişlerdir (YA.5.1, 5.2, 5.3). Araştırmanın kuramsal bölümünde de açıklandığı gibi Türkiye’de yapılan program değerlendirme çalışmaları incelendiğinde değerlendirme sonuçlarının genellikle karar vericiler tarafından dikkate alınmadığını, çünkü çalışmaların bireysel olarak yapıldığı ve sınırlı olduğu düşünülünce uzmanlar

tarafından önerilen bu standart ve göstergesinin Türkiye için alanda bir yeniliği başlatabileceği söylenebilir. Öte yandan değerlendiricinin değerlendirme sonuçlarının kullanımından sorumlu olmaması gerektiği, bu durumun değerlendiricinin yükünü daha da arttıracığı da savunulmaktadır (Jang, 2000; Lewy, 1984).

Uzmanlar tarafından kapsam geçerliği bağlamında veri toplama aracının *Uygunluk* standartları bölümünden çıkarılması önerilen standartı *Çıkar Çatışması*dır. Diğer ülkeler tarafından da oldukça tartışılan bu standartın uzmanlar tarafından çıkarılma önerisinin temelinde standartın Türk değerlendirme kültürüne uygun olmaması yer alabilir. Uzmanlar, bu standartın Türkiye'deki çalışmaları yansıtmadığı, yapılan çalışmalarda çıkar çatışmasının yaşanmadığını ya da ortaya çıkmadığını belirtmişlerdir. Bu standartın temelinde kültürel öğeler olduğunu belirten Taut (2001) ise standartın mesleki dayanışmanın ve bireysel arkadaşlıkların olduğu kültürlerde uygulanmasının zor olduğunu savunmaktadır. Smith, Chircop ve Mukherjee (1984) de Taut'un bu görüşünü destekleyerek standartın (1981 sürümü) Hindistan gibi siyasi ortama sahip ülkelerde uygulanmasının zor olduğunu, çünkü yapılan değerlendirme çalışmalarının federal hükümet tarafından görevlendirilen belirli bir görüşe sahip değerlendiriciler tarafından yürütüldüğünü ve bu durumda karşıt bir sonucun alınmasının zor olduğunu ifade etmektedir. Smith ve diğ. (1984), Hindistan'da politik sistemin temel kontrol ögesi olduğunu, ABD gibi az partili bir yönetim sistemi olmadığını ve bağımsız bakış açılarının çalışmalarda yansıtılmasının zor olduğunu vurgulamaktadır. Jang (2000) ise her kültürde çıkar çatışmalarının olabileceğini ve her kültürün bu sorunu çözmede kendine uygun bir yol bulduğunu belirtmektedir. Jang (2000), Korelilerin bireyden çok grubun iyiliği için en iyisini seçtiğini, halkın çıkarlarını her şeyin üzerinde tuttıklarını, ailede baba, okulda müdür, sınıfta öğretmeni otorite figür olarak algıladıklarını ve bu nedenlerle çıkar çatışmalarıyla oldukça az karşılaştıklarını ifade etmektedir. Patel ve Russon (1999) da, bu standartın tamamen kültürel özellikler taşıdığını savunarak, kavramın Afrika kültüründe farklı, batı kültüründe farklı değerlendirildiğini belirtmektedir. Patel ve Russon, Afrika'da değerlendiricilerin pek çok kez Birleşmiş Milletler'in programlarını değerlendirmekle görevlendirildiğini, ancak aynı değerlendiricilerin aynı kurumdan ödül vb. beklediğini de ifade etmekte ve yapılan değerlendirme sonuçlarının hiçbir zaman kurumun aleyhine olmadığını savunmaktadır.

Sonuç olarak, Türk uzmanlar, bu standartın Türkiye’de ki siyasi yapıyı, kültürel yapıyı dikkate alarak uygun olmadığı yönünde öneride bulunmuşlardır.

Tez İzleme Komitesi’nce uygun görülen ancak kapsam geçerliği bağlamında uzmanların veri toplama aracından çıkarılmasını önerdikleri standart *Meta-Değerlendirme* standartıdır. Uzmanlar değerlendirmenin değerlendirilmesi olarak tanımlanan bu standartın gereksiz olduğunu, çünkü diğer tüm standartların bu amaçla oluşturulduğunu, buradaki meta-değerlendirme kavramının meta-analiz kavramıyla yorumlanabileceğini ifade etmişlerdir. Uzmanların bu görüşlerine paralel olarak Jang (2000) Korelilerin de meta-değerlendirme standartı ile ilgili olumlu görüş bildirmediklerini, standartı gereksiz bulduklarını, değerlendirmenin değerlendirilmesinin önceki çalışmanın sonuçlarının tatmin edici olmadığı anlamına geldiğini düşündüklerini ifade etmektedir.

Taslak olarak hazırlanan veri toplama aracında uzmanların görüşleri ve önerileri doğrultusunda standartlar ve göstergeleri üzerinde yapılan değişiklikler ve düzeltmelerle son biçimi verilen veri toplama aracı, Tez İzleme Komitesi üyelerinin de olumlu görüş bildirmeleri üzerine veri toplama aracında yer alan standartların ve göstergelerin Türkiye için uygun olup olmadığı konusunda Türkiye’deki tüm eğitim fakültelerinde görev yapan eğitim programları ve öğretim alanındaki uzmanların görüşlerine sunulmuştur. Veri toplama aracını yanıtlayan alan uzmanları, görüşlerine sunulan program değerlendirme standartları ve göstergelerinden 23 standartın 23’ünü, 110 göstergenin 94’ünü Türkiye için uygun kabul etmişlerdir. Uzmanlar, 16 gösterge ile ilgili olarak ya uygun olmadığı için çıkarılması, ya da değiştirilmesi/düzeltilmesi gerektiği veya yerine önerdikleri göstergenin konulmasını önermişlerdir. Türkiye için program değerlendirme standartlarının ve göstergelerinin oluşturulmasında temel alınan ülkelerdeki program değerlendirme standartlarında genellikle göstergeler yer almadığından, alan uzmanları tarafından Türkiye için uygun bulunmayan göstergeler, altında yer aldıkları standart kapsamında tartışılacaktır. Bu bağlamda tartışılacak standartlar *Paydaşların Tanımlanması, Değerlendirici Yeterliliği ve Güvenirliği, Grup Ve Kurum Desteği, Maliyet Yarar Çözümlemesi, Değerlendirme Araştırmasının Yönetimi, Resmi Sözleşme, Adil Değerlendirme, Sonuçların Açıklanması, Verilerin Geçerliliği ve Güvenirliği, Verilerin Çözümlemesi ve Sonuçların Gerekçelenirilmesi* standartlarıdır.

Paydaşların Tanımlanması: Yararlık standartlarının ilki olan ve değerlendirmede yer alacak ve değerlendirmeden etkilenecek tüm bireylerin tanımlanmasını amaçlayan bu standart altında toplam altı gösterge yer almıştır. Bu göstergelerden ikisi uzmanlar tarafından uygun görülmemiştir. Uygun bulunmayan göstergeler standartın beşinci ve altıncı göstergeleri olan “Program değerlendirme araştırmasında sonuçların uygulanmasından sorumlu grup ya da bireyleri tanımlama” ve “Program değerlendirme araştırmasında sonuçların yayınlanmasından sorumlu grup ya da bireyleri tanımlama” göstergeleridir. Bu göstergelerdeki paydaşlar program değerlendirme araştırmasının sonuçlarının uygulanmasından ve yayınlanmasından sorumlu olan kişi ya da kurumlardır. Özellikle bu standartın geliştirildiği ve uyarlandığı ülkelerin değerlendirme kültürü göz önüne alındığında bu ülkelerde değerlendirmenin bireysel olarak değil de grup tarafından ve genellikle bir kurum adına yapıldığı görülmektedir (Fitzpatrick, Sanders ve Worthen, 2004). Bu nedenle bu ülkelerde paydaş kavramı oldukça önemsenmiştir. Taut (2001) paydaş kavramının tüm standartların merkezinde yer aldığını ifade ederek, değerlendirmenin tüm paydaşlarının da bu nedenle tanımlanması gerektiğini belirtmektedir. McLaughlin (1990) değerlendirilecek programla ilgili tüm bireylerin sürece katılımının hayati olduğunu; Sanders (1987) ise değerlendirmenin bilgi toplanacak ve bilgilendirilecek tüm yasal grupları temsil ettiği sürece uygun bir değerlendirme olacağını belirtmektedir. Türkiye’deki uzmanların bu göstergeleri onaylamamalarının nedeni ise Türk değerlendirme kültüründe bu tür paydaşlara olan gereksinimle ilgili olabilir. Çünkü Türkiye’de değerlendirme çalışmaları genellikle yüksek lisans ve doktora tez çalışmaları kapsamında bireysel ve bağımsız yürütülmekte ve bu nedenle sonuçların uygulanmasına ilişkin bilgilere araştırmaların Öneriler bölümünde yer verilmekte; sonuçların paydaşlara, karar vericilere ya da kamuoyuna duyurulması için yayınlanması amacı güdülmemektedir. Uzmanlar bu nedenlerle bu göstergeleri uygun bulmamış olabilir. Patel ve Russon (1999), tüm paydaşların tanımlanması konusunda bir görüş birliği olduğunu, ancak bazı paydaş gruplarının özellikle kültürlere dayalı olarak belirtilmesinin gerekli olmadığını ifade etmektedir. Lee’ye (1995) göre de değerlendirici, paydaşları ne kadar az bilirse, o kadar nesnel sonuçlara ulaşır.

Değerlendirici Yeterliği ve Güvenirliği: Değerlendirmenin güvenirligi ve kabul edilebilirliđi için deđerlendiricinin güvenilir ve alanında yeterli olmasını amaçlayan bu standart alan uzmanları tarafından Türkiye için uygun bulunmuştur. Bununla birlikte uzmanlar, bu standart altında yer alan altı göstergeden “Bireysel sınırlıkların farkına varma, bu sınırlıkları aşan durumlarda uzman yardımı alma” göstergesinin Türkiye için uygulanabilir olmadığına karar vermişlerdir. Bu göstergenin uzmanlar tarafından uygun bulunmamasının nedeni, göstergede yer alan “sınırlık” kavramının deđerlendiricinin güvenirligi ile mi, yoksa yeterliđi ile mi ilgili olduğunun belirgin olmaması ya da kavramla anlatılmak istenenin belirsiz olması olabilir. Jang (2000), deđerlendiricinin yeterliđine ilişkin saptamalarda bulunurken, deđerlendiricinin sınırlılıđının kendisine sunulan kaynakların sınırlılıđıyla ilgili olduğunu belirtmekte ve Kore’de deđerlendiricinin adının çok önemli olduğunu bu nedenle deđerlendiricin neler yapabileceđinin ve yapamayacağıının az çok kestirilebileceđini belirtmektedir. AEA (2005) ise, deđerlendiricinin mesleki eğitimi ve yeterliđi sınırları içinde uygulama yapması gerektiđini ve sınırın aşılması durumunda deđerlendirmeden çekilmesi gerektiđini belirterek deđerlendiricinin sınırlarının çıkar çatışmalarıyla son bulacağını ima etmektedir. AEA (2005) yine bu konuda deđerlendiricinin buradaki sınırlılıđının kültürel yeterlikler konusunda eksikliklerinden de kaynaklanabileceđini ifade etmektedir. Wahba (1990) deđerlendiricinin alanı kadar kültürel bağlamı da bilmesi gerektiđini savunarak özellikle deđerlendiricilerin bu konudaki eksikliklerini gidermek için bir eğitimden geçmeleri gerektiđini belirtmektedir. Bewyl ve Speer (2004), standart bir iş profiline sahip olmadıkları ve üniversitelerde deđerlendirme üzerine çok az ders olduğu için deđerlendiricinin yeterliđinin önemli olduğunu belirtmekte ve özellikle sürekli kendi alanında çalışan ve alanında iyi deneyim sahibi olan bireylerin programın kötü yönlerini görmede körleşebileceđini ve programın yalnızca olumlu yönlerini dikkate alabileceđi tehlikesini öngörmektedir. Bu durum deđerlendiricinin de bazı durumlarda yardıma gereksinim duyabileceđini ve deđerlendirme takımının en azından bir alan uzmanı ve bir deđerlendirme uzmanından oluşması gerektiđini ortaya koymaktadır. Bu nedenle Bewyl ve Speer (2004) kesin bir ifadeyle deđerlendiricilerin kendi bilgi ve becerilerinin sınırlarının farkında olmaları gerektiđini vurgulamaktadırlar. Yazarlar bir deđerlendirme çalışmasında katılımcıların güdülenmesi gerektiđi bir

durumda eğer deęerlendiricinin psikoloji deneyimi yetersiz ise verilerin toplanmasından ya da arařtırmadan çekilmesini buna örnek olarak vermektedirler.

Grup ve Kurum Desteęi: Yürütülebilirlik standartlarının ikincisi olan deęerlendiricinin paydařlar ve karar vericiler arasında denge saęlaması ve her iki grubun beklenti ve taleplerini diplomatik yollarla karřılamasını içeren bu standart veri toplama aracının oluřturulması sürecinde de oldukça tartıřılmış ve standart Tez İzleme Komitesi'nce politik ifadelerden arındırılmıştır. Bu biçimiyle uzmanların görüşlerine sunulan ve “Kısmen Uygun” bulunan bu standart dięer tüm standartlar arasında en düşük aritmetik ortalamaya sahip standart olmuřtur. Alan uzmanları, bu standart altında yer alan dört göstergeden ikisinin Türkiye için uygulanabilir olmadığına karar vermişlerdir. Bu göstergelerin uygun bulunmamasının nedenleri arasında daha önce de açıklandığı gibi standartın ve göstergelerinin hiyerarşik kültüre sahip toplumlarda uygulanmasının güç olması olabilir. Çünkü ast-üst ilişkisi olan kültürlerde ast tarafından yapılan bir çalışmanın sonucunun üstü olumsuz etkilemesi düşünülemez. Bewyl ve Speer (2004), standartın aşırı kültürel özellikler taşıdığını belirterek özellikle Almanya'da kurumların deęerlendirici üzerindeki baskısının hissedilebilir olduğunu vurgulamaktadır. Wholey (1981) deęerlendirme bulgularının nadiren güç yapısını (power structure) deęiřtirebileceğini ifade etmektedir. Kore'de ise tam tersi bir durum görülmektedir. Kore'de uygunluk düzeyi en yüksek çıkan standartlar arasında yer alan bu standarta ilişkin olarak Jang (2000) bazı durumlarda Kore'de bir projenin sonuçlarının politik amaçlarla kullanılabileceğini, deęerlendirme çalışmasının başarıyla tamamlanmasının politik sisteme baęlı olabileceğini ifade etmekte ve sorumlu bir deęerlendirici olmanın karar vericilerle mükemmel bir iletişime baęlı olduğunu vurgulamaktadır. Çünkü çoęu durum da karar vericilerin üst grup olduğunu belirtmektedir. Bewyl ve Speer (2004) ise tüm paydařların resmi sözleşme çerçevesinde deęerlendirme sürecine dahil edilerek sorunun çözülebileceğini belirtmektedir.

Maliyet Yarar Çözümlemesi: Alan uzmanları, deęerlendirmede kullanılacak kaynakların etkili ve verimli kullanılmasını amaçlayan bu standartı, Türkiye bağlamında uygulanabilir olarak deęerlendirmişler ancak standartın üçüncü göstergesi olan “Program deęerlendirme arařtırmasının yarar-maliyet ilişkisini ortaya koyma” göstergesinin uygun olmadığını belirtmişlerdir. Uzmanlarının bu görüşünün nedeni,

standartın diğerk göstergeleri ile anlatılmak istenenin bu gösterge ile özetlendiğı olabilir. Standartın temel amacı dikkate alındığında standartta zaten maliyet-yarar ilişkisinin değerlendirildiğı görülmektedir. Dolayısıyla uzmanlar bu göstergenin gereksiz olduğunu düşünmüş olabilirler. Öte yandan araştırmanın sağlayacağı yararların da ne düzeyde olabileceğı, öngörülen sonuçların ne gibi yenilikler getireceğı ve bunun maliyetinin ne olacağına da bilinmesi önemlidir. Levin ve McEvan (2000) etkili ve verimli bir değerlendirmenin yararının maliyetiyle ölçülebileceğini ifade ederek bu çözümlemenin zorunlu olduğunu vurgulamaktadır. Patel ve Russon (1999) ise program yöneticilerinin çoğunlukla değerlendirmeden çok programın uygulanmasına dönük harcamalarda bulunduğunu ve bu nedenle değerlendirme bütçesinin çok dikkatli hazırlanması gerektiğini belirtmektedir. Bunun için de değerlendirmeden beklenen yararlar ile bu yararların ortaya çıkarılması maliyetlerinin ayrıntılı bir çözümlemesinin yapılması gerektiğini vurgulamaktadır. Sonuç olarak uzmanlar tarafından uygun bulunmamasına karşın oldukça önemli olan göstergenin yeniden gözden geçirilmeye ve ifadelendirilmeye gereksinimi vardır.

Değerlendirme Araştırmasınının Yönetimi: Diğerk ülke standartları arasında yer almayan ve değerlendirme çalışmasının proje yönetim stratejileri kullanılarak yürütülmesini amaçlayan bu standart uzmanlar tarafından Türkiye için uygun bulunmuş ancak standartın son göstergesi olan “Değerlendirme çalışmasının sonuçlarının paydaşlarla paylaşılmasını ve raporlaştırılmasını sağlama” göstergesi uzmanlar tarafından uygulanabilir bulunmamıştır. Bunun temel nedeni göstergenin içeriğinin *Sonuçların Açıklanması* standartının göstergeleriyle binişik olması olabilir.

Resmi Sözleşme: Türk değerlendirme kültüründe pek uygulanmayan ancak batılı ülkeler tarafından önemsenen bu standart Türk uzmanlar tarafından da uygulanabilir bulunmuştur. Bu durum Türkiye’deki program değerlendirme çalışmaları adına önemli bir yeniliktir. Çünkü özellikle araştırmalarda resmi sözleşme geleneğinin yaygın olmadığı Türkiye’de bu standartın ve göstergelerinin bir yenilik getirmesi beklenmektedir. Türkiye’de ki bu durumun benzerini Kore’nin yaşadığını belirten Jang (2000), günlük yaşamda Korelilerin her şeyi resmi sözleşme olmaksızın yaptıklarını, batılı ülkelerden farklı olarak iş ortamında resmi sözleşmelere ve beyannamelere daha az bağlı kaldıklarını belirtmektedir. Jang, bu standartın tümüyle kültürel olduğunu

belirterek özellikle Konfüçyüs felsefesine bağlı Kore’de sözün her şeyden önemli olduğunu ve bu nedenle de resmi sözleşmelere çok itibar edilmediğini ifade etmektedir. Aynı durumun Hindistan’da da yaşanabileceğini belirten Smith ve diğ. (1994) de resmi sözleşmelerin gücünün ABD’dekinden daha sınırlı olduğunu belirtmektedir. Patel ve Russon (1999) ise Afrika’daki durumu özetlerken resmi sözleşmelerin bağlayıcılığının yasalarla kontrol altına alındığını ancak yasanın her zaman gelenekten önemli olmadığını vurgulayarak sözleşmelerin araştırmalarda pek dikkate alınmadığını belirtmektedir. Öte yandan Stufflebeam (2000) resmi sözleşmelerin değerlendirici ile değerlendirmeyi talep edenler arasında yasal bir bağ olduğunu, bu sözleşme çerçevesinde yasal sorumluluklar taşıdıklarını, sözleşme ile hem değerlendiriciyi, hem de talep edenleri güvence altına aldığını ve bu nedenle oldukça önemli olduğunu vurgulamaktadır. Sonuç olarak yukarıdaki örneklerin Türkiye kültürü için de geçerli olabileceği düşünüldüğünde standartın araştırma adına oldukça önemli olduğu görülmektedir. Nitekim son yıllarda Türkiye’de bilimsel araştırma kurumları/birimleri (TUBİTAK, EARGED vb.) ya da üniversitelerin proje birimleri tarafından desteklenen araştırma projelerinin bir resmi sözleşmeye bağlı kalınarak yürütülmesi giderek yaygınlaşmaktadır.

Adil Değerlendirme: Değerlendirilen programın güçlü ve zayıf yanlarının paydaşların etkisi altında kalmadan ortaya çıkarılmasını amaçlayan bu standart uzmanlar tarafından Türkiye için uygulanabilir bulunmuş, ancak standartın üçüncü göstergesi olan “Paydaşların beklenti ve isteklerinden etkilenmemesi” göstergesi kabul edilmemiştir. Bunun temel nedeni gösterge ile anlatılmak istenen yargının tam olarak ifade edilememesidir. Araştırma kapsamında uzmanlardan gelen eleştiriler arasında yer alan bu göstergede paydaşlardan etkilenmemesi gerekenin değerlendirici mi, değerlendirme mi olduğu açıkça belirtilmemiştir. Göstergenin kastettiği, örneğin Türkiye’de MEB, EARGED tarafından desteklenen bir program değerlendirme çalışması yapan bir araştırmacının MEB’in beklentilerinden ya da isteklerinden etkilenmemesi gereğidir. Gerçi Bewyl ve Speer (2004), değerlendirme sonuçlarının paydaşların beklentilerinden bağımsız olmasının da zor olduğunu belirtmektedir.

Sonuçların Açıklanması: Sonuçların açıklanması standartı, ABD yasalarınca garanti altına alınmış ve Türkiye’de de son yıllarda gündemde olan bilgi edinme hakkını (right

to know) ifade etmektedir (Taut, 2001). Bu nedenle tüm sonuçların açık bir biçimde tüm ilgili kesimlere iletilmesinin zorunlu olduğu vurgulanmaktadır. Çünkü bu durumun değerlendiricinin güvenilirliği ile de yakından ilişkisi vardır (Morris ve Cohn, 1993). Bu standardın Türkiye’de uygulanabilirliğine ilişkin olarak uzmanlar olumlu görüş bildirmişler, ancak standartın ikinci göstergesi olan “Ulaşılan sonuçların ve sınırlıkların dayanaklarını belirtme” göstergesinin uygun olmadığı yönünde görüş bildirmişlerdir. Genellikle araştırmaların tartışma bölümlerinin temelini oluşturan sonuçların kanıtlarını/dayanakları belirtme araştırmanın güvenilirliğini ve özgünlüğünü ortaya koyma yönünden de oldukça önemli görünmektedir. Uzmanların bu göstergenin uygulanabilir olmadığı yönünde görüş bildirmelerinin nedeni, göstergenin ifade biçimi olabilir. Uzmanlar ifadede bir anlatım bozukluğu olduğunu düşünmüş olabilirler. Göstergede yer alan “sınırlıklar” kavramının sonuçların sınırlığı olarak değil bağımsız bir sözcük olarak kullanılması bu duruma neden olmuş olabilir. Bu göstergeye benzer şekilde *Doğruluk* standartlarından *Sonuçların Gerekçelenirilmesi* standardı altında yer alan “Ulaşılan sonuçlara ilişkin kanıtlar ortaya koyma” göstergesi de uzmanlar tarafından Türkiye için uygun görülmemiştir. Uzmanlar, genel olarak araştırmada sınırlılıklar ayrı başlık altında belirtildiğinden ve sonuçlar da araştırmada elde edilen bulgulara dayalı olduğundan, araştırmacının ayrıca kanıt göstermesine gerek olmadığını da düşünmüş olabilirler. Her iki göstergenin de birbirine benzer bir içeriğe sahip olduğu, ancak buldukları bağlam içinde oldukça önemli oldukları görülmektedir. Özellikle son yıllarda araştırmalarda ulaşılan sonuçların karşılaştırmalı olarak tartışılması ve ortaya konulan sonuçların kanıtlarla savunulması araştırmanın güvenilirliği için vazgeçilmez olarak görülmektedir. Bewyl ve Speer (2004) sonuçların kanıtlarının ortaya konulmasının değerlendiricinin ayrı ve temel bir görevi olduğunu vurgulamakta, yalnızca verileri sunan, diyagramları gösteren, özetlenen bulgulardan çıkarımlarda bulunmayan çalışmaların kabul edilemez olduğunu savunmaktadır. Rieper (1997) Avrupa’da da tartışma kültürünün zayıf olduğunu belirterek birçok raporun yorumların hangi temel bilgiye dayandığını ve bilgilerin nasıl edindiğini açıklamayı ihmal ettiğini vurgulamaktadır.

Verilerin Geçerliliği ve Güvenirliği: Program değerlendirme çalışmasında kullanılacak verilerin geçerli ve güvenilir olmasını amaçlayan bu standart diğer standartlar göz

önüne alındığında alan uzmanları tarafından en düşük aritmetik ortalamaya sahip standartlardan biridir. Bunun temel nedeni uzmanların bu standartın iki göstergesini Türkiye için uygulanabilir bulmamasıdır. “Program değerlendirme araştırmasının amaçlarına uygun nitel ve nicel araştırma süreçleri belirleme” göstergesine ilişkin olarak uzmanların uygun bulmama nedeni göstergenin geçerlik ve güvenilirlik kavramlarıyla ilişkisinin kurulmaması olabilir. Uzmanlar bu göstergenin bir sonraki standartın göstergeleri ile de bir bütünlük sağladığını ve bu standart altında yer almaması düşüncesinde de bulunmuş olabilirler. Bewyl ve Speer (2004) geçerlik ve güvenirliliğin deneysel araştırmalarda oldukça önemli olduğunu, bu bağlamda iç geçerlik-dış geçerlik, yapı geçerliği-kapsam geçerliğine dikkat edilmesi gerektiğini vurgulamaktadır.

Verilerin Çözümlemesi: Nitel ve nicel verilerin çözümlenmesinin doğru olarak yapılmasını amaçlayan bu standartı uzmanlar uygulanabilir bulmuş, ancak standart altında yer alan yedi göstergenin üç göstergesinin (Veri girişlerinin uzmanlarca yapılmasını sağlama, çözümlenmeler sonucu ortaya çıkan veri tablolarını gözden geçirme ve eksiklikleri giderme, Olanaklar ölçüsünde verilerin doğrulanması ve veri tablolarının gözden geçirilmesi için farklı uzman görüşlerine başvurma) uygun olmadığı yönünde görüş bildirmişlerdir. Uygun bulunmayan bu göstergeler incelendiğinde değerlendirici dışında farklı uzmanların görevlendirilmesinin/işe koşulmasının amaçlandığı görülmektedir. Dış değerlendirici kavramıyla tanımlanabilecek bu uzmanlar gelişmiş ülkelerde sık kullanılmakla birlikte Türkiye’de genellikle programın içinden değerlendiriciler seçilmektedir. Fitzpatrick ve diğ. (2004) dış değerlendiricinin araştırmanın güvenirliliğini önemli ölçüde arttırdığını, sonuçların nesnellliğini sağlayabileceğini ifade etmektedir. Schmidt (2001) de dış değerlendiricinin daha bilimsel ve bağımsız olduğunu ifade ederek bir dış değerlendiricinin değerlendirmede görev alması gerektiğini savunmaktadır. Öte yandan kültürel özellikle taşıdığı açıkça görülen bu standart maddesinde veri girişlerinin farklı uzmanlarca yapılmasının, veri tablolarının incelenmesi için farklı bir uzmana başvurulmasının nedeni, değerlendirme çalışmalarının disiplinler arası yaklaşımla bir değerlendirme grubu tarafından yürütülmesi olarak görülebilir. Bewyl ve Speer (2004) birçok değerlendirme projesinin akademisyenleri, eğitim uzmanlarını, yönetici ve katılımcılar tarafından yürütülmesi

gerektiğini ifade ederken, bu bireylerin farklı eğitsel geçmişlerini ön plana çıkarmaktadır. Oysa Türkiye’de değerlendirme çalışmaları alan uzmanı bireyler tarafından yürütülmekte, araştırma verileri genellikle araştırmacı tarafından sisteme girilmekte ve bir istatistikçi tarafından çözümlenmektedir. Verilerin araştırmacı tarafından sisteme girilmesi kuşkusuz araştırmanın güvenilirliğini etkilemektedir. Bu nedenle araştırmanın güvenilirliği ve sonuçların nesnelliği yönünden önemli ve gerekli görülen değerlendiriciden farklı bir uzmanın görevlendirilmesi olgusunun Türkiye için yeniden tartışılması yerinde olacaktır.

5.3. Öneriler

Bu araştırma ile elde edilen sonuçlara dayalı olarak uygulamaya ve ileri araştırmalara dönük geliştirilen öneriler şunlardır:

Uygulamaya dönük öneriler:

- Bu araştırma ile ortaya çıkan program değerlendirme standartları ve göstergeleri, ulusal veya yerel düzeyde bireysel veya kurumsal katılımı yapılmış program değerlendirme çalışmalarını ya da projelerini değerlendirme çalışmalarında kullanılabilir.
- Bu araştırma ile ortaya çıkan program değerlendirme standartları ve göstergeleri, bir program değerlendirme araştırması veya projesi planlayan bir uzmana araştırmanın/projenin planlanması, yürütülmesi ve değerlendirilmesi süreçlerinde yardımcı olacak bir yönerge olarak kullanılabilir.
- Bu araştırma ile ortaya çıkan program değerlendirme standartları ve göstergeleri YÖK, MEB ve diğer eğitim kurumları tarafından program değerlendirme çalışmalarının/projelerinin değerlendirilmesinde geçerli ölçütler olarak kullanılabilir.
- Bu araştırma ile ortaya çıkan program değerlendirme standartları ve göstergeleri bugüne kadar çeşitli alanlarda program değerlendirmeye ilişkin yapılmış çalışmaların yararlık, yürütülebilirlik, uygunluk ve doğruluk yönünden niteliğinin belirlenmesinde kullanılabilir.

- Program değerlendirme arařtırmalarının bir program geliştirme bir program değerlendirme ve bir konu alan uzmanını içeren bir ekip tarafından yürütülmesi daha yararlı olabilir.
- Özellikle kurumlar tarafından maddi destek sağlanan program değerlendirme arařtırmalarında verilerin geçerliğinin ve güvenilirliğinin sağlanması adına verilerin sisteme girilmesi ve çözümlenmesi amacıyla arařtırmacının yanı sıra bir istatistik uzmanının da yer alması yaygınlaştırılabilir.

İleri arařtırmalara dönük öneriler:

- *Değerlendirici Güvenirliđi ve Yeterliđi* standardı kapsamında program değerlendirme uzmanlarının yeterliklerinin belirlenmesine dönük çalışmalar desenlenebilir ve program değerlendirme uzmanlarının yetiştirilmesine dönük lisansüstü eğitim programları açılabilir.
- Program değerlendirme çalışmalarında uyulması gereken etik ilkeler daha kapsamlı bir şekilde arařtırılabilir.
- Bu arařtırmada eğitim programları ve öğretim uzmanlarının görüşlerine dayanılarak oluşturulan program değerlendirme standartlarının ve göstergelerinin uygunluđunu ve kullanılabilirliğini sınamak için farklı alanlardan uzmanların (ölçme değerlendirme uzmanları vb.) görüşlerinin alındığı yeni çalışmalar desenlenebilir.
- Program değerlendirme standartları ve göstergelerine dayalı olarak Delfi tekniđi gibi farklı arařtırma teknikleri kullanılarak yeni standartlar ve göstergeler geliştirme çalışmalarını tasarlanabilir.
- Bu arařtırma ile ortaya çıkan program değerlendirme standartları ve göstergelerine dayalı olarak program geliştirme ve değerlendirme alanındaki gelişmeler ve deđişimler doğrultusunda bu standartları güncelleme ya da yeniden geliştirme çalışmaları desenlenebilir.

- Bu araştırma ile ortaya çıkan program değerlendirme standartları ve göstergelere dayalı olarak geliştirilen bir kontrol listesi aracılığıyla tamamlanmış program değerlendirme arařtırmalarının niteliğini belirleme arařtırması desenlenebilir.
- Bu araştırma ile ortaya çıkan program değerlendirme standartlarına ve göstergelerine dayalı olarak bir program değerlendirme ürün ölçeđi geliştirme çalışması yapılabilir.
- Bu araştırma kapsamında Türkiye için uygun bulunmayan göstergeler anlaşılrlık, uygulanabilirlik, açıklık, işlevsellik vb. yönlerden yeniden ifadelendirilerek alan uzmanlarının görüşlerine sunulabilir.

EKLER

EK 1. Türkiye Program Değerlendirme Standartları

YARARLIK STANDARTLARI (YA)

Program değerlendirmede “Yararlık Standartları”, bir program değerlendirme araştırmasının değerlendirme amaçları ve paydaşların gereksinimleri doğrultusunda yürütüldüğünün güvence altına alınmasını amaçlar.

YA.1. Paydaşların Tanımlanması

YA.1.1. Program değerlendirme araştırmasını destekleyenleri tanımlama

YA.1.2. Program değerlendirme araştırmasını planlayacak, uygulayacak ve değerlendirecek değerlendirme uzman/ uzmanlarını tanımlama

YA.1.3. Program değerlendirme araştırmasının araçlarının uygulanacağı grup ya da bireyleri tanımlama

YA.1.4. Program değerlendirme araştırmasının sonuçlarından doğrudan ya da dolaylı etkilenecek grup ya da bireyleri belirtme

YA.2. Değerlendirici Yeterliği ve Güvenirliği

YA.2.1. Program değerlendirme konusunda akademik bir süreçten geçmiş olma

YA.2.2. Program değerlendirme araştırmasının planlanma, uygulama ve değerlendirme süreçlerinde gerekli bilgi, beceri ve deneyimlere sahip olma

YA.2.3. Farklı gruplar ile çalışmak için uygun iletişim stratejilerini ve becerilerini kullanma

YA.2.4. Program değerlendirme sürecinde gerekli durumlarda dış değerlendirmeci yardımına başvurma

YA.2.5. Program değerlendirme sürecinde tüm ilgililer karşısında tarafsızlığını koruma

YA.3. Verilerin Kapsamı ve Seçimi /Yararlığı

YA.3.1. Program değerlendirme araştırmasının planlanmasında araştırma sorularının yanıtlanması için gerekli/yararlı bilgileri belirleme/seçme

YA.3.2. Toplanan/toplanacak verilerden paydaşların gereksinimlerini karşılayacak olanlarına öncelik verme

YA.4. Raporun Kapsamlılığı ve Anlaşılabilirliği

YA.4.1. Raporda program değerlendirme araştırmasının amacını, içeriğini, araştırma sorularını, süreçte kullanılan yöntemleri ve değerlendirme bulgularını açık ve kapsamlı olarak betimleme

YA.4.2. Raporu hedeflenen tüm kitlelerin özelliklerine göre özet rapor, ana rapor, teknik rapor ve sözel sunumlar halinde hazırlama

YA.4.3. Raporda sonuçların tüm ilgililer tarafından anlaşılabilirliğini sağlama için örnekler kullanma

YA.4.4. Raporda ilgililerin anlayabileceği nitelikte bir dil kullanma/ teknik dili az kullanma

YA.4.5. Raporda kullanılan önemli kavramları tanımlayan bir sözlük hazırlama

YA.4.6. Raporun dili, biçimi, zamanlaması gibi konularda paydaşların durumlarını/konularını dikkate alma

YA.4.7. Raporun gözden geçirilmesi için bağımsız değerlendirmeci görüşüne sunma

YA.5. Sonuçların Yararlılığı/Etkisi

YA.5.1. Program değerlendirme sonuçlarını nasıl, nerede ve hangi amaçla kullanılacağını belirleme

YA.5.2. Program değerlendirme sonuçlarının paydaşlar yönünden olası yararlarını/etkilerini belirleme

YA.5.3. Program değerlendirme sonuçlarının paydaşlarda yeni bakış açısı/anlayış geliştirmesi

YÜRÜTÜLEBİLİRLİK STANDARTLARI (YÜ)

Program değerlendirmede “Yürütülebilirlik Standartları”, bir program değerlendirme araştırmasının gerçekçi, mantıklı, grup ve kurumlarca destekli ve uygun maliyetli planlandığının ve yürütüldüğünün güvence altına alınmasını amaçlar.

YÜ.1. Uygun Yöntemler

- YÜ.1.1. Program değerlendirme araştırmasında seçilen yöntemin maliyet yönünden uygunluğunu inceleme
- YÜ.1.2. Program değerlendirme araştırmasında seçilen yöntemin zaman/süre yönünden uygunluğunu inceleme
- YÜ.1.3. Program değerlendirme araştırmasında seçilen yöntemin etik ilkeler yönünden uygunluğunu inceleme
- YÜ.1.4. Program değerlendirme araştırmasında seçilen yöntemin güçlü ve zayıf yönlerini ortaya koyma

YÜ.2. Grup ve Kurum Desteği

- YÜ.2.1. Program değerlendirme araştırması sürecine doğrudan ya da dolaylı katılacak kişi, kurum ve kuruluşların desteğini alma
- YÜ.2.2. Program değerlendirme araştırmasının sonuçlarını kullanacak kişi, kurum ve kuruluşların desteğini alma

YÜ.3. Maliyet-Yarar Çözümlemesi

- YÜ.3.1. Program değerlendirme araştırmasının öngörülen süresini belirleme
- YÜ.3.2. Program değerlendirme araştırmasının öngörülen maliyetini (insan kaynakları, araç-gereçler, eğitim, seyahat vb. yönünden) belirleme
- YÜ.3.3. Çalışma bütçesini ayrıntılı olarak hazırlama

YÜ.4. Değerlendirme Araştırmasının Yönetimi

- YÜ.4.1. Değerlendirilmesi amaçlanan programın değerlendirilebilirliğini ortaya koyma
- YÜ.4.2. Değerlendirme araştırmasının önerisini hazırlama
- YÜ.4.3. Zaman, süre, işlem ve kaynaklar yönünden değerlendirme çalışmasının yönetim planını hazırlama
- YÜ.4.4. Değerlendirme araştırmasının yürütme stratejilerini belirleme

UYGUNLUK STANDARTLARI (U)

Program deęerlendirmede ‘‘Uygunluk Standartları’’, bir program deęerlendirme arařtırması s¼recinde yasal, etik, yansızlık ve d¼r¼st¼luk ilkelerine uygun davranıldıęının g¼vence altına alınmasını amaçlar.

U1. Resmi S¼zleřme (program deęerlendirmenin arařtırmasının bir kurum adına ya da proje kapsamında yapıldıęı durumlarda)

- U.1.1. S¼zleřmede paydař bilgilerine yer verme
- U.1.2. S¼zleřmede program deęerlendirme arařtırmasının gerekçesini, amacını ve yöntemini belirtme
- U.1.3. S¼zleřmede program deęerlendirme arařtırmasının zaman çizelgesine yer verme
- U.1.4. S¼zleřmede tarafların (destekleyici ve kullanıcıların) saęlayacaęı hizmetleri (personel, bilgi ve araç-gereçler) belirtme
- U.1.5. Deęerlendirme s¼recinde kullanılacak hizmet (personel, danıřmanlar, seyahat, kaynaklar, posta ve telefon, verilerin ç¼z¼mlenmesi, baskı ve yayınlama) harcamalarını ayrıntılı olarak belirtme

U2. Etik Kurallara Uygunluk

- U.2.1. Deęerlendirme uygulamaları için resmi ve özel izin alma
- U.2.2. Katılımcılara deęerlendirme s¼recinde sahip oldukları temel hakları bildirme
- U.2.3. Katılımcıların kiřisel özelliklerine, k¼lt¼rel ve sosyal deęerlerine saygı g¼sterme
- U.2.4. Veri saęlayıcıların ve katılımcıların gizlilięini (kimlik vb) g¼vence altına alma
- U.2.5. Arařtırma s¼recinde eęitim hakkına özen g¼sterme (özellikle deneysel arařtırmalarda)
- U.2.6. Katılımcı bilgilerini deęerlendirme amacı dıřında kurum, grup ya da birey hizmetine sunmama g¼vencesi verme
- U.2.7. Verileri yalnızca bilimsel s¼reç ve sonuçlar için kullanma
- U.2.8. Finansal kaynakların kullanımı s¼recine iliřkin bilgi ve belgeleri paydařlarla paylařma

U3. Adil Deęerlendirme

- U.3.1. Programın g¼çlü yanlarını açık bir biçimde belirtme
- U.3.2. Programın zayıf yanlarını açık bir biçimde belirtme
- U.3.3. Uygulamanın paydař taleplerinden baęımsız y¼r¼t¼lmesi
- U.3.4. Program deęerlendirmecinin uygulama s¼recinde etik ilkelere baęlı kalması
- U.3.5. Uygulama s¼recinde deęerlendirme uzmanının politik, ideolojik ve k¼lt¼rel anlayıřını paydařlara yansıtması

U4.řeffaf Raporlařtırma

- U.4.1. Program deęerlendirme raporunun paydař taleplerinden baęımsız olması
- U.4.2. Program deęerlendirme raporunun deęerlendiricinin nesnellilięini yansıtması
- U.4.3. Program deęerlendirme raporunun tarafsız bir deęerlendirmeci tarafından incelenmesi

U5. Sonuçların Açıklanması

- U.5.1. Ç¼z¼mlenmeler sonucu ulařılan tüm bulguları eksiksiz olarak yazılı ifade etme
- U.5.2. Sonuçları alan uzmanları ile tartıřma
- U.5.3. Sonuçları paydařlarla tartıřma
- U.5.4. Sonuçları karar vericilerle tartıřma
- U.5.5. Sonuçları ilgili kesimlere ulařtırma

U6. Finansal Uygunluk

- U.6.1. Deęerlendirme gelir-gider b¼tçesini belirtme
- U.6.2. Deęerlendirme için saęlanan finansman kaynaklarını belirtme
- U.6.3. Deęerlendirme s¼reci içinde yapılan harcamaları belirtme
- U.6.4. Gelir-gider b¼tçesinin raporlarını gerektięinde deęerlendirme raporuna ekleme

DOĞRULUK STANDARTLARI (D)

Program değerlendirmede “Doğruluk Standartları”, program değerlendirme araştırmasının araştırma amaçlarıyla ilgili olarak geçerli ve kullanışlı bilgiler ve bulgular ortaya çıkardığını güvence altına almayı amaçlar.

D1. Değerlendirilecek Programın Tanımlanması

- D.1.1. Programla ilgili farklı kaynaklardan bilgi ve belgeleri toplama
- D.1.2. Programı açık, anlaşılır ve eksiksiz tanımlama
- D.1.3. Programın amaçlarını belirleme
- D.1.4. Programın var olan görevlerini belirleme

D2. Durum Çözümlemesi

- D.2.1. Kurum kayıtları, paydaş ve destekleyici görüşleri, raporlar vb. kaynaklardan yararlanarak programın durumunu betimleme
- D.2.2. Programın felsefi ve kuramsal temelleri ile yasal, sosyal ve ekonomik boyutlarını belirleme
- D.2.3. Program değerlendirmeyi etkileyebilecek olası etmenleri belirleme

D3. Değerlendirme Araştırmasının Amacının Belirlenmesi

- D.3.1. Paydaş ve destekleyici katılımıyla değerlendirme amaçlarını belirleme
- D.3.2. Paydaş ve destekleyici katılımıyla değerlendirme sorularını belirleme
- D.3.3. Program değerlendirmenin kim için ve neden yapıldığını tanımlama
- D.3.4. Belirlenen amaçların hangi gereksinimleri karşılayacağını gerekçeleriyle ifade etme

D4. Değerlendirme Araştırmasının Yönteminin Belirlenmesi

- D.4.1. Program değerlendirme araştırmasında temel alınan program değerlendirme modelini ve sınırlılıklarını belirleme
- D.4.2. Program değerlendirme araştırmasının desenini ve sınırlılıklarını belirleme
- D.4.3. Program değerlendirme araştırmasının evren ve örneklemini belirleme
- D.4.4. Program değerlendirme araştırmasının veri toplama ve çözümleme yöntem/tekniklerini belirleme

D5. Veri Kaynaklarının Tanımlanması

- D.5.1. Program değerlendirme ile ilgili geçerli ve güvenilir bilgileri toplama
- D.5.2. Program değerlendirme araştırmasında kullanılacak veri kaynaklarını (kurumlar, bireyler, tv-video, araştırmalar vb.) belirleme
- D.5.3. Veri kaynaklarının seçiminde kullanılan yöntem ve ölçütleri belirleme
- D.5.4. Veri kaynaklarını ayrıntılı inceleme (evreni, örnekleme, araştırma deseni, veri toplama aracı vb. yönlerden)

D6. Verilerin Geçerliliği ve Güvenirliği

- D.6.1. Program değerlendirme araştırmasının amaçları ile veri toplama yöntemlerinin uygunluğunu sağlama
- D.6.2. Veri toplama araçlarının geçerlik ve güvenilirlik çalışmaları için pilot uygulamalar yapma
- D.6.3. Bulguların geçerlik ve güvenirliliğini etkileyebilecek katılımcı özelliklerini ifade etme (okuma yetersizliği, dil beceri düzeyi, fiziksel engeller vb.)

D7. Verilerin Çözümlemesi

- D.7.1. Nicel verileri çözümlenmek için uygun nicel veri çözümleme tekniklerini belirleme
- D.7.2. Nitel verileri çözümlenmek için uygun nitel veri çözümleme tekniklerini belirleme
- D.7.3. Verileri doğru kodlama ve puanlama
- D.7.4. Tablo, grafikler vb. görsellerden yararlanarak çözümlemelerin sunumunu görselleştirme

D8. Sonuçların Gerekçelenmesi

- D.8.1. Program değerlendirme araştırmasının sorularının yanıtlarını açıklama
- D.8.2. Sonuçlar doğrultusunda önerilerde bulunma
- D.8.3. Bulguların alternatif yorumlamalarına yer verme
- D.8.4. Zaman, ortam, birey ve değerlendirme amaçları ve süreçleri yönünden sonuçların sınırlılıklarını ifade etme

EK 2. Alan Uzmanları İçin Veri Toplama Aracı

“TÜRKİYE İÇİN PROGRAM DEĞERLENDİRME STANDARTLARI OLUŞTURMA” ARAŞTIRMASI ANKETİ

Sayın Uzman;

Bu anket, “Türkiye İçin Program Değerlendirme Standartları Oluşturma Çalışması” konulu doktora tez çalışmasına gerekli verileri sağlamak, elde edilen sonuçlar doğrultusunda program değerlendirme çalışması yapacak araştırmacılara, bu sonuçları kullanacak ve buna dayalı karar verecek yetkililere değerlendirme hakkında sağlam kanıtlar sunacak, yine bu çalışmaların geçerliği ve güvenilirliği hakkında ilgililere yol gösterecek standartların oluşturulması amacıyla hazırlanmıştır.

Türkiye’de eğitimde program değerlendirme alanında önemli bir eksikliği gidermesi düşünülen bu çalışmada standartların belirlenmesi için siz değerli program geliştirme ve değerlendirme uzmanlarının görüşlerine gereksinim duyulmaktadır. Ankette maddeler halinde sıralanan program değerlendirme standartları olarak düşünülebilecek ifadeler, ABD’de The Joint Committee tarafından geliştirilmiş olan ve daha sonra İsviçre, Almanya, Avustralya, Kanada ve Kenya (Afrika ülkelerinde kullanılmak üzere) tarafından kendi ülkelerine uyarlanan program değerlendirme standartları temel alınarak ve Fransa, Kanada, Avustralya ve Amerika’da Değerlendirme Dernekleri ile OECD ve Birleşmiş Milletler tarafından belirlenen program değerlendirme ilkeleri, yönergeleri ve etik kurallarından da bazı maddeler eklenerek oluşturulmuştur.

Ankette yer alan her bir ifadenin program değerlendirme standardı olarak uygunluk düzeyini belirlemek için sizden istenen, standart ve gösterge ifadelerinin karşısında yer alan seçeneklerden size en uygun olanını ilgili sütuna (X) işareti koyarak belirtmenizdir. Ankete vereceğiniz yanıtlar yalnızca bu alanla ilgili bilimsel çalışmalara katkı sağlamak amacıyla kullanılacaktır.

Katkılarınız için içten teşekkürlerimizi sunarız.

Danışman:

Prof.Dr. Mustafa SAĞLAM
Anadolu Üniversitesi
Yunussemre Kampüsü
Eğitim Fakültesi
Eğitim Bilimleri Bölümü
26470 Eskişehir

Araş.Gör. İsmail YÜKSEL
Anadolu Üniversitesi
Yunussemre Kampüsü
Eğitim Fakültesi
Eğitim Bilimleri Bölümü
26470 Eskişehir

0222 335 0580 (3526)

e-mail: msaglam@anadolu.edu.tr

0222 335 0580 (3526)

e-mail: ismailyuksele@anadolu.edu.tr

EK 2 Devam

YARARLIK STANDARTLARI (YA)					
Program değerlendirmede “Yararlık Standartları”, bir program değerlendirme araştırmasının değerlendirme amaçları ve paydaşların gereksinimleri doğrultusunda yürütüldüğünün güvence altına alınmasını amaçlar.					
Aşağıda program değerlendirmede yararlık standartları olarak düşünülen ifadeler verilmiştir. Her ifadeyi inceleyerek, bu ifadenin yararlık standardı olarak uygunluğuna ilişkin görüşünüzü verilen seçeneklerden birisini (X) ile işaretleyerek belirtiniz.					
Standartlar ve Göstergeler	Uygunluk Düzeyi				
	Tamamen Uygun	Oldukça Uygun	Kısmen Uygun	Çok Az Uygun	Hiç Uygun Değil
YA1. Paydaşların Tanımlanması					
YA.1.1. Program değerlendirme araştırmasını destekleyenleri tanımlama					
YA.1.2. Program değerlendirme araştırmasını planlayacak, uygulayacak ve değerlendirecek değerlendirme uzman/uzmanlarını tanımlama					
YA.1.3. Program değerlendirme araştırmasının araçlarının uygulanacağı grup ya da bireyleri tanımlama					
YA.1.4. Program değerlendirme araştırmasının sonuçlarından doğrudan ya da dolaylı etkilenecek grup ya da bireyleri belirtme					
YA.1.5. Program değerlendirme araştırmasında sonuçların uygulanmasından sorumlu grup ya da bireyleri belirtme					
YA.1.6. Program değerlendirme araştırmasında sonuçların yayınlanmasından sorumlu grup ya da bireyleri belirtme					
YA2. Değerlendirici Yeterliliği ve Güvenirliliği					
YA.2.1. Program değerlendirme konusunda akademik bir süreçten geçmiş olma					
YA.2.2. Program değerlendirme araştırmasının planlanma, uygulama ve değerlendirme süreçlerinde gerekli bilgi, beceri ve deneyimlere sahip olma					
YA.2.3. Bireysel sınırlılıkların farkına varma, bu sınırlılıkları aşan durumlarda uzman yardımı alma					
YA.2.4. Farklı gruplar ile çalışmak için uygun iletişim stratejilerini ve becerilerini kullanma					
YA.2.5. Program değerlendirme sürecinde gerekli durumlarda dış değerlendirmeci yardımına başvurma					
YA.2.6. Program değerlendirme sürecinde tüm ilgililer karşısında tarafsızlığını koruma					
YA3. Verilerin Kapsamı ve Seçimi /Yararlığı					
YA.3.1. Program değerlendirme araştırmasının planlanmasında araştırma sorularının yanıtlanması için gerekli/yararlı bilgileri belirleme/seçme					
YA.3.2. Toplanan/toplanacak verilerden paydaşların gereksinimlerini karşılayacak olanlarına öncelik verme					
YA4. Raporun Kapsamlılığı ve Anlaşılabilirliği					
YA.4.1. Raporunda program değerlendirme araştırmasının amacını, içeriğini, araştırma sorularını, süreçte kullanılan yöntemleri ve değerlendirme bulgularını açık ve kapsamlı olarak betimleme					
YA.4.2. Raporu hedeflenen tüm kitlelerin özelliklerine göre özet rapor, ana rapor, teknik rapor ve sözel sunumlar halinde hazırlama					
YA.4.3. Raporunda sonuçların tüm ilgililer tarafından anlaşılabilirliğini sağlama için örnekler kullanma					

EK 2 Devam

YA.4.4. Raporla ilgili ilerin anlayabileceđi nitelikte bir dil kullanma/ teknik dili az kullanma					
YA.4.5. Raporla kullanılan önemli kavramları tanımlayan bir sözlük hazırlama					
YA.4.6. Raporun dili, biçimi, zamanlaması gibi konularda paydaşların durumlarını/konularını dikkate alma					
YA.4.7. Raporun gözden geçirilmesi için bağımsız değerlendirme görüşüne sunma					
YA5. Sonuçların Yararlılığı/Etkisi					
YA.5.1. Program değerlendirme sonuçlarını nasıl, nerede ve hangi amaçla kullanılacağını belirleme					
YA.5.2. Program değerlendirme sonuçlarının paydaşlar yönünden olası yararlarını/etkilerini belirleme					
YA.5.3. Program değerlendirme sonuçlarının paydaşlarda yeni bakış açısı/anlayış geliştirmesi					
Madde No	Yararlık Standartları İle İlgili Varsa Önerileriniz (Lütfen madde numarasını belirterek yazınız)				

YÜRÜTÜLEBİLİRLİK STANDARTLARI (YÜ)					
Program değerlendirme “Yürütülebilirlik Standartları”, bir program değerlendirme araştırmasının gerçekçi, mantıklı, grup ve kurumlarca destekli ve uygun maliyetli planlandığının ve yürütüldüğünün güvence altına alınmasını amaçlar.					
Aşağıda program değerlendirmede uygulanabilirlik standartları olarak düşünülen ifadeler verilmiştir. Her ifadeyi inceleyerek, bu ifadenin yürütülebilirlik standardı olarak uygunluğuna ilişkin görüşünüzü verilen seçeneklerden birisini (X) ile işaretleyerek belirtiniz.					
Standartlar ve Göstergeler	Uygunluk Düzeyi				
	Tamamen Uygun	Oldukça Uygun	Kismen Uygun	Çok Az Uygun	Hiç Uygun Değil
YÜ1. Uygun Yöntemler					
YÜ.1.1. Program değerlendirme araştırmasında seçilen yöntemin maliyet yönünden uygunluğunu inceleme					
YÜ.1.2. Program değerlendirme araştırmasında seçilen yöntemin zaman/süre yönünden uygunluğunu inceleme					
YÜ.1.3. Program değerlendirme araştırmasında seçilen yöntemin etik ilkeler yönünden uygunluğunu inceleme					
YÜ.1.4. Program değerlendirme araştırmasında seçilen yöntemin güçlü ve zayıf yönlerini ortaya koyma					
YÜ2. Grup ve Kurum Desteği					
YÜ.2.1. Program değerlendirme araştırması sürecine doğrudan ya da dolaylı katılacak kişi, kurum ve kuruluşların desteğini alma					
YÜ.2.2. Program değerlendirme araştırmasının sonuçlarını kullanacak kişi, kurum ve kuruluşların desteğini alma					
YÜ.2.3. Program değerlendirme sürecinden doğrudan ya da dolaylı etkilenebilecek kişi, kurum ve kuruluşların desteğini alma					
YÜ.2.4. Veri toplama kaynaklarının (kişi, grup, kurum ve kuruluşlar) desteğini alma					
YÜ3. Maliyet-Yarar Çözümlemesi					
YÜ.3.1. Program değerlendirme araştırmasının öngörülen süresini belirleme					
YÜ.3.2. Program değerlendirme araştırmasının öngörülen maliyetini (insan kaynakları, araç-gereçler, eğitim, seyahat vb. yönünden) belirleme					
YÜ.3.3. Program değerlendirme araştırmasının yarar-maliyet ilişkisini ortaya koyma					
YÜ.3.4. Çalışma bütçesini ayrıntılı olarak hazırlama					
YÜ4. Değerlendirme Araştırmasının Yönetimi					
YÜ4.1. Değerlendirilmesi amaçlanan programın değerlendirilebilirliğini ortaya koyma					
YÜ4.2. Değerlendirme araştırmasının önerisini hazırlama					
YÜ4.3. Zaman, süre, işlem ve kaynaklar yönünden değerlendirme çalışmasının yönetim planını hazırlama					
YÜ4.4. Değerlendirme araştırmasının yürütme stratejilerini belirleme					
YÜ4.5. Değerlendirme çalışmasının sonuçlarının paydaşlarla paylaşılmasını ve raporlaştırılmasını planlama					
Madde No	Yürütülebilirlik Standartları İle İlgili Varsa Önerileriniz (Lütfen madde numarasını belirterek yazınız)				

EK 2 Devam

UYGUNLUK STANDARTLARI					
<p>Program değerlendirme “Uygunluk Standartları”, bir program değerlendirme araştırması sürecinde yasal, etik, yansızlık ve dürüstlük ilkelerine uygun davranıldığına güvence altına alınmasını amaçlar.</p> <p>Aşağıda program değerlendirmede uygunluk standartları olarak düşünülen ifadeler verilmiştir. Her ifadeyi inceleyerek, bu ifadenin <u>uygunluk standardı</u> olarak uygunluğuna ilişkin görüşünüzü verilen seçeneklerden birisini (X) ile işaretleyerek belirtiniz.</p>					
Standartlar ve Göstergeler	Uygunluk Düzeyi				
	Tamamen Uygun	Oldukça Uygun	Kısmen Uygun	Çok Az Uygun	Hiç Uygun Değil
U1. Resmi Sözleşme (program değerlendirmenin araştırmasının bir kurum adına ya da proje kapsamında yapıldığı durumlarda)					
U.1.1. Sözleşmede paydaş bilgilerine yer verme					
U.1.2. Sözleşmede program değerlendirme araştırmasının gerekçesini, amacını ve yöntemini belirtme					
U.1.3. Sözleşmede program değerlendirme araştırmasının zaman çizelgesine yer verme					
U.1.4. Sözleşmede tarafların (destekleyici ve kullanıcıların) sağlayacağı hizmetleri (personel, bilgi ve araç-gereçler) belirtme					
U.1.5. Değerlendirme sürecinde kullanılacak hizmet (personel, danışmanlar, seyahat, kaynaklar, posta ve telefon, verilen çözümlenmesi, baskı ve yayınlama) harcamalarını ayrıntılı olarak belirtme					
U.1.6. Süreç içinde hizmetlerin ve harcamaların ayrıntılı kayıtlarını tutma					
U2. Etik Kurallara Uygunluk					
U.2.1. Değerlendirme uygulamaları için resmi ve özel izin alma					
U.2.2. Katılımcılara değerlendirme sürecinde sahip oldukları temel hakları bildirme					
U.2.3. Katılımcıların kişisel özelliklerine, kültürel ve sosyal değerlerine saygı gösterme					
U.2.4. Veri sağlayıcıların ve katılımcıların gizliliğini (kimlik vb) güvence altına alma					
U.2.5. Araştırma sürecinde eğitim hakkına özen gösterme (özellikle deneysel araştırmalarda)					
U.2.6. Katılımcı bilgilerini değerlendirme amacı dışında kurum, grup ya da birey hizmetine sunmama güvencesi verme					
U.2.7. Verileri yalnızca bilimsel süreç ve sonuçlar için kullanma					
U.2.8. Finansal kaynakların kullanımı sürecine ilişkin bilgi ve belgeleri paydaşlarla paylaşma					
U3. Adil Değerlendirme					
U.3.1. Programın güçlü yanlarını açık bir biçimde belirtme					
U.3.2. Programın zayıf yanlarını açık bir biçimde belirtme					
U.3.3. Paydaşların beklenti ve isteklerinden etkilenmemesi					
U.3.4. Uygulamanın paydaş taleplerinden bağımsız yürütülmesi					
U.3.5. Program değerlendirmecinin uygulama sürecinde etik ilkelere bağlı kalması					
U.3.6. Uygulama sürecinde değerlendirme uzmanının politik, ideolojik ve kültürel anlayışını paydaşlara yansıtması					
U4.Şeffaf Raporlaştırma					

U.4.1.	Program değerlendirme raporunun paydaş taleplerinden bağımsız olması					
U.4.2.	Program değerlendirme raporunun değerlendircinin nesnellliğini yansıtması					
U.4.3.	Program değerlendirme raporunun tarafsız bir değerlendirmeci tarafından incelenmesi					
U5. Sonuçların Açıklanması						
U.5.1.	Çözümlenmeler sonucu ulaşılan tüm bulguları eksiksiz olarak yazılı ifade etme					
U.5.2.	Ulaşılan sonuçların ve sınırlılıkların dayanaklarını belirtme					
U.5.3.	Sonuçları alan uzmanları ile tartışma					
U.5.4.	Sonuçları paydaşlarla tartışma					
U.5.5.	Sonuçları karar vericilerle tartışma					
U.5.6.	Sonuçları ilgili kesimlere ulaştırma					
U6. Finansal Uygunluk						
U.6.1.	Değerlendirme gelir-gider bütçesini belirtme					
U.6.2.	Değerlendirme için sağlanan finansman kaynaklarını belirtme					
U.6.3.	Değerlendirme süreci içinde yapılan harcamaları belirtme					
U.6.4.	Gelir-gider bütçesinin raporlarını gerektiğinde değerlendirme raporuna ekleme					
Madde No	Uygunluk Standartları İle İlgili Varsa Önerileriniz (Lütfen madde numarasını belirterek yazınız)					

EK 2 Devam

DOĞRULUK STANDARTLARI					
<p>Program değerlendirilmede “Doğruluk Standartları”, program değerlendirme araştırmasının araştırma amaçlarıyla ilgili olarak geçerli ve kullanışlı bilgiler ve bulgular ortaya çıkardığını güvence altına almayı amaçlar.</p> <p>Aşağıda program değerlendirilmede doğruluk standartları olarak düşünülen ifadeler verilmiştir. Her ifadeyi inceleyerek, bu ifadenin <i>doğruluk standardı</i> olarak uygunluğuna ilişkin görüşünüzü verilen seçeneklerden birisini (X) ile işaretleyerek belirtiniz.</p>					
	Uygunluk Düzeyi				
	Tamamen Uygun	Oldukça Uygun	Kısmen Uygun	Çok Az Uygun	Hiç Uygun Değil
Standartlar ve Göstergeler					
D1. Değerlendirilecek Programın Tanımlanması					
<i>D.1.1. Programla ilgili farklı kaynaklardan bilgi ve belgeleri toplama</i>					
<i>D.1.2. Programı açık, anlaşılır ve eksiksiz tanımlama</i>					
<i>D.1.3. Programın amaçlarını belirleme</i>					
<i>D.1.4. Programın var olan görevlerini belirleme</i>					
D2. Durum Çözümlemesi					
<i>D.2.1. Kurum kayıtları, paydaş ve destekleyici görüşleri, raporlar vb. kaynaklardan yararlanarak programın durumunu betimleme</i>					
<i>D.2.2. Programın felsefi ve kuramsal temelleri ile yasal, sosyal ve ekonomik boyutlarını belirleme</i>					
<i>D.2.3. Program değerlendirmeyi etkileyebilecek olası etmenleri belirleme</i>					
D3. Değerlendirme Araştırmasının Amacının Belirlenmesi					
<i>D.3.1. Paydaş ve destekleyici katılımıyla değerlendirme amaçlarını belirleme</i>					
<i>D.3.2. Paydaş ve destekleyici katılımıyla değerlendirme sorularını belirleme</i>					
<i>D.3.3. Program değerlendirmenin kim için ve neden yapıldığını tanımlama</i>					
<i>D.3.4. Belirlenen amaçların hangi gereksinimleri karşılayacağını gerçekleriyle ifade etme</i>					
D4. Değerlendirme Araştırmasının Yönteminin Belirlenmesi					
<i>D.4.1. Program değerlendirme araştırmasında temel alınan program değerlendirme modelini ve sınırlılıklarını belirleme</i>					
<i>D.4.2. Program değerlendirme araştırmasının desenini ve sınırlılıklarını belirleme</i>					
<i>D.4.3. Program değerlendirme araştırmasının evren ve örneklemini belirleme</i>					
<i>D.4.4. Program değerlendirme araştırmasının veri toplama ve çözümleme yöntem/tekniklerini belirleme</i>					
D5. Veri Kaynaklarının Tanımlanması					
<i>D.5.1. Program değerlendirme ile ilgili geçerli ve güvenilir bilgileri toplama</i>					
<i>D.5.2. Program değerlendirme araştırmasında kullanılacak veri kaynaklarını (kurumlar, bireyler, tv-video, araştırmalar vb.) belirleme</i>					
<i>D.5.3. Veri kaynaklarının seçiminde kullanılan yöntem ve ölçütleri belirleme</i>					
<i>D.5.4. Veri kaynaklarını ayrıntılı inceleme (evreni, örnekleme, araştırma deseni, veri toplama aracı vb. yönlerden)</i>					
D6. Verilerin Geçerliliği ve Güvenirliği					
<i>D.6.1. Program değerlendirme araştırmasının amaçlarına uygun nitel ve nicel araştırma süreçleri belirleme</i>					

EK 2 Devam

D.6.2.	Program değerlendirme araştırmasının amaçları ile veri toplama yöntemlerinin uygunluğunu sağlama						
D.6.3.	Veri toplama araçlarının geçerlik ve güvenilirlik çalışmaları için pilot uygulamalar yapma						
D.6.4.	Veri toplama araçları ile toplanacak verilerin kapsamını tanımlama						
D.6.5.	Bulguların geçerlik ve güvenilirliğini etkileyebilecek katılımcı özelliklerini ifade etme (okuma yetersizliği, dil beceri düzeyi, fiziksel engeller vb.)						
D7. Verilerin Çözümlemesi							
D.7.1.	Nicel verileri çözümlmek için uygun nicel veri çözümleme tekniklerini belirleme						
D.7.2.	Nitel verileri çözümlmek için uygun nitel veri çözümleme tekniklerini belirleme						
D.7.3.	Veri girişlerinin uzmanlarca yapılmasını sağlama						
D.7.4.	Verileri doğru kodlama ve puanlama						
D.7.5.	Tablo, grafikler vb. görsellerden yararlanarak çözümlmelerin sunumunu görselleştirme						
D.7.6.	Çözümlmeler sonucu ortaya çıkan veri tablolarını gözden geçirme ve eksiklikleri giderme						
D.7.7.	Olanaklar ölçüsünde verilerin doğrulanması ve veri tablolarının gözden geçirilmesi için farklı uzman görüşlerine başvurma						
D8. Sonuçların Gerekçelenirilmesi							
D.8.1.	Program değerlendirme araştırmasının sorularının yanıtlarını açıklama						
D.8.2.	Ulaşılan sonuçlara ilişkin kanıtlar ortaya koyma						
D.8.3.	Sonuçlar doğrultusunda önerilerde bulunma						
D.8.4.	Bulguların alternatif yorumlamalarına yer verme						
D.8.5.	Zaman, ortam, birey ve değerlendirme amaçları ve süreçleri yönünden sonuçların sınırlılıklarını ifade etme						
Madde No	Doğruluk Standartları İle İlgili Varsa Önerileriniz (Lütfen madde numarasını belirterek yazınız)						

EK 3. Uzmanlara Gönderilen E-davet Mektubu

Değerli Meslektaşımız,

Günümüzde, bilim teknoloji ve ekonomi alanındaki gelişmeler her ülkede bireysel ve toplumsal yaşamı derinden etkilemekte ve köklü değişimlere neden olmaktadır. Hemen her ülke bu değişimlere uyum sağlayabilmek için eğitim başta olmak üzere tüm toplumsal sistemlerde etkililik, verimlilik ve kalitenin artırılması için çaba göstermektedir. Bu çabaların önemli bir kesimini de sistemlerin girdi, süreç ve ürün boyutlarına ilişkin standartların geliştirilmesi oluşturmaktadır.

Eğitim sistemlerinde niteliğin ve verimliliğin artırılması, örgün ve yaygın eğitim kurumlarındaki eğitim programlarının toplumsal gereksinimler ve koşullar doğrultusunda geliştirilmesi ve bu programların uygulanması sonucunda elde edilecek çıktılarının standartlarının belirlenmesiyle sağlanabilir. Bu nedenle özellikle gelişmiş pek çok ülkede eğitim kurumlarının hizmet ve ürünlerine dönük standartların geliştirilmesi ve bu standartlar doğrultusunda kurumların akredite edilmesi çalışmalarına önem verilmektedir. Türkiye’de de bu yönde bazı çalışmaların olduğu bilinmektedir. Ancak bu girişimlerin yetersiz olduğu da bir gerçektir.

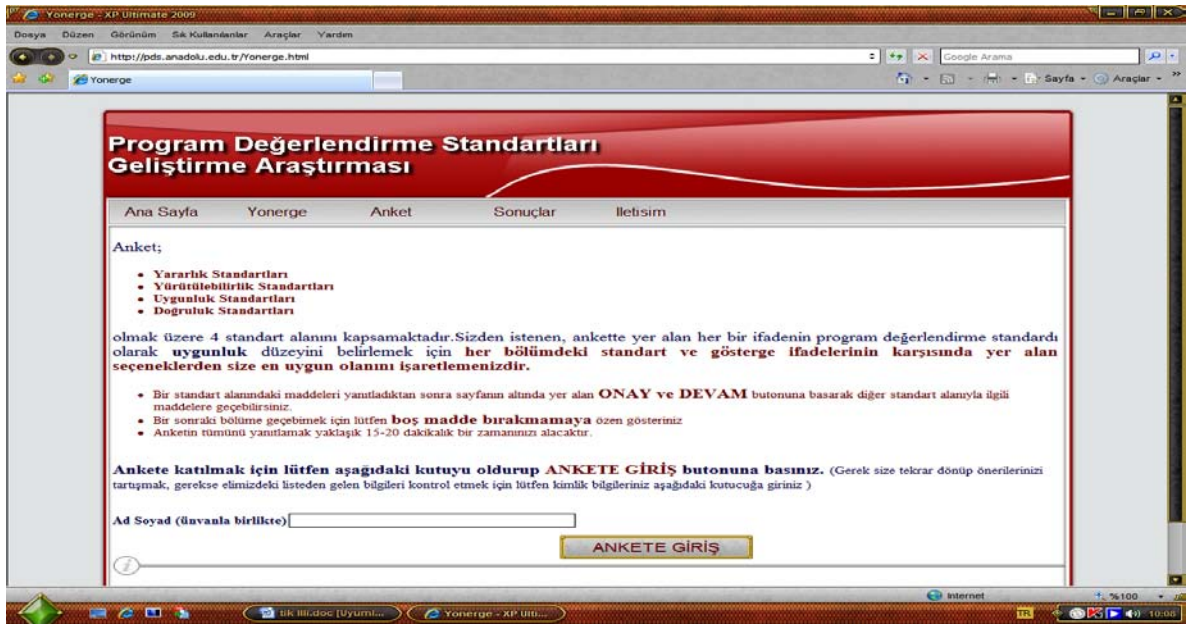
Eğitim programların geliştirilmesindeki temel dayanaklardan biri de program değerlendirme çalışmalarıdır. Bu nedenle ülke, kurum ve ders düzeyinde yapılan program değerlendirme çalışmalarının daha nitelikli ve verimli yürütülebilmesi, yapılan program değerlendirme çalışmalarının niteliğinin ortaya konulabilmesi için program değerlendirme standartlarına gereksinim duyulmaktadır. Bu gereksinimi karşılamak amacıyla **“Türkiye İçin Program Değerlendirme Standartları Oluşturma Çalışması”** konulu bir doktora tez çalışması yapılacaktır. Bu çalışmanın amacına uygun olarak gerçekleştirilebilmesi için gerekli olan verilerin toplanmasında Eğitim Programları ve Öğretim alanında çalışan bir akademisyen ve program geliştirme ve değerlendirme uzmanı olarak sizlerin görüşlerine ve katkılarına gereksinim duyulmaktadır. Bu alanda Türkiye’deki uzmanların sayısının azlığı ve tez konusunun önemi nedeniyle görüş ve katkılarınız bizim için son derece önemlidir.

Alanımızdaki bir eksikliği gidereceğini düşündüğümüz bu araştırma için hazırlanan aşağıda web adresi verilen anketi biraz zaman ayırarak ve gereken hassasiyeti göstererek yanıtlamanızı rica eder, çalışmalarınızda başarılar dileriz.

Ankete <http://pds.anadolu.edu.tr> adresine tıklayarak ulaşabilirsiniz.

Prof.Dr. Mustafa SAĞLAM,
Araş. Gör. İsmail YÜKSEL
Anadolu Üniversitesi Eğitim Fakültesi
Eğitim Bilimleri Bölümü

EK 4. Program Değerlendirme Standartları Oluşturma Çalışması Anketinin Web Sayfası Görüntüleri



EK 4 Devam

Yararlılık Standartları - XP Ultimate 2009

Yararlılık Standartları

Program Değerlendirme Standartları Geliştirme Araştırması

Ana Sayfa Yonerge Anket Sonuçlar İletisim

Yararlılık Standartları
Program değerlendirilmede "**Yararlılık Standartları**", bir program değerlendirme araştırmasının değerlendirme amaçları ve paydaşların gereksinimleri doğrultusunda yürütüldüğünün güvence altına alınmasını amaçlar

Aşağıda program değerlendirmede yararlılık standartları olarak düşünülen ifadeler verilmiştir. Her ifadenin **yararlılık standardı** olarak uygunluğuna ilişkin görüşünüzü verilen seçeneklerden birisini işaretleyerek belirtiniz.

STANDARTLAR ve GÖSTERGELER	Tamamen Uygun	Oldukça Uygun	Kısmen Uygun	Çok Az Uygun	Hiç Uygun Değil
YA1. Paydaşların Tanımlanması					
1.1 Program değerlendirme araştırmasını destekleyenleri tanımlama	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1.2 Program değerlendirme araştırmasını planlayacak, uygulayacak ve değerlendirecek değerlendirme uzman/uzmanlarını tanımlama	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1.3 Program değerlendirme araştırmasının araçlarının uygulanacağı grup ya da bireyleri tanımlama	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1.4 Program değerlendirme araştırmasını sonuçlarından doğrudan ya da dolaylı etkilenecek grup ya da bireyleri belirtme	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1.5 Program değerlendirme araştırmasında sonuçların uygulanmasından sorumlu grup ya da bireyleri belirtme	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bitir

Internet %100

10:08

İletisim - XP Ultimate 2009

İletisim

Program Değerlendirme Standartları Geliştirme Araştırması

Ana Sayfa Yonerge Anket Sonuçlar İletisim

Anketin herhangi bir bölümüyle ya da genelleyle ilgili önerilerinizi ve eleştirilerinizi pds@anadolu.edu.tr mail adresine ya da bu pano yardımıyla bize gönderebilirsiniz.

Adınız :

e-Mail :

Konu :

Mesaj :

Internet %100

10:09

KAYNAKÇA

- Aktan, C. ve Gencel, U. (2007). *Yüksek öğretimde akreditasyon*. İzmir: Yaşar Üniversitesi Yayını.
- Alkin, M. C. (2004). *Evaluation roots: tracing theorists' views and influences*. Thousand Oaks, CA: SAGE.
- Alkin, M. C. ve Christina, C. (2004). An evaluation theory tree. A. Marvin (Ed.). *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage.
- American Evaluation Association (AEA). (2005). Guiding principles for evaluators. *American Journal of Evaluation* , 26 (1), 5-7.
- Arslan, M. (2000). Cumhuriyet dönemi ilköğretim programları ve belli başlı özellikleri. *Milli Eğitim Dergisi*, 146.
- Australasian Evaluation Society. (2009). *Guidelines for the Ethical Conduct of Evaluations*. The Australasian Evaluation Society (AES): <http://www.aes.asn.au/> adresinden 24 Ocak 2009 tarihinde edinilmiştir.
- Berry, J., Segall, M.ve Kagitçibasi, C. (. (1997). *Handbook of cross-cultural psychology*. Needham Heights, NC: Allyn & Bacon.
- Beywl, W. (2000). Standards for evaluation: On the way to guiding principles in German evaluation. C. Russon (edt), *The Program Evaluation Standards in International Settings* (s. 60-67). Kalamazoo: Evaluation Center, Western Michigan University.
- Beywl, W.ve Speer, S. (2004). Developing standards to evaluate vocational education and training programs. P. D. Tessaring (edt), *The Foundations of Evaluation and Impact Research: Third Report on Vocational Training Research in Europe—Background Report* (s. 51-131). Luxembourg: Office for Official Publications of the European Communities.

- Bilen, M. (1999). *Plandan uygulamaya öğretim*. Ankara: Anı Yayıncılık.
- Birsözlük. (2009). *Türkçe sözlük*. <http://www.birsozluk.com/> adresinden 10 Ocak 2009 tarihinde edinilmiştir.
- Bloom, B. S. (1968). *Learning for mastery. Evaluation comment*. Los Angeles: University of California.
- Bloom, B. S., Hastings, J. T. ve Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw Hill Book Company.
- Brinkerhoff, R. O., Brethower, D. M., Hluchyj, T. ve Nowakowski, J. R. (1983). *Program evaluation: a practitioner's guide for trainers and educators*. Boston: Kluwer/Nijhoff.
- Burke, B. (1998). Evaluating for a change: reflections on participatory methodology. *New Directions for Evaluation* (80), 43-56.
- CERI. (2005). *Formative assessment: improving learning in secondary classrooms*. Paris: OECD Publishing.
- Chatterji, M. (2005). Applying the Joint Committee's 1994 standards in international contexts: a case study of education evaluations in Bangladesh. *Teachers College Record*, 107 (10), 2372-2400.
- Chelmsky, E. ve Shadish, W. R. (1997). *Evaluation for the 21st century: a handbook*. Thousand Oaks, CA: Sage.
- Cheng, M. (2003). *Medical device regulations: Global overview and guiding principles*. World Health Organization.
- Conley, D. (1997). Performance standards. *OSSC Bulletin*, 40 (3), s. 8-33.
- Cousins, J. ve Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation* (80), 5-23.

- CSA. (2009). *Standards development*. <http://www.csa.ca/cm/ca/en/standards-development> adresinden 15 Haziran 2009 tarihinde edinilmiştir.
- Çelenk, S., Tertemiz, N. ve Kalaycı, N. (2000). *İlköğretim programları ve gelişmeler*. Ankara: Nobel Yayınevi.
- Demirel, Ö. (2006). *Kuramdan uygulamaya eğitimde program geliştirme (10.Baskı)*. Ankara: PegemA Yayıncılık.
- Demirel, Ö. (1992). Türkiye'de program geliştirme uygulamaları. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* (7), s. 27-43.
- Edelfelt, R. ve Raths, J. (1998). *A brief history of standards in teacher education*. Virginia: Association of Teacher Educators.
- Eisner, E. W. (1976). Educational connoisseurship and criticism: Their form and functions in educational evaluation. *Journal of Aesthetic Education*, 10 (3/4), 135-150.
- Eisner, E. W. (1985). *The art of educational evaluation: A personal view*. Philadelphia: The Falmer Press, Taylor & Francis Inc.
- Erden, M. (1998). *Eğitimde program değerlendirme*. Ankara: Anı Yayıncılık.
- Ertürk, S. (1993). *Eğitimde program geliştirme*. Ankara: Meteksan Yayınları.
- Fer, S. (2005). Cumhuriyet dönemi ilköğretim programı üzerine bir değerlendirme. *Cumhuriyet Dönemi Eğitim Politikaları Sempozyumu*. İstanbul: Marmara Üniversitesi Atatürk Eğitim Fakültesi ve Başbakanlık Atatürk Kültür, Dil Ve Tarih Yüksek Kurumu Atatürk Araştırma Merkezi.
- Fidan, N. (1997). *Okulda öğrenme ve öğretme*. Ankara: Alkım Yayınevi.
- Fink, A. (1995). *How to design surveys*. Thousand Oaks, CA: Sage.

- Fitzpatrick, J., Sanders, J. R. ve Worthen, B. (2004). *Program evaluation: alternative approaches and practical guidelines*. Boston: Pearson Education.
- Fontaine, C. ve Monnier, E. (2002). *Evaluation in France*, J.-E. Furubo, R. C. Rist ve R. Sandahl (edt) *International Atlas of Evaluation*, ss: 63–75. New Brunswick, NJ and London: Transaction Publishers.
- Fournier, D. M. (1994). Review of the joint committee on standards for educational evaluation. *Journal of Educational Measurement*, 31 (4), 363-367.
- Gencil, U. (2001). Yükseköğretim hizmetlerinde toplam kalite yönetimi ve akreditasyon. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 3 (3), 164-218.
- Giraffe, V. (1997). *A Metaevaluation of the assessment standards applied to government funded adult vocational training programs*. Yayınlanmamış Doktora Tezi. Saybrook Institute San Francisco.
- Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement*, (15), 237-261.
- Gould, R. B. (1995). *The development, validation, and applicability of "the program evaluation standards: how to assess evaluations of educational programs*. Validation Panel Commissioned by the Joint Committee on Standards for Educational Evaluation.
- Gözütok, D. (2001). Program değerlendirme. M. Gültekin (Ed.). *Öğretimde planlama ve değerlendirme*. Eskişehir: Anadolu Üniversitesi Açıköğretim Fakültesi Yayınları.
- Gözütok, D. (2003). Türkiye’de program geliştirme çalışmaları. *Milli Eğitim Dergisi* (160), 44-64.
- Greene, J. (2002). With a splash of soda, please: towards active engagement with difference. *Evaluation* (8), 259–266.

- Guba, E. G. ve Lincoln, Y. S. (1981). *Effective evaluation: improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass.
- Gürkan, T. (2005). Programın yapısal boyutları ve program geliştirme süreci. M. Gültekin (Ed.). *Öğretimde planlama ve değerlendirme* (s. 15-30). Eskişehir: Anadolu Üniversitesi Yayınları.
- IDS. (1998). Participatory monitoring and evaluation: learning from change. *IDS Policy Briefings* (12).
- Jang, S. (2000). The Appropriateness of Joint Committee standards in non-western settings: A Case study of South Korea. C. Russon (Ed.). *The program evaluation standards in international settings*. Kalamazoo: Western Michigan University.
- Johnson, J. J. (1983). *Evaluation of the reading program of the Guilford County*. Yayınlanmamış Yüksek Lisans Tezi. The University of North Carolina at Greensboro.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards. (2nd ed.)*. California: Thousand Oaks, Sage.
- Kabukçuoğlu, S. (2005). *Herkes için fizibilite*. (Ders Notları). Ankara
- Karasar, N. (2008). *Bilimsel araştırma yöntemi: Kavramlar, ilkeler, teknikler*. Ankara: Nobel Yayın Dağıtım.
- Kaya, Z. (1997). Eğitimde program değerlendirme sürecinin temel işlemleri. *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, s. 5.
- Keith, G. (2003). The Canadian evaluation society (CES) experience in developing standards for evaluation & ethical issues. *The 5th European Conference on the Evaluation of Structural Funds*, (s. 1-25). Budapest, Hungary.

- Kellaghan, T., Stufflebeam, D. L. ve Wingate, L. A. (2003). *International handbook of educational evaluation*. Dordrechts: Springer.
- King, J. A. (2003). Evaluating educational programs and projects in the United States. T. Kellaghan, D. L. Stufflebeam ve L. A. Wingate (Ed.). *International handbook on educational evaluation*. Dordrechts: Springer.
- Koetting, J. R. (1988). Educational connoisseurship and educational criticism: Pushing beyond information and effectiveness. *Fifth Annual Open Forum The Foundational Issues Of The Field* (s. 442-457). Orleans, LA: Annual Meeting of the Association for Educational Communications and Technology.
- Lachat, M. A. (1994). *High standards for all students: Opportunities and challenges*. New Hampshire: Center For Research Management.
- Larsen, N. B. (1985). *Implementation and metaevaluation of an experimental method for evaluating an administrator training program*. Yayınlanmamış Yüksek Lisans Tezi. Western Michigan University.
- Lee, B. L. (1995). *A meta-evaluation of Taiwan ministry of education's national technology institutes' evaluation: A study of evaluation team's and stakeholders' judgments on the evaluation practice*. Yayınlanmamış Doktora Tezi. Virginia Polytechnic Institute and State University.
- Levin, H.ve McEwan, P. (2001). *Cost-effectiveness analysis (2. baskı)*. Thousand Oaks, CA: Sage Publications.
- Lewy, A. (1984). Evaluation Standards: Comments from Israel. *Evaluation and Program Planning* , 7, 377-381.
- Maher, J. A. (1997). *Planning and implementing an evaluation of a program for young children with developmental disabilities in a public school system: A case investigation*. Yayınlanmamış Doktora Tezi. The State University of New Jersey.

- Mann, K. B. (1991). *Evaluation standards for selected student activity programs in Nebraska high schools*. Yayınlanmamış Yüksek Lisans Tezi. The University of Nebraska - Lincoln.
- Martin, P. H. (1981). *Evaluation in bilingual education programs, a metaevaluation of title VII projects: Fiscal year 1979-80*. Yayınlanmamış Yüksek Lisans Tezi. University of San Francisco.
- Mathison, S. (2005). *Encyclopaedia of evaluation*. Londra: SAGE Publications.
- McCain, D. V. (2005). *Evaluation basics*. Alexandria: American Society for Training & Development.
- McLaughlin, J. A. (1990). *A guide for planning and conducting an evaluation of local special education programs in Kansas*. Blacksburg, VA: Virginia Polytechnic Institute & State University
- McNeil, J. D. (1977). *Curriculum: A comprehensive introduction. Fifth edition*. Los Angeles: Harper Collins College Publishers.
- MEB. (2005). *Temel eğitim destek programı öğretim programlarının değerlendirme raporu*. Ankara: MEB-EARGED.
- Mertens, D. M. (2004). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Thousands Oaks, California: SAGE Publications.
- Miles, M. B. ve Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook (2. baskı.)*. Thousand Oaks, CA: Sage.
- Mitchell, V. J. (1998). *A metaevaluation of the Texas education agency's district effectiveness and compliance review a record of study*. Yayınlanmamış Doktora Tezi. Texas A&M University.

- Morris, M.ve Cohn, R. (1993). Program evaluators and ethical challenges. *Evaluation Review* , 17 (6), 621-642.
- Morrison, G. S. (1993). *Contemporary curriculum K-8*. Boston, Mass.: Allyn and Bacon.
- NISO. (2009). *Creating NISO standards*. <http://www.niso.org/creating/index.html> adresinden 15 Mayıs 2009 tarihinde edinilmiştir.
- Oğuzkan, F. (1993). *Eğitim terimleri sözlüğü*. Ankara: Emel Matbaacılık.
- Oliva, P. (1988). *Developing the curriculum. Second edition*. United States: Scott, Foresman / Little, Brown College Division.
- Ötüken sözlük. (2009). *Ötüken sözlük*. <http://www.otukensozluk.com/> adresinden 10 Ocak 2009 tarihinde edinilmiştir.
- Patel, M.ve Russon, C. (1999). Appropriateness of the program evaluation standards for use in African cultures. *African Evaluation Association*. Nairobi: Kenya.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. California: SAGE.
- Patton, M. Q. (1994). The program evaluation standards: How to assess evaluation of educational programs. *Evaluation Practice*, 17 (2), 193-199.
- Pearson. (2008). What is a formative assessment. *Educational Measurement Research Bulletin* (5).
- Plake, B. S. (2008). Standard setters: Stand up and take a stand! *Educational Measurement: Issues and Practice*, 27 (1), 3-9.
- Popham, W. J. (2006). Defining and enhancing formative assessment. *Assessment for Learning*.
- Provus, M. M. (1969). *The discrepancy evaluation model: An approach to local program improvement*. PA: Pittsburgh Public Schools.

Quinion, M. (1996). *World wide words*.

<http://www.worldwidewords.org/topicalwords/tw-curl.htm> adresinden 10 Mayıs 2009 tarihinde edinilmiştir.

Richardson, V. (1994). Standards and assessments: What is their educative potential? M. Diez, V. Richardson ve P. Pierson (Ed.). *Setting Standards and Educating Teachers*. Washington DC: American Association of Colleges for Teacher Education.

Rieper, O. (1997). Evaluation practice of European Structural Funds. *Evaluation of European training, employment and human resource programmes*. (s. 37-43). Luxembourg: Office for Official Publications of the European Communities.

Riles, P. (2002). *Ralph W. Tyler*. <http://wredu.com/~wriles/Tyler.html> adresinden 12 Mayıs 2009 tarihinde edinilmiştir.

Rossi, P. H. (1982). Standards for evaluation practice. *New Directions for Program Evaluation* (15).

Rossi, P. H., Lipsey, M. ve Freeman, H. (2004). *Evaluation: A systematic approach*. California: Sage Publications.

Rossmann, G. B. (2008). *Participatory monitoring & evaluation*. http://www.umass.edu/cie/Themes/participatory_evaluation.htm adresinden 20 Aralık 2008 tarihinde adresinden edinilmiştir

Rouge, J. C. (2004). The origin and development of the African evaluation guidelines. *New Directions for Evaluation* (104), 55-66.

Sağlam, M. ve Yüksel, İ. (2007). Program değerlendirmede meta analiz ve meta değerlendirme. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi* (18), 175-187.

Sanders, J. R. ve Nafziger, D. N. (1976). *A basis for determining the adequacy of evaluation design*. *Occasional paper No.6*. Kalamazoo: Western Michigan University Evaluation Center.

- Saylor, G. J. ve Alexander, W. M. (1974). *Planning curriculum for schools*. New York: Halt, Rinehart and Winston Inc.
- Schmidt, C. (2001). Knowing what works: the case for rigorous program evaluation. London : CEPR (CEPR Working paper).
- Scriven, M. (1991b). Beyond summative and formative evaluation. M. W. McLaughlin ve D. Phillips (Ed.). *Evaluation and Education: At Quarter Century* (s. 19-64). Chicago: University of Chicago Press.
- Scriven, M. (1991a). *Evaluation thesaurus*. London: SAGE.
- Scriven, M. (1967). The methodology of evaluation. R. E. Stake (Ed.). *Curriculum Evaluation*. Chicago: Rand McNally.
- Scriven, M. (1972). The methodology of evaluation. C. Weiss (Ed.). *Evaluating Action Programs: Readings in Social Action and Education*. Boston: Allyn & Bacon.
- Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice*, 17, 151-162.
- Shadish, W. R., Cook, T. D. ve Leviton, L. C. (1995). *Foundations of program evaluation: theories of practice* (3. baskı). Londra: SAGE.
- Smith, N., Chircop, S.ve Mukherjee, P. (1993). Considerations on the development of culturally relevant evaluation standards. *Studies in Educational Evaluation* , 19, 3-13.
- Sönmez, V. (2001). *Program geliřtirmede öğretmen el kitabı*. Ankara: Anı Yayıncılık.
- Sözer, E. (1998). *Kuramdan uygulamaya sosyal bilimlerin öğretimi*. Eskişehir: Anadolu Üniversitesi Yayınları.
- Sözlüknet. (2009). *Sözlüknet*. <http://www.sozluk.net/> adresinden 10 Ocak 2009 tarihinde edinilmiştir

- Stake, R. E. (1975). *Program evaluation particularly responsive evaluation (Occasional Paper No: 5)*. Kalamazoo: Western Michigan University Evaluation Center.
- Stake, R. E. (1972). *Responsive evaluation*. US Department of Health, Education and Welfare.
- Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record*, 68 (7), 523-540.
- Straton, R. G. (2001). Evaluation research in Australasia: Moving forward. *Evaluation Journal of Australasia*, 1 (2), 24-26.
- Struhkamp, G. (2005). Evaluation in Germany: An overview. *Journal of MultiDisciplinary Evaluation*. 3, 180-195
- Stufflebeam, D. L. (1974). *Metaevaluation (Occasional Paper No: 3)*. Kalamazoo: Western Michigan Evaluation Center.
- Stufflebeam, D. L. (1992). Professional standards for educational evaluation. T. Husén ve P. Neville (Ed.). *International Encyclopaedia of Education*. Pergamon Press.
- Stufflebeam, D. L. (2000). Lessons in contracting for evaluation. *American Journal of Evaluation* , 21 (3), 293-314
- Stufflebeam, D. L. (2003). The CIPP model for evaluation. *Annual Conference of the Oregon Program Evaluators Network*. Portland, Oregon: Oregon Program Evaluators Network.
- Stufflebeam, D. L. ve Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco, CA: Jossey- Bass A Wiley Imprint.
- Stufflebeam, D. L., Madaus, G. F. ve Kellaghan, T. (2002). *Evaluation models: Viewpoints on educational and human services evaluation*. Dordrechts: Springer.
- Sweeny, B. (1999). Content standards: Gate or bridge? *Kappa Delta Pi Record*, 35 (2), s. 64-67.

- Talmage, H. (1982). Evaluation of programs. H. E. Mitzel (Ed.). *Encyclopaedia of Educational Research*. New York: Free Press.
- Taut, S. (2000). Cross-cultural transferability of the program evaluation standards. C. Russon (edt), *The program evaluation standards in international settings* (s. 5-27). Kalamazoo, MI: The Evaluation Center (Occasional Papers Series).
- TDK. (2009). *Türk Dil Kurumu*. <http://tdkterim.gov.tr/bts/> adresinden 10 Ocak 2009 tarihinde edinilmiştir
- Torres, E. B. (2005). *A metaevaluation of an evaluation of a second language course*. Yayınlanmamış Yüksek Lisans Tezi. Brigham Young University.
- Turgut, M. F. (1983). Program değerlendirme. *Cumhuriyet Döneminde Eğitim*. İstanbul: Milli Eğitim Basımevi.
- Tyler, R. (1969). *Basic principles of curriculum and instruction*. Chicago: University Of Chicago Press.
- Tyler, R. W. (1981, Kasım). *An interview with Ralph Tyler*. (J. R. Nowakowski, Röportajı Yapan)
- Variş, F. (1996). *Eğitimde program geliştirme: Teoriler-teknikler*. Ankara: Alkım Yayıncılık.
- W.K. Kellogg Foundation. (1998). *Evaluation Handbook*. W.K. Kellogg Foundation.
- Wahba, A. F. (1990). *Assessing meta-evaluation in practice: Application of the joint committee standards to final evaluations of traditional birth attendants (tbas) in four developing countries*. Yayınlanmamış Doktora Tezi. University of Pittsburgh.
- Weiss, C. H. (1972). *Evaluation research: Methods of assessing program effectiveness*. Upper Saddle River, N.J: Prentice Hall.

- Widmer, T. (2004). The Development and status of evaluation standards in Western Europe. *New Directions for Evaluation*, 31-42.
- Wiles, J. ve Bondi, J. (1993). *Curriculum development*. U.S.A.: Macmillan Publishing Company.
- Wholey, J. S. (1987). Evaluability assessment: Developing program theory. L. Bickman (edt), *Using program theory in evaluation. New Directions for Program Evaluation* (33) (s. 77-92). San Francisco : Jossey-Bass.
- Worthen, B. R., Borg, W. R. ve White, K. (1993). *Measurement and evaluation in the schools*. NY: Longman Publishing Group.
- Yarbrough, D. B., Shulha, L. M. ve Caruthers, F. (2004). Background and history of the Joint Committee's Program Evaluation Standards. *New Directions for Evaluation* (104), 15-30.
- Yaşar, S., Gültekin, M., Köse, N., Girmen, P. ve Anagün, Ş. (2005). The Meta-evaluation of teacher training programs for elementary education in Turkey. *ATEA: Australian Teacher Education Association 33. Annual Conference* (s. 498-504). Gold Coast, Queensland, Australia: ATEA.
- Yaşar, Ş. (1998). Evaluation of educational programmes in Turkey. *Aera Annual Meeting Session*. CA-USA.
- Yıldırım, C. (1999). *Eğitimde ölçme ve değerlendirme*. Ankara: ÖSYM Yayınları.
- Zukoski, A. ve Luluquisen, M. (2002). Participatory evaluation: What is it? why do it? what are the challenges? *Policy and Practice, Nisan* (5), 3-7