

Geniş Açı: Modeller ve Yaklaşımlar Açısından Türkiye’de Program Değerlendirme Çalışmaları

The Wide Angle: Program Evaluation Studies in Turkey in Terms of Models and Approaches

Tuba GÖKMENOĞLU*

Öz

Çalışmanın amacı Türkiye’de son on yılda yürütülen program değerlendirme çalışmalarının hangi değerlendirme yaklaşım ve modellerine dayandığını ve ne ölçüde paydaş katılımını desteklediğini incelemektir. Bir derleme çalışması olarak desenlenen araştırmada son on yılda yürütülen 18 makale ve 34 doktora tez çalışması betimsel analiz yöntemi ile incelenmiştir. Bulgulara göre dört farklı tema ortaya çıkmıştır; kullanılan program değerlendirme modeli veya araştırma yaklaşımı, değerlendirme çalışmalarının odağı, çalışmalardaki paydaş katılımı ve çalışmaların yürütüldüğü disiplinler. Bulgulara göre 52 çalışmanın yalnızca beşinde program değerlendirme modeli kullanılırken, çalışmaların çoğunda farklı veri kaynakları ve yöntemlerinin kullanıldığı gözlenmiştir. Program değerlendirme çalışmalarının çoğunun Eğitim Bilimleri bölümleri dışındaki bölümlerde yürütüldüğü ise bir başka bulgudur. Program değerlendirme çalışmalarının incelenmesi sonucunda, yürütülen çalışmaların program hakkında karar verme veya program geliştirme sürecinin bir parçası olarak değil de daha çok yayına dönüştürülecek bir araştırma odağı ile yürütüldüğü tartışılmaktadır.

Anahtar sözcükler: Program değerlendirme, program değerlendirme modelleri, derleme çalışması

Abstract

The purpose of the present study is to examine the contents of the current program evaluation studies on the approaches and models in Turkey. Furthermore, this study also aims to explore the extent of stakeholder participation in the program evaluation process. Being designed as a review study, this study examined 18 journal papers and 34 Ph.D. dissertations through descriptive analysis. According to findings, four categories were observed namely used program evaluation models and research approaches, the scope of evaluation studies, stakeholder participation in studies, and disciplines. The results of the study showed that only five studies out of 52 studies prefer using program evaluation model. Furthermore, the studies conducted by the departments other than Educational Sciences, and Curriculum and Instruction departments in different universities. The general conclusion of the presented study is that the studies reviewed in this paper were not a part of curriculum development or program evaluation studies rather they seem to be carried out with purpose of publishing research.

Keywords: Program evaluation, program evaluation models, review study

* Yrd. Doç. Dr., Uluslararası Kıbrıs Üniversitesi, e-posta: tgokmen@ciu.edu.tr

Giriş

Değerlendirmenin öncelikli amacı programlar hakkında kararlar verilirken ve ilgili politikalar düzenlenirken bu süreçte katkıda bulunacak anlamlı bilgi ve verileri sağlamaktır (Patton, 1997; Stufflebeam ve Shinkfield, 1985; Worthen ve Sanders, 1973); ancak araştırmalar bu verilerin uzun bir süre gerektiği gibi kullanılmadığını ortaya koymaktadır (Alkin ve Taut, 2003; Patton ve diğerleri, 1977). Son yıllarda ise bu algı dünya genelinde değişim göstermektedir ve program değerlendirme verilerine yönelik ihtiyaç ve istek her geçen gün artmaktadır. Özellikle yasama organları, kamu kurum ve kuruluşları, vakıflar, kar amacı gütmeyen kuruluşlar ve diğer fon sağlayan kurumlar fon desteklerinin nasıl kullanıldığı, desteklenen programların nasıl geliştirildiği ve kullanılan fonlarla etkin programlar yürütülüp yürütülmediği hakkında daha fazla ve daha detaylı bilgi edinme talebindedir (Wholey, Hatry ve Newcomer, 2010). Bu bilgi talebine yönelik yapılacak değerlendirme programın çalışıp çalışmadığını görmek için değil aynı zamanda Brinkerhoff ve diğerlerinin (1983; p.i) de *“Evaluation is for making it work. If it works ... Notice and nurture. If it doesn't work ... Notice and change.”* ifade ettiği gibi programı daha verimli çalışır hale getirmek amacına yöneliktir. Burada aslında program değerlendirmenin işlevinden söz edilmektedir. Değerlendirme sonunda eğer programın çalıştığı belirlenmişse bu önemli bir bilgidir ve bunun farkında olunup programın daha da geliştirilmesi yönünde çalışılması gereklidir. Eğer değerlendirme sonunda programın çalışmadığı ortaya çıktıysa yine bunun farkında olunup değişikliğe gidilmesi gereklidir.

Yukarıda özetlenen değerlendirme arzı ve işlevine bağlı olarak eğitim dünyasında çeşitli program değerlendirme çalışmaları yapılmaktadır. Benzer konularda yapılan program değerlendirme çalışmaları farklı araştırma yaklaşım ve modelleri ile yapıldığından bir takım farklılıklar göstermektedir. Bu çalışmanın temel odağı ülkemizde yapılan program değerlendirme çalışmalarının kurgusu, yaklaşımı ve tasarımı olduğundan öncelikle program değerlendirme yaklaşım ve modellerine daha sonra da program değerlendirme arzının kaynağı olan paydaş gereksinim, beklenti ve katılımına yönelik alan yazın verileri tartışıldıktan sonra araştırma bulgularının paylaşımına geçilecektir.

Araştırmanın değerlendirmedeki yeri: Program değerlendirme yaklaşım ve modelleri

Değerlendirmeden elde edilecek veri, politika ve programların iyileştirilmesinde kullanılacağından var olan zaman ve kaynakların kullanımına yönelik alınacak kararlar, sosyal bilimlerin sanatsal boyutu olarak nitelendirilebilir (Wholey, Hatry ve Newcomer, 2010). Sanatsal boyutu oluşturan bu kararlar, değerlendirmede hangi sorulara yanıt aranacağına, hangi verinin toplanacağına, verinin nasıl analiz edileceğine ve sonuçlardan elde edilen bilginin nasıl kullanılacağına yön verir. Tüm bu boyutlar ve sorular her araştırmanın başında aynı gibi görünse de, araştırmacının yaklaşımına, felsefik bakış açısına ve yaptığı tasarıma bağlı olarak farklılık gösterecektir. Program değerlendirme ile araştırma benzer işler olarak görülmekle birlikte temelde amaç ve sonuçlarında ayrılmaktadırlar (Fitzpatrick, Sanders ve Worthen, 2004). Araştırma daha çok yeni bilgiyi keşfetmek ve ortaya koymak için yürütülürken program değerlendirme bir programın etkililiği hakkında bilgi edinip yargılarda bulunmak için yapılmaktadır. Yargıda bulunma süreci program değerlendirmenin temel odağıyken araştırma bu yargılara varmak için program değerlendirmeciye veri sağlar (Oliva, 1997). Araştırma süreçlerinde olduğu program değerlendirme çalışmalarının da sistematik ilerlemesi ve geçerlilik ve güvenilirlik konusunda gerekli çalışmalarının yapılması, bu çalışmaların kalite ve değerini arttırmaktadır.

Program değerlendirme çalışmalarının program geliştirme çalışmalarında olduğu gibi bir sistem içerisinde yürütülebilmesi amacıyla Amerika Birleşik Devletleri başta olmak üzere birçok gelişmiş ülkede program değerlendirme uzmanları bir dizi model önermektedir (Yüksel, 2010). Bu modeller temel aldıkları felsefe ve yaklaşımlara göre farklı uzmanlar tarafından farklı şekilde gruplandırılmaktadır. Örneğin Gredler (1996) program değerlendirme modellerini yarırcı (utilitarian) ve çoğulcu (intuitionist/pluralist) olmak üzere ikiye ayırırken Worthen, Sanders ve Fitzpatrick (1997) değerlendirme yaklaşımlarını temel aldıkları öğeye göre amaç odaklı, yönetim odaklı, tüketici odaklı, uzman odaklı, katılımcı odaklı ve rakip (adversary) odaklı olarak altı gruba ayırır. Amaca odaklı

değerlendirme yaklaşımları amaçların belirlenmesine ve belirlenen amaçların çıktılara odaklanırken; yönetim odaklı modeller planlama, uygulama ve değerlendirmeden sorumlu yöneticilere bilgi sağlamaya odaklanır (Worthern ve diğerleri, 1997). Tüketici odaklı modeller ise daha çok eğitsel ürün veya hizmetlerin değerlendirilmesinde yer alan bağımsız değerlendiriciler tarafından geliştirilmiştir (Fitzpatrick ve diğerleri, 2004). Uzman odaklı modeller programı değerlendiren kişilerin değerlendirilen program alanındaki uzmanlığına odaklanmaktadır. Katılımcı odaklı modeller ise programdan etkilenen birden fazla katılımcıya odaklanarak farklı kişilerden farklı yollarla veri toplanması gerektiğini savunmaktadırlar. Son olarak rakip odaklı modeller birden fazla değerlendirme grubunun programa yönelik görüşlerinden yola çıkarak modelin devam edip etmemesi yönünde net kararlara varan bir yaklaşımı benimsemektedir (Gredler, 1996). Bu yaklaşımları benimseyen modellere yönelik bazı örnekler aşağıda yer verilmiştir:

Tyler’in Davranış Merkezli Modeli

Tyler’in (1949) yayımlanmış olduğu *Basic Principles of Curriculum and Instruction* adlı ders izlencesinde yer alan Amaç Odaklı Değerlendirme Modeli, alandaki ilk modellerden biri olarak tanımlanır ve daha sonra geliştirilen birçok modelin de temel çıkış noktası olmuştur (Glatthorn, Boschee ve Whitehead, 2009). Geniş ölçekli birçok değerlendirme çalışmasının ardından rasyonel ve sistematik bir yaklaşıma ulaşan Tyler modeli şu basamaklardan oluşmaktadır: (1) Programın önceden belirlenmiş hedeflerini ortaya koyma (2) Öğrencilerin hedeflere ulaşip ulaşmadığını gösterebilecekleri durumları belirleme (3) Belirlenen hedeflerle uyumlu ölçme ve değerlendirme araçlarını seçme, değiştirme veya geliştirme (4) Ölçme araçlarını kullanarak veri toplama (5) Elde edilen verilerle belirlenen hedefleri karşılaştırma (Tyler, 1949). Tyler’in modeli öğrenci davranışlarının ölçülerek belirlenen hedeflere ulaşılma derecesi hakkında bilgi edinmeyi hedeflemektedir. Sistematik olması bu modelin kullanımını kolaylaştırmaktadır. Ayrıca program sonunda izleme testlerinin önemine de vurgu yapmaktadır. Bu özellikleri nedeniyle Amaç Odaklı Değerlendirme Modeli deneysel çalışmalarda sıklıkla kullanılabilmektedir (Demirel, 2013). Guba ve Lincoln (1981) ise Tyler Modelini hedeflerin nasıl değerlendirileceğinin belirtilmemesi ve standartların nasıl belirleneceği konusunda önerilerde bulunmaması gibi nedenlerle eleştirmektedir.

Stufflebeam’in Bağlam, Girdi, Süreç ve Ürün Modeli

Yukarıda bahsedilen Tyler Modelindeki bazı eksiklikler 1960-1970lerde başka modellerin ortaya çıkmasını hızlandırmıştır (Glatthorn ve diğerleri, 2009). Phi Delta Kappan komitesi başkanı olan Stufflebeam’in yönetim odaklı *Bağlam, Girdi, Süreç ve Ürün* (CIPP) modeli alana büyük bir etkiye bulunmuştur. Bu model değerlendirme sonunda karar vermeyi temele aldığından büyük ölçüde eğitim liderlerine hitap etmektedir. Modelin dört temel boyutu vardır: (1) Mevcut durumun ve ihtiyaçların analiz edildiği *Bağlam*; (2) Belirlenen hedeflere ulaşmada var olan kaynakların ve bunların nasıl kullanıldığının tespit edildiği *Girdi*; (3) Program uygulanırken gerçekleştirilen etkinliklerin incelendiği *Süreç*; (4) Programın çıktılarının tespit edilerek beklenen ile ulaşılan ürün arasında karşılaştırmanın yapıldığı *Ürün* (Stufflebeam, 1971). Bu model de kullanıcılar için adım adım rehberlik oluşturmaya rağmen karar alma sürecinde politikayı gözden kaçırmaması ve sürdürülebilirliğinin zor olması nedeniyle eleştirilmektedir (Guba ve Lincoln, 1981).

Stake’in İhtiyaca Cevap Verici Modeli

Stake (1975) program değerlendirme alanına geliştirmiş olduğu Katılımcı Odaklı İhtiyaca Yönelik modeli ile bu alana en büyük katkıda bulunanlardan biri olmuştur (Glatthorn ve diğerleri, 2009). Çünkü bu model tamamen paydaşları odağa almakta ve onların ihtiyaçlarına yönelik olarak bir değerlendirme yöntemi oluşturmaktadır. Genel olarak ifade etmek gerekirse Stake’in interaktif ve tekrarlı olan modeline göre (1) Değerlendirme uzmanı, paydaşlarla bir araya gelerek onların yaklaşım ve bakış açıları hakkında bilgi edinir. (2) Değerlendiren, değerlendirme projesinin odağını belirleyecek tartışmayı yönetir ve belgeleri analiz eder. (3) Değerlendiren, programın nasıl çalıştığını anlamak için programı yakından gözlemler, beklenen duruma karşıt bir durum varsa not eder. (4) Değerlendiren, projenin belirtilen ve

gerçekte var olan amaçlarını keşfeder, hedef kitleyi bu amaçlar ve değerlendirme hakkında bilgilendirir. (5) Değerlendiren, değerlendirmenin odaklanacağı konuları ve problemleri belirler. Her konu ve problem için değerlendirme deseni oluşturur. (6) Değerlendiren, ihtiyacı olan veriyi elde etmek için gerekli olan araçları belirler. Bu araçlar genelde gözlemciler veya karar vericilerdir. (7) Değerlendiren, veri toplama prosedürlerini uygular. (8) Değerlendiren, bilgiyi tema ve kategorilere ayırır. Yaptığı betimlemeler örnek olaylar, videolar ya da eşdeğer temsillerden oluşabilir. (9) Paydaşlara hassasiyet gösteren bu model, sonuç raporunu, bu raporu sunacağı hedef kitlenin özelliklerine en uygun formatı seçerek hazırlar (Glatthorn, 1987). Aşamalarından da anlaşılacağı üzere bu model paydaş katılımına ve hedef kitleye odaklı olması ve değerlendiren kişiye yöntem ve desen seçme konusunda esneklik sağlaması bakımından avantajlıdır; ancak model doğası gereği hedef kitlenin program değerlendirmede odağı istedikleri doğrultuda değiştirme veya tahrif etme riskini doğurmaktadır (Glatthorn ve diğerleri, 2009).

Eisner'in Eğitsel Uzmanlık ve Eleştiri Modeli

Uzmanlık ve eleştiri olmak üzere iki bileşenden oluşan Eisner'in modelinde uzmanlık bilgi ve tecrübeyi kullanarak bir durumun önemli boyutları ve özellikleri hakkında karara varılması; eleştiri ise olumsuz bir yargılamadan öte temel niteliklerin tekrar değerlendirilmesi şeklinde tanımlanabilir (Gredler, 1996). Eisner (1979) tarafından geliştirilen bu model daha çok sınıf içi etkileşimlere odaklanmaktadır. Sınıf içindeki durumların nitel analizlerine daha çok yer veren bu modelde eğitsel eleştirileri bu konuda geniş bilgisi ve tecrübesi olan kişiler yapabilir. Bu modele göre (1) Öncelikle sınıf içi uygulamalar, kurallar ve düzenlemeler yani eğitsel yaşantıların tümü betimlenir. (2) Programın uygulanmasına dair ortaya çıkacak sonuçlar hakkında yorumlama yapılır ve (3) Betimleme ve yorumlama sonuçlarına göre program hakkında bir değerlendirmede bulunulur. Sınıf etkinlikleri ve yaşantılara ağırlık vermesinden dolayı diğer modellere göre bir üstünlüğü olduğu kabul edilebilmekle birlikte, bu model fazlaca uzmanlık gerektirmesi ve yönteminin net olmamasından dolayı eleştirilmektedir (Glatthorn ve diğerleri, 2009).

Görüldüğü gibi yaklaşımları farklı olan değerlendirmecilerin amaçları ve yöntemleri de birbirinden farklı olmaktadır. Yukarıda dört farklı modele değinilmiş olup, alan yazında çok sayıda başka modellerinde de var olduğunu belirtmekte yarar vardır. Her modelin değerlendirme boyutlarına ilişkin sorduğu sorular farklılık gösterebildiğinden her değerlendirme tasarımı, yaklaşımı veya modeli kendine özgülük göstermektedir. Uygun durum ve program üzerinde kullanıldığında, gerçekte var olan durumun önemli bir boyutunu ortaya çıkarması nedeniyle her model değerlidir (Wholey, Hatry ve Newcomer, 2010). Özetle, değerlendirme modelleri, araştırmacılara değerlendirme çalışmalarının nasıl planlanacağı ve nasıl yürütüleceği konusunda akademik danışmanlık yaparak çok önemli bir yol haritası sunmaktadır (Madaus ve Kellaghan, 2000).

Kimin değerlendirmesi? Paydaş ihtiyaç ve beklentileri

Paydaş, değerlendirme süreç ve sonuçlarından etkilenen kişiler, gruplar ya da organizasyonlar şeklinde tanımlanabilir (Wholey ve diğerleri, 2010). Anahtar paydaşlar ise Patton'a (2008) göre öncelikli belirlenen kullanıcılar olarak tanımlanmaktadır. Alan yazındaki deneyimler, anahtar paydaş katılımını sağlamanın doğru araştırma sorularının sorulmasını, uygun değerlendirme deseninin belirlenmesini, uygulanma sürecini kolaylaştırmasını ve değerlendirme sonuçlarının karar alma aşamasında kullanımını büyük ölçüde desteklediğini göstermiştir (Patton, 2008; Taylor-Powell, Steele ve Douglass, 1996). Değerlendirmenin bu boyutlarına yönelik desteğinin yanı sıra daha demokratik ve eşitlikçi bir değerlendirme yapılması açısından da çoklu paydaş katılımının önemi yadsınamaz (Alex, 1995). Ayrıca toplanan verinin geçerli ve güvenilir olması açısından da veri kaynağına yönelik üçlemenin (data triangulation) yapılarak çoklu veri toplanması gerektiği konusunda araştırmacılar hemfikirler (Fraenkel ve Wallen, 2005).

Değerlendirme modelleri tasarım ve yöntem boyutlarında değerlendirene rehberlik ederken aslında araştırmanın paydaşlarının kimler olacağı konusunda da fikir vermektedir. Örneğin Eisner’ın Eğitsel Uzmanlık ve Eleştiri modelinde alan uzmanları en önemli paydaşlardan biri kabul edilirken Stake’in İhtiyaca Cevap Verici modelinde öğretimi yapanlar ve öğrenenler öncelikli paydaşlar olarak kabul edilmektedir. Paydaş katılımı genel olarak değerlendirenin felsefik bakış açısına ve dolayısıyla tercih ettiği yaklaşım veya modele göre değişiklik göstermekle birlikte araştırmacılar genel olarak çoklu paydaş katılımının desteklenmesini önermektedir. Ancak değerlendirenler çoklu paydaş katılımını destekleyen bir bakış açısına sahip olsalar da birçok değerlendirme uzmanı yalnızca politika yapıcılarının ve yöneticilerin ihtiyaç ve beklentilerine yönelik desenleme yaparak çoğu zaman birçok paydaşı çalışma dışında bırakılmaktadırlar (Alex, 1995; Wholey ve diğerleri, 2010).

Yukarıda program değerlendirmenin yaklaşım ve modellerine, kimler tarafından yürütüldüğüne ve paydaş katılım yöntemine yönelik özet bir bilgilendirme yapılmaya çalışılmıştır. Ülkemizde son yıllarda çok sayıda program değerlendirme çalışmasının yürütüldüğü gözlemlenmektedir. Bu çalışmaların genel bir değerlendirmesine yönelik ise hem ülkemizde hem de uluslararası alan yazında yeterli sayıda kapsamlı çalışmaların yürütülmemiş olduğu ileri sürülebilir. Değerlendirme çalışmalarının genel değerlendirmesine yönelik yürütülen çalışmalar, programcılara, politika yapıcılara, yerel yöneticilere, öğretmenlere ve hatta öğrencilere önemli dönütler verdiğinden bu tür çalışmaların yapılması çok önemli görülmektedir. Değerlendirmenin genel değerlendirmesine yönelik olmamakla birlikte alan yazında çok sayıda bir veya iki tür programın değerlendirme çalışmalarının değerlendirmelerine yönelik oldukça kapsamlı projeler ve çalışmalarla karşılaşılacaktır. Örneğin Clewell ve diğerleri (2004) oldukça kapsamlı olan çalışmalarında 400 tane fen ve matematik programının değerlendirme çalışmalarını içerik analizi yaparak alan yazına önemli dönütlerde bulunmuşlardır. Bir diğer örnekte ise Rolstad, Mahoney ve Glass (2005) İngilizce öğrenimde program etkililiğine yönelik 17 çalışmayı kapsayan bir meta analiz çalışması yürütmüşlerdir. Tobler ve diğerleri (2000) ise okullarda madde kullanımını önleme programlarının etkinliği konusunda 207 çalışmanın sonuçlarını meta analiz ile değerlendirmişlerdir. Ülkemizde ise Özdemir (2009) 2004 ilköğretim programlarının değerlendirilme sürecine yönelik nitel bir çalışma yayınlamıştır. Tüm bu ve benzeri değerlendirme çalışmalarının değerlendirilmesine yönelik çalışmalar yukarıda da belirttiği gibi öncelikle program geliştirenlere ve diğer paydaşlara önemli dönütler sağlamaktadır.

Bu noktada, *program değerlendirme çalışmaları ülkemizde nasıl yapılmaktadır?* sorusu araştırmamıza yön verecektir. Daha özel olarak, bu çalışmanın amacı, Türkiye’de son on yılda yürütülen program değerlendirme çalışmalarının hangi değerlendirme yaklaşım ve modellerine dayandığını belirlemek ve ne ölçüde paydaş katılımını desteklediğini incelemektir. Bu amaca ulaşmak için aşağıdaki araştırma soruları çalışmaya rehberlik etmektedir:

1. Türkiye’de son on yılda yapılan program değerlendirme çalışmalarında hangi araştırma yaklaşımları veya değerlendirme modelleri kullanılmıştır?
2. Türkiye’de son on yılda yapılan program değerlendirme çalışmaları araştırma odaklı mı değerlendirme odaklı mı yapılmıştır?
3. Türkiye’de son on yılda yapılan program değerlendirme çalışmalarında paydaşlar ve öğrenen ihtiyaçları ne ölçüde değerlendirmeye dahil edilmiştir?
4. Türkiye’de son on yılda yapılan program değerlendirme çalışmaları hangi disiplinler ve bölümler tarafından yürütülmüştür?

Yöntem

Derleme çalışması niteliğinde olan bu araştırmada Türkiye’de son on yılda yapılmış doktora tezleri ve hakemli dergilerde yayınlanan program değerlendirme çalışmaları betimsel analiz yöntemi ile analiz edilerek yorumlanmıştır. Betimsel analiz yöntemi nitel verinin düzenlenmesi, sınıflandırılması, karşılaştırılması ve bütüncül bir anlayışla özetlenmesidir (Cohen, Manion ve Morrison, 2007; Fraenkel ve Wallen, 2005). Analiz yapılırken araştırmacı kategoriler belirler ve her kategoriye giren örneklerin sayısını tespit eder (Silverman, 2006). Aslında bu yöntem ile benzer veriler bir araya getirilerek belirlenen

temalar altında veri araştırmacı tarafından yorumlanır (Yıldırım ve Şimşek, 2006). Derleme çalışmalarını güçlü ve etkili yapan yanı ise aynı veya benzer alanlarda yapılmış çalışmaları bir araya getirerek genel bir değerlendirilmeye gidilmesinin yanında gözden kaçmış ve bulut altında kalmış noktaların açığa çıkarılmasıdır.

Veri kaynağı

Türkiye’de son on yılda yapılmış program değerlendirme çalışmaları taranırken öncelikle YÖK tarafından onaylanmış doktora tezlerine yer verilmiştir. YÖK tez tarama motorundan indirilen 34 doktora tezi ve eğitim alanında yayın yapan sekiz hakemli dergiden toplam 18 makale çalışmanın örneklemini oluşturmaktadır. Akademik dergilerin belirleme aşamasında Orta Doğu Teknik Üniversitesi tarafından dergi sürekliliği, hakem ve yazar çeşitliliği, makale sayısı, bilimsel nitelik gibi kriterlere bağlı olarak hazırlanan ulusal dergi gruplaması temel alınmış ve burada A, B ve aday B gruplarında bulunan eğitim dergileri çalışmanın örnekleme dahil edilmiştir (ODTÜ, 2014). Buna göre Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, Eğitim ve Bilim, Kuram ve Uygulamada Eğitim Yönetimi Dergisi, Gazi Eğitim Fakültesi Dergisi, Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, İlköğretim Online ve Kuram ve Uygulamada Eğitim Bilimleri dergilerinin arşivleri veri kaynağı olarak belirlenmiştir. Listedeki eğitim dergilerinden Eğitim Araştırmaları Dergisi’nin arşivine ulaşamadığından örnekleme alınamamıştır. Bu listeye ek olarak çalışmanın konusu ve disiplini göz önünde bulundurularak Eğitim Programları ve Öğretim Derneği tarafından yayımlanan Uluslararası Eğitim Programları ve Öğretim Çalışmaları Dergisi de örnekleme dahil edilmiştir. Yayımlanan makale ve doktora tezlerinin güncel olması amacıyla 2004-2014 yılları arasında yayınlanmış olmasına dikkat edilerek örneklem ölçüte bağlı olarak sınırlandırılmıştır.

Toplanan veri, çalışmaların odaklandığı program türlerine göre gruplandırıldığında üç ana kategori ortaya çıkmaktadır. Buna göre çalışmalar ilköğretim ve orta öğretim programları, yükseköğretim programları ve diğer olmak üzere gruplandırılmıştır. Çalışmaların odaklandığı program türleri Tablo 1’de özetlenmektedir.

Tablo 1

Çalışmaların Odaklandığı Program Türleri (N=53)

<i>Program Türü</i>	<i>f</i>	<i>%</i>
İlköğretim Eğitim Programları	5	9.43
İlköğretim Matematik Öğretim Programları	1	1.89
Teknoloji ve Tasarım Öğretim Programları	1	1.89
Fen ve Teknoloji Öğretim Programları	5	9.43
Kimya Öğretim Programları	3	5.66
İlköğretim ve Orta Öğretim Türkçe Öğretim Programları	4	7.55
İngilizce Öğretim Programları	4	7.55
Coğrafya Öğretim Programları	1	1.89
Din Kültürü ve Ahlak Bilgisi Öğretim Programları	1	1.89
Vatandaşlık ve İnsan Hakları Öğretim Programları	1	1.89
Biyoloji Öğretim Programları	1	1.89
Hayat Bilgisi Öğretim Programları	1	1.89
Sosyal Bilgiler Öğretim Programları	2	3.77
Orta Öğretim Matematik Öğretim Programları	1	1.89
İlköğretim ve Orta Öğretim Sınıf Rehberlik Programı	1	1.89

Yüksek Öğretim	Gelişim ve Öğrenme Öğretim Programı	1	1.89
	Orta Öğretim Coğrafya Eğitimi Programı	6	11.32
	Beden Eğitimi ve Spor Eğitimi Programı	1	1.89
	Öğretmenlik Uygulaması/ Okul Deneyimi Öğretim Programı	2	3.77
	Dil Geliştirme Öğretim Programı	1	1.89
	Güzel Sanatlar Fakültesi Geleneksel Tekstil Ürünlerinin Konservasyon ve Restorasyonu Öğretim Programları	1	1.89
	Yüksek Lisans Programı	1	1.89
	Eğitim Fakültesi Öğretmenlik Eğitimi Programları	1	1.89
	Doktora Programları	1	1.89
Diğer	ERASMUS Programı	1	1.89
	BIYEP ve OYP	1	1.89
	Birleşmiş Milletler Eğitim Programı	1	1.89
	Bilim ve Sanat Merkezi Eğitim Programları	1	1.89
	Öğretmenlere yönelik Hizmet-içi Eğitim Programları	2	3.77

Tablo 1’e göre son 10 yılda yürütülen program değerlendirme çalışmalarının büyük kısmının ilk ve orta öğretim programlarına yönelik olduğu gözlenmektedir.

Veri toplama süreci ve analizi

Makale ve tez seçimi yapılırken öncelikle arşivdeki tüm makalelerin özleri okunmuş, program değerlendirme çalışması olduğu düşünülen makale ve tezlerin tam metinleri indirilmiştir. Bu makaleler detaylı şekilde incelenerek konunun belirli bir bölümüne odaklanmış çalışmalar yerine hedeflenen programı kapsamlı bir şekilde değerlendiren çalışmalar örnekleme dahil edilmiştir. Örneğin 8. sınıf Matematik öğretim programının fraktal geometri konusu açısından değerlendirmesine yönelik yapılmış bir çalışma araştırmanın örnekleme dahil edilmezken, 8. sınıf Matematik öğretim programının öğretmen görüşleri açısından değerlendirilmesine yönelik yürütülmüş daha kapsamlı bir çalışma örnekleme dahil edilmiştir. Bu aşamadan sonra her çalışmanın künyesi, araştırma modeli, yaklaşım ve yöntemi, odağı, örnekleme, konu alanı, veri toplama süreci ve analiz yöntemleri hazırlanan bir tabloya kaydedilmiştir. Bu aşamada veri detaylı incelenerek araştırma kapsamında kategoriler ve temalar oluşturulmuştur (Miles ve Huberman, 1994). Olası hataların önüne geçilmesi amacıyla veriden elde edilen kategori ve temalar araştırmacı ile birlikte Eğitim Programları ve Öğretim alanından gönüllü iki meslektaş tarafından kontrol edilmiştir. Toplanan veri betimsel istatistiki yöntemler ile analiz edilmiş; yüzde ve frekans cinsinden yorumlanmıştır.

Geçerlik ve güvenirlik

Araştırma kapsamına alınan tüm makale ve tezler üç araştırmacıyla da paylaşılmıştır. Çalışmaların sınıflandırılması ve karşılaştırılması iki aşamada gerçekleştirilmiştir. İlk aşamada tez arama ve dergi arşivlerine girme işleri araştırmacılar tarafından paylaşılmıştır. İkinci aşamada araştırmacılar kendi bölümlerinde topladıkları veriyi diğer araştırmacıların kontrolüne sunmuştur. Veri, diğer araştırmacılar tarafından da diğer araştırmacılardan bağımsız bir şekilde detaylı olarak incelenmiştir. Böylece çapraz kontrol sağlanarak çalışmanın iç geçerlilik ve güvenirliği sağlanmaya çalışılmıştır.

Bulgular

Bu çalışmada Türkiye’de yürütülen program değerlendirme çalışmalarının, kullanılan program değerlendirme modelleri ve araştırma yöntemleri açısından, incelenmesi amaçlanmıştır. Bu amaçla çalışma kapsamına alınan 18 makale ve 34 doktora tezinin çeşitli program ve dersleri değerlendirdiği

gözlenmiştir. Buna göre ilköğretim, ortaöğretim ve yükseköğretim eğitim programlarından, çeşitli öğretim programlarına ve öğretmenlere yönelik düzenlenen hizmet içi eğitim kurslarına kadar farklı türde programların değerlendirildiği gözlenmiştir. Çalışma sonunda elde edilen bulgular araştırma sorularına paralel olarak *Çalışmalarda Kullanılan Değerlendirme Model ve Yaklaşımları*, *Program Değerlendirme Çalışmalarının Odağı*, *Çalışmalarda Paydaş Katılımı* ve *Disiplinler* olmak üzere dört ana başlıkta sunulmaktadır.

Çalışmalarda kullanılan değerlendirme model ve yaklaşımları

Son 10 yılda yapılan program değerlendirme çalışmaları incelendiğinde örnekleme dahil edilen 18 makale ve 34 tez arasından yalnızca iki makale ve üç tez çalışmasında literatürde yer alan *program değerlendirme modellerinin* kullanıldığı görülmüştür. Bu beş çalışmanın üçünde Stufflebeam'ın CIPP (Bağlam, Girdi, Süreç ve Ürün) modeli kullanılırken; bir çalışmada Stake'in *Uyumluluk* modeli ile Provus'un *Farklılıklar* modeli bir arada kullanılmıştır. Diğer iki çalışmada ise Hammond'ın *Değerlendirme Modeli* ve Bellon ve Handler'in *Müfredat Değerlendirme* modelinin kullanıldığı belirlenmiştir.

Tablo 2

Program Değerlendirme Çalışmalarında Kullanılan Araştırma Yöntemleri (N=53)

<i>Araştırma Türü</i>	<i>Veri Toplama Yöntemi</i>	<i>f</i>	<i>%</i>
Nitel	Görüşme ve Doküman Analizi	6	11.32
	Gözlem, Görüşme ve Doküman Analizi	6	11.32
	Doküman Analizi	4	7.55
	Görüşme	4	7.55
	Gözlem ve Görüşme	2	3.77
Nicel	Anket/ Ölçek	10	18.87
	Anket ve Başarı Testi	2	3.77
Karma	Anket ve Görüşme	8	15.09
	Anket ve Doküman Analizi	2	3.77
	Anket, Görüşme, Gözlem ve Doküman Analizi	2	3.77
	Anket, Görüşme ve Gözlem	1	1.89
	Anket, Görüşme ve Başarı Testi	1	1.89
Deneysel	Deney ve Kontrol Gruplu Desenler	2	3.77

Tüm çalışmalarda tercih edilen araştırma desen ve veri toplama yöntemleri incelendiğinde ortaya dört kategori çıkmaktadır. Buna göre ilk kategori olan *Nitel* ($n=22$) araştırmalarda beş tür veri toplama yönteminden bahsedilebilir. Bunlardan en sık gözlenenleri görüşme ve doküman analizlerinin birlikte kullanıldığı ($n=6$) ve gözlem, görüşme ve doküman analizlerinin ($n=6$) bir arada tercih edilmesiyle oluşan desenlerdir. İkinci kategori olan *Nicel* ($n=12$) araştırma türüne yönelik olarak da iki farklı desenden söz edilmektedir. Bunlar yalnızca anket/ ölçek kullanılan çalışmalar olup, ($n=10$) genel sonuçlarda da en yüksek orana sahiptir. *Karma* ($n=14$) desen olan üçüncü kategoride ise farklı beş veri toplama kombinasyonu gözlenmiştir. Bu desenler arasında en sık tercih edilen hem anket uygulamasının hem de birebir görüşmelerin ($n=8$) yapıldığı çalışmalardır. Son kategoride ise deneysel iki çalışma yer almaktadır. Bu çalışmalarda hem deney hem de kontrol grubun olduğu desenler tercih edilmiştir. Tüm sonuçlar Tablo 2'de özetlenmektedir.

Program değerlendirme çalışmalarının odağı

Çalışmada ikinci araştırma sorusuna yanıt vermek üzere Türkiye’de son on yılda yapılan program değerlendirme çalışmalarının temel amacı veya odaklandığı noktalar incelenmiştir. Buna göre çalışmaların tümü araştırma odaklı olarak yani tez veya makale ürününe dönüşen araştırmalar olarak yürütülmüştür. Buna karşın incelenen çalışmalarından iki doktora tezinde EARGED (2012 yılında kapatılmış olan Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı) desteğinden bahsedilmektedir. Geniş örnekleme sahip çalışmaların dahi tam metni incelendiğinde bu örneklemelere ulaşmada bakanlıktan veya üniversitelerden herhangi bir destek alınıp alınmadığına yönelik bir bilgiye ulaşılamamıştır. Dolayısıyla bu noktada, bu çalışmalar elbette ki bir veri sunmuşlardır ancak elde edilen bu verilerin ilgili kişilerce kullanılıp kullanılmadığı, bu çalışmaların bir program geliştirme sürecinin parçası olarak ya da yürütülen bir programın etkililiği hakkında bilgi edinerek yargılarda bulunmak -program değerlendirme- amacıyla mı yürütüldüğü konusunda kesin yargılara varmak mümkün değildir.

Program değerlendirme çalışmalarında paydaş katılımı

Üçüncü araştırma sorusuna yanıt vermek üzere program değerlendirme çalışmalarında hangi kaynaklardan veri toplandığı incelenmiştir. Analiz sonucunda çalışmaların %9.43’ünün ($n=5$) yalnızca var olan dokümanların analizine bağlı olarak yapıldığı gözlenmektedir. Diğer çalışmalarda ise hedeflenen programın değerlendirilmesinde genellikle programdan yararlanan paydaşlardan değil programın doğrudan kullanıcılarından veri toplanmıştır. Buna göre çalışmaların çoğunda ($n=19$) hem programı yürüten öğretmenlerden hem de programa dahil olan öğrencilerden veri toplanırken çalışmaların %25.53’ünde ($n=12$) yalnızca programı yürüten öğretmenlerden veri elde edildiği görülmüştür. Çalışmaların %14.89’unda ($n=7$) ise çoklu paydaş katılımı ile veri sağlandığı gözlenmiştir. Buna göre öğretmen ve öğrencilerin yanı sıra çalışmaya veliler, yöneticiler, program geliştirme uzmanları, alan uzmanları, bakanlık personeli ve müfettişler de katılmıştır. Değerlendirme modeli kullanılan beş çalışma ayrıca incelendiğinde bu çalışmaların tümünün birden fazla paydaş katılımını desteklediği gözlemlenmiştir. Sonuçlar Tablo 3’de özetlenmektedir.

Tablo 3

Program Değerlendirme Çalışmalarında Veri Kaynakları (N=47)

<i>Paydaşlar</i>	<i>f</i>	<i>%</i>
Programda Eğitimci/ Öğretmen	12	25.53
Programdan Yararlanan/ Öğrenci	6	12.76
Programdan Yararlanan/ Veli	1	2.13
Hem Öğretmen Hem Öğrenci	19	40.43
Hem Öğretmen Hem Veli	1	2.13
Çoklu Paydaş Katılımı	7	14.89

Disiplinler

Tablo 4’te son on yılda yapılan program değerlendirme çalışmalarının hangi bölümlerde yürütüldüğü özetlenmektedir. Tablo 4 genel olarak incelendiğinde çalışmaların yarısından azının (toplamda $n=18$, %33.96) benzer ancak ismi farklı dallar olan *Eğitim Bilimleri* ve *Eğitim Programları ve Öğretim* bölümlerinde tamamlandığı sonucuna varılabilir. Son on yılda *Eğitim Programları ve Öğretim* anabilim dalında sonra ise en yüksek sayıda program değerlendirme çalışmasını Ortaöğretim Sosyal Alanlar Eğitimi ABD Coğrafya Öğretmenliği ($n=6$) ve Ortaöğretim Fen ve Matematik Eğitimi ABD Matematik Öğretmenliği ($n=5$) bölümlerinden araştırmacıların yürüttüğü söylenebilir.

Tablo 4*Program Değerlendirme Çalışmalarının Yürütüldüğü Alanlar/Bölümler (N=52)*

<i>Alan/Bölüm</i>	<i>f</i>	<i>%</i>
Eğitim Bilimleri ABD Eğitim Programları ve Öğretim BD	10	18.87
Eğitim Bilimleri ABD	6	11.32
Eğitim Programları ve Öğretim ABD	1	1.89
Eğitim Programları ABD Eğitimde Program Geliştirme BD	1	1.89
Ortaöğretim Sosyal Alanlar Eğitimi ABD Coğrafya Öğretmenliği BD	6	11.32
Ortaöğretim Fen ve Matematik Eğitimi ABD Matematik Öğretmenliği BD	5	9.43
İlköğretim ABD Matematik Eğitimi BD	3	5.66
Yabancı Dil Eğitimi BD	3	5.66
Kimya Eğitimi ABD	2	3.77
Türkçe Eğitimi ABD	2	3.77
İlköğretim ABD Fen Bilgisi Eğitimi BD	2	3.77
Eğitim Bilimleri ABD Eğitim Yönetimi ve Politikası BD	1	1.89
Ortaöğretim Fen ve Matematik Eğitimi ABD Biyoloji Öğretmenliği BD	1	1.89
Eğitim Bilimleri ABD Rehberlik ve Psikolojik Danışmanlık BD	1	1.89
İlköğretim ABD Sosyal Bilgiler Öğretmenliği BD	1	1.89
Beden Eğitimi ve Spor ABD	1	1.89
Güzel Sanatlar Eğitimi ABD Resim is Öğretmenliği BD	1	1.89
İlahiyat ABD Din Eğitimi BD	1	1.89
Yaşam boyu Öğrenme ve Yetişkin Eğitimi ABD	1	1.89
Özel Eğitim ABD	1	1.89
Matematik ABD	1	1.89

Araştırma bulgularına ek olarak model kullanılan çalışmalar ayrıca incelendiğinde program değerlendirme modeli kullanılan çalışmaların farklı üniversitelerin Eğitim Bilimleri ve Eğitim Programları anabilim dalı bölümlerinde yürütüldüğü gözlenmiştir.

Tartışma ve Sonuç

Bu araştırmada, Türkiye’de son on yılda yürütülen program değerlendirme çalışmalarının hangi değerlendirme yaklaşım ve modellerine dayandığı ve ne ölçüde paydaş katılımını desteklediği incelenmiştir. Araştırma sonuçlarına göre son on yılda çok sayıda program değerlendirme çalışmasının farklı alanlar ve disiplinler tarafından yürütüldüğüne dikkat çekmek gerekmektedir. Bu incelemede sadece genel ve kapsamlı program değerlendirme çalışmaları örnekleme dahil edilmesine rağmen son on yıllık sürede gerçekleştirilmiş oldukça fazla sayıda (52) çalışmaya ulaşılmıştır. Alan yazında Türkiye’de yürütülen program değerlendirme çalışmalarının daha çok ders değerlendirme boyutunda kaldığı ve yükseköğretim programlarında daha çok derslerde yapılan dar kapsamlı ve çoğunlukla anket odaklı çalışmaların yayına dönüştürüldüğü yönünde eleştiriler bulunmaktadır (Yüksel, 2010). Benzer şekilde Yaşar ve diğerleri (2005) de öğretmen eğitimi programlarının değerlendirilmesine yönelik yürütmüş olduğu çalışmada, Türkiye’de yürütülen çalışmaların daha çok nicel yöntemlerle, programların yalnızca tek bir boyutuna yönelik, program değerlendirme sürecini genel olarak yansıtmayan ve küçük

örneklemelerle ile yürütülmüş genellenemeyen çalışmalar olduğunu vurgulamaktadırlar. Alan yazında belirtilenin aksine, bu çalışmanın bulgularına göre örneklemde yükseköğretim derslerinin değerlendirilmesine yönelik bazı çalışmaların bulunmasına rağmen örneklemdeki çalışmaların büyük çoğunluğunun nitel araştırma yöntemleri veya karma desen kullanılarak yürütüldüğünü ve üniversite derslerinin değerlendirilmesinden çok ilköğretim ve ortaöğretim programlarının değerlendirilmesinin yapıldığı belirlenmiştir. Bulguların alan yazındaki eleştirileri desteklemez nitelikte olması seçilen dergilerin belirli bir kalite düzeyinde olması ve tezlerin doktora düzeyinden seçilmesi ile açıklanabilir.

Çalışmanın temel odağı olan program değerlendirme model ve yaklaşımlarına yönelik araştırma sorusu bulgular tarafından Türkiye’de son on yılda yürütülen program değerlendirme çalışmalarının çoğunda model kullanılmadığı şeklinde yanıtlanmıştır. Alan yazın tarafından da model kullanımının önemi destekleniyor olmasına rağmen 52 çalışmanın sadece beşinde model kullanılmış olması ilginç bir bulgudur. Bu bulgu yayımlanan çalışmaların 34’ünün Eğitim Bilimleri ve Program Geliştirme veya Eğitim Programları ve Öğretim alanlarından farklı alanlarda yürütülmüş olması ile ilişkilendirilebilir. Eğitim Bilimleri ve Program Geliştirme veya Eğitim Programları ve Öğretim alanlardakilerin aksine eğitim fakültelerinin diğer bölümlerindeki araştırmacıların program değerlendirmeye yönelik olarak almış oldukları dersler ve bu konudaki uzmanlaşma düzeyleri bu alana yönelik olarak yaptıkları çalışmaların teorik temel açısından zayıf ve model kullanımının sınırlı olması neden olabilmektedir. Yine bu çalışmalarda program geliştirme uzmanlarına paydaş olarak yer verilmeyişi de program değerlendirme modellerinin kullanılmama nedenleri arasında gösterilebilir. Oysaki program değerlendirme sürecinin en başında oluşturulacak değerlendirme komitesi üyelerinin seçimi en önemli adımlardan biridir (Hage, 2013). Çünkü programı değerlendiren uzmanın alt yapısı ve yeterlilikleri nasıl bir araştırma deseni tercih edeceğinden kullanacağı yöntem ve modellere kadar değişiklik göstermesinde önemli değişkenlerden biridir (Azzam, 2011). Paydaşlar, programdan doğrudan etkilenen aynı zamanda aldıkları kararlar ile programın geleceğini doğrudan etkileyen kişiler olarak tanımlandığından (Upreti, Liaupsin ve Koonce, 2010), program değerlendirme sürecinin en önemli paydaşlarından biri de program geliştirme uzmanlarıdır. Araştırmalar da göstermektedir ki bu uzmanların ve diğer çeşitli paydaşların yer almadığı bir program değerlendirme sürecinin geçerliliğinin program geliştirme uzmanlarının ve diğer paydaşların katılımı ile yürütülen değerlendirme çalışmalarına göre geçerliliği çok daha düşüktür (Brandon, 1998). Yine bu alanın önemli isimlerinden olan Kaplan (1964), Patton (1987) ve Cronbach (1980) değerlendirme yapan uzmanların her bağlama uyacak belli desenleri ve tasarımları tercih etmek yerine verilen bağlama en uygun desen ve tasarımı seçebilecek şekilde eğitilmeleri konusunu şiddetle savunmaktadırlar.

Türkiye’de son on yılda yapılan program değerlendirme çalışmaları araştırma odaklı mı değerlendirme odaklı mı yapılmıştır? olarak ortaya konulan ikinci araştırma sorusunda ise çalışmaların ne amaçla yürütüldüğü sorgulanmaktadır. Alan yazında da belirtildiği gibi bazı çalışmalar araştırma odaklı yapılırken bazıları ise program hakkında yargılara varmak amacıyla yürütülmektedir (Fitzpatrick ve diğerleri, 2004). Alan yazındaki bir başka ve daha sert bir görüş ise program değerlendirme sonuçlarının ancak programa uygulandığında bir anlam ifade ettiği yönündedir (Westbury, 1970). Bulgular yapılan program değerlendirme çalışmaları yürütücülerinin büyük çoğunluğunun değerlendirilen programlardan bağımsız araştırmacılar olduğunu ve değerlendirdikleri programın geleceğine yön verme anlamında karar verme yetkisine sahip olmadığını ortaya koymaktadır. Bir diğer söylemle araştırmacıların yaptığı çalışmalar ilgili otoriteler tarafından program hakkında karar vermeye yönelik olarak işe koşulmamışlardır. Bu bulgu ilk araştırma sorusunun yanıtını da açıklar niteliktedir. Buna göre incelenen çalışmalarda değerlendirme modellerinin kullanılmama nedenlerinden birinin de, değerlendirilen programların devam ettirilip ettirilmeyeceği ya da hangi noktalarının değiştirileceği hakkında araştırmacıların gerekli yetkiye sahip olmaması, olduğu öne sürülebilir. Çünkü program değerlendirme modelleri temel olarak programın geleceği hakkında karar verme amacı güder (Fitzpatrick ve diğerleri, 2004). Oysaki örneklemdeki çalışmalar genel olarak yayına dönüştürülecek bir araştırma olarak yürütülmüşlerdir. Bu durumda Westbury (1970)’nin belirttiği gibi programlar üzerinde etkisi olmayan bu çalışmaların program alanındaki değeri tartışılmalıdır.

Değerlendirme modeli kullanılan az sayıdaki çalışmaların tümünde veri toplama yöntemi, örneklem ve veri çeşitliliğinin zengin olduğu görülmüştür. Model kullanılmayan program değerlendirme çalışmalarının çoğunda yine paydaş katılımının ve çeşitliliğinin yüksek olduğu gözlenmiştir. Buna rağmen tek bir yöntemle ve tek tür katılımcı gruplarıyla yapılan çalışmaların hatırı sayılır miktarda olduğunu da belirtmekte yarar vardır. Bu şekilde yapılan araştırma desenleri de güvenilirlik ve iç geçerlilik açısından bazı sorunlar teşkil etmektedir (Miles ve Huberman, 1994; Patton, 2008). Denzin (1978)'e göre üçlemenin (triangulation) dört farklı boyutu vardır; veri kaynağı (katılımcılar), teoriler, yöntemler ve araştırmacılar. Bu üçlemeler yapılmadığında yürütülen çalışmaların sonuçlarının var olan durumu gerçekten yansıtıp yansıtmadığı konusu şüphe uyandırmaktadır (Creswell ve Miller, 2010; LeCompte, 2010). Bu durumda, Türkiye’de yapılan program değerlendirme çalışmalarının bir kısmının sonuçları hakkında yorum yapılırken dikkatli olmak gerekmektedir. Yine araştırmalar göstermektedir ki paydaş katılımının desteklenmediği değerlendirme çalışmalarında çıkan sonuçlar ve yapılacak olan düzeltmelere programın kullanıcılarının programı uygulamaya yönelik isteği önemli ölçüde olumsuz etkilenmektedir. Bunun aksine eğer paydaşlar bu süreçlerde yer alırlarsa programa sahip çıkarak düzeltme ve reformları uygulama yönünde daha istekli olmaktadır (Greene, 1988).

Bu çalışmada toplanan veriden, örneklemdeki program değerlendirme çalışmalarından elde edilen sonuçların ve hazırlanan raporların ne şekilde ve kim tarafından kullanıldığına dair bir bilgi edinilmesi mümkün olmamıştır. Ayrıca aynı programların farklı şehirlerde yeniden değerlendirilmesinin yapıldığı çalışmalarda, bunun neden gerekli olduğu ve yazına nasıl bir katkı sağlayacağı konusunda herhangi bir bilgilendirici bölüm bulunmamaktadır. Örneğin Fen ve Teknoloji ders programları son on yılda beş kez program değerlendirmeye tabi tutulmuştur. Neden bir önceki yıl farklı şehirde fakat aynı şekilde yapılmış bir çalışmanın farklı araştırmacılar tarafından yeniden yapıldığına dair açıklayıcı bilgi çalışmaların içeriğinde ve alan yazın bölümlerinde bulunmamaktadır. Araştırmacılar yanlış zamanda, yanlış yöntemle ve yanlış çıkış noktaları ile yapılan program değerlendirme çalışmalarının eğitim reformlarının hızını keseceği ve yönünü saptıracağı konusunda uyarmaktadırlar (Westbury, 1970). Bu nedenle, eğitim programlarının değerlendirilme çalışmaları incelenirken okuyucular ve diğer araştırmacılar bu tür çalışmaların yöntemi, değerlendirme gerekçeleri ve zamanlaması ve dolayısıyla sonuçlarının kullanımı konusunda daha temkinli yaklaşmalıdırlar.

Sonuç olarak, bu çalışmada örnekleme dahil edilmeyen, daha detay durumunda olan veya farklı boyutlar açısından incelenen başka değerlendirme çalışmalarının da mevcut olduğu düşünüldüğünde, program değerlendirmenin araştırmadan ayrılan iki temel farkı soru işareti yaratmaktadır. Bu farklardan biri çalışma sonunda varılan yargılarla ilgilidir. Buna göre program değerlendirmenin amacı değerlendirme sonunda program geleceği hakkında bir yargıya varmaktır; ancak bulgulardan da görüldüğü üzere çalışmalar daha çok araştırma boyutunda kalarak yargıda bulunmak yerine çıkan sonuçların paylaşımı yapılmıştır. Yapılan bu araştırmalar sonunda çıkan sonuçların yetkililerce değerlendirmeye alınıp alınmadığı ise ayrı bir tartışma konusudur. On yılda beş kez değerlendirmesi yapılan bir programın normal şartlar altında araştırma sonuçlarına göre düzenlenerek iyi bir seviyeye ulaşması beklenirken hala her değerlendirme sonunda büyük ölçüde düzeltmeye ihtiyaç duyduğu yönünde sonuçlara ulaşılması ise araştırmacılar tarafından yürütülen değerlendirmeye yönelik araştırmaların bulgularından yeterince yararlanılmadığını işaret etmektedir. Dolayısıyla bu da program değerlendirme ve program geliştirmede zaman, enerji ve bütçe kaybı gibi ek maliyetler olarak karşımıza çıkmaktadır. Tam da bu noktada *“Daha kaç kez aynı programı, benzer şekilde ve teorik modeller kullanmadan değerlendireceğiz?”* sorusu gündeme gelmektedir. Bu durumu en aza indirmek adına, yayınlanacak olan program değerlendirme çalışmalarında araştırma değerinin yanı sıra çalışmanın *“geliştirme değeri”* ve *“yaygın etki”*sine yönelik zorunlu bir tartışma bölümünün dergi yayın politikalarına eklenmesi dergi editörlerine politika önerisi olarak tavsiye edilebilir. Yine editörler ve hakemler tarafından dergi politikası gereği program değerlendirme çalışmalarında alan uzmanlarının yanı sıra Program Geliştirme veya Eğitim Programları ve Öğretim gibi ilgili alanlardan özellikle program değerlendirmenin yaklaşım ve yönetimine yönelik danışmanlık alınıp alınmadığının sorgulanması da program değerlendirme çalışmalarında komitelerin daha sağlıklı kurulması yönünde olumlu bir etki yaratabilir. Tüm bu öneriler

program değerlendirmede bir takım standartların oluşması yönünde küçük adımlar gibi görünürken uzun vadede yapılan çalışmaların daha bilimsel, geçerli ve güvenilir olmasını, aynı zamanda etki değerinin daha yüksek olmasını ve kaynakların daha etkin kullanılmasını sağlayabilir.

Kaynakça

- Alex, J. P. S. (1995). *Understanding the dynamics of stakeholder participation in evaluation research: A document study*. Master’s thesis. The University of British Columbia.
- Alkin, M. C., & Taut, S. M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29, 1 - 12.
- Azzam, T. (2011). Evaluator characteristics and methodological choice. *American Journal of Evaluation*, 32(3), 376-391.
- Brandon, P. R. (1998). Stakeholder participation for the purpose of helping ensure evaluation validity: Bridging the gap between collaborative and non-collaborative evaluations. *American Journal of Evaluation*, 19(3), 325-337.
- Brinkerhoff, R. O., Brethower, D. M., Hluchyj, T., & Nowakowski, J. R. (1983). *Program evaluation: A practitioner’s guide for trainers and educators*. Boston: Kluwer-Nijhoff Publishing.
- Clewell, B. C., Cohen, C. C., Campbell, P. B., Deterding, N., Manes, S., Tsui, L., Rao, S. N. S., Branting, B., Hoey, L., & Carson, R. (2004). *Review of evaluation studies of mathematics and science curricula and professional development models*. Retrieved August 12, 2014, from <http://www.urban.org/UploadedPDF/411149.pdf>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. New York, NY: Routledge.
- Creswell, J. W. & Miller, D. L. (2010). Determining validity in qualitative inquiry. *Theory Into Practice*, 39(3), 1 Edebi Sanatlar-Eda LALE24-130.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R.D., Hornik, R. C., Phillips, D. C. ... , Weiner, S. S. (1980). *Toward reform of program evaluation*. San Francisco, CA: Jossey-Bass.
- Demirel, Ö. (2013). *Eğitimde program geliştirme*. Ankara: PegemA.
- Denzin, N. K. (1978). *The research act: A theoretical orientation to sociological methods* (2nd ed.). New York: McGraw-Hill.
- Eisner, E. W. (1979). *The educational imagination: On the design and evaluation of school programs*. New York: Macmillan.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines*. NY: Pearson Education.
- Fraenkel, J. R., & Wallen, N. (2005). *How to design and evaluate research in education*. NY: McGraw-Hill.
- Glatthorn, A. A. (1987). *Curriculum leadership*. New York: HarperCollins.
- Glatthorn, A. A., Boschee, F., & Whitehead, B. M. (2009). *Curriculum leadership: Strategies for development and implementation*. US: Sage.
- Gredler, M. E. (1996). *Program evaluation*. NJ: Pearson Education.
- Greene, J. G. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review*, 12(2), 91-116.
- Guba, E., & Lincoln, Y. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.

- Hage, D. D. (2013). *A review of the levels of stakeholder involvement in program evaluation and the effects of the level of involvement on school culture assessment and improvements*. Yayınlanmamış doktora tezi. Marshall University, ABD.
- Kaplan, A. (1964). *The conduct of inquiry*. San Francisco, CA: Chandler Press.
- LeCompte, M. D. (2010). Analyzing qualitative data. *Theory Into Practice*, 39(3), 146-154.
- Madaus, G. F., & Kellaghan, T. (2000). Models, metaphors, and definitions in evaluation. In D. L. Stufflebeam, G. F. Madaus ve T. Kellaghan (Eds.), *Evaluation models viewpoints on educational and human services evaluation* (2nd ed.). Dordrecht: Kluwer Academic Publishers Group.
- Miles, M., & Huberman, M. (1994). *Qualitative Data Analysis*. London: Sage.
- ODTÜ (2014). *Sosyal bilimler ulusal A, B ve B grubu aday dergi listesi*. 24.03.2014 tarihinde <http://old.fedu.metu.edu.tr/web/documents/other/anasayfa/AveB.htm> adresinden alınmıştır.
- Oliva, P. F. (1997). *Developing the curriculum*. New York: Longman.
- Özdemir, M. S. (2009). Eğitimde program değerlendirme ve Türkiye’de eğitim programlarını değerlendirme çalışmalarının incelenmesi. *Yüzüncü Yıl Üniversitesi, Eğitim Fakültesi Dergisi*, 5(2), 126-149.
- Patton, M. Q. (1987). Evaluation’s political inherency: Practical implications for design and use. In D. Palumbo (Ed.), *The politics of program evaluation* (pp. 100-145). Beverly Hills, CA: Sage.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2008) *Qualitative research and evaluation methods*. Thousand Oaks: SAGE.
- Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on english language learners. *Educational Policy*, 19(4), 572-594.
- Silverman, D. (2006). *Interpreting qualitative data*. London: SAGE.
- Stake, R. E. (1975). *Evaluating the arts in education: A responsive approach*. Columbus, OH: Bobbs-Merrill.
- Stufflebeam, D. L. (1971). *Educational evaluation and decision making*. Itasca, IL: Peacock.
- Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic evaluation*. Boston: Kluwer-Nijhoff.
- Taylor-Powell, E., Steele, S., & Douglass, M. (1996). *Planning a program evaluation*. Madison, WI: University of Wisconsin Cooperative Extension.
- Tobler, N. S., Roona, M. R., Ochshorn, P., Marshall, D. G., Streke, A. V., & Stackpole, K. M. (2000). School-based adolescent drug prevention programs: 1998 Meta-analysis. *The Journal of Primary Prevention*, 20(4), 275-336.
- Tyler, R. (1949). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.
- Upreti, G., & Koonce, D. (2010). Stakeholder utility: Perspectives on school-wide data for measurement, feedback, and evaluation. *Education and Treatment of Children*, 33(4), 497-511.
- Westbury, I. (1970). Educational evaluation. *Review of Educational Research*, 40(2), 239-260.
- Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2010). *Handbook of practical program evaluation*. San Francisco: Jossey-Bass
- Worthen, B. R. & Sanders, J. R. (1973). *Educational evaluation: Theory and practice*. Worthington, OH: Charles A. Jones.

- Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practice guidelines* (2nd ed.). New York: Longman.
- 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.
- Yıldırım, A. ve Şimşek, H. (2005). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.
- Yüksel, I. (2010). *Türkiye için program değerlendirme standartları oluşturma çalışması*. Yayımlanmamış doktora tezi, Anadolu Üniversitesi, Eskişehir.

Extended Abstract

The Wide Angle: Program Evaluation Studies in Turkey in terms of Models and Approaches

Program evaluation is not a type of formal research but it is a kind of discipline. Thus, describing the evaluation just as a research underestimates its importance. The main purpose of this study is to explore the evaluation approaches and models employed in recent curriculum evaluation studies. Thus, the following report presents the findings of a review of fifty-two studies evaluating different types of curricula in Turkey in the last decade. To reach this purpose following research questions guide the study:

1. Which evaluation approaches and models have been mostly used in the curriculum evaluation studies in the last ten years in Turkey?
2. Are the curriculum evaluation studies research-oriented or evaluation-oriented studies in the last ten years in Turkey?
3. To what extent have stakeholders and learner needs been involved in evaluation studies in the last ten years in Turkey?
4. At which department and disciplines have the program evaluation studies been conducted in the last ten years in Turkey?

Descriptive analysis was employed in the present study to review the Turkey-addressed program evaluation studies in the last ten years. A criteria set was developed to guide the selection of evaluation studies to be included in our review. Studies were expected to (1) evaluate the big picture of any curriculum in Turkey (2) be published between 2004 and 2014 (3) be either a journal paper from the high ranking Turkish journals from the national list -developed by Middle East Technical University- of groups A and B journals (4) or be a Ph.D. dissertation. Therefore, 34 Ph.D. dissertations, and 18 journal papers composed of the sample.

47 studies reported in this study did not employ evaluation approaches and models. Only five of the 52 curriculum evaluation studies used models which are Stufflebeam’s CIPP model ($n=3$), Stake’s Countenance model ($n=1$), Provus discrepancy model ($n=1$), Hammond’s evaluation model ($n=1$), and Bellon and Handler’s curriculum evaluation model ($n=1$).

Regarding the orientations of evaluation studies, it was observed that almost all of the studies were research-oriented, that is, these studies were carried out as a dissertation or a research to be published as a journal paper. Only two of the studies stated to be granted by EARGED as a curriculum evaluation project. However, it is not clear how other studies even those wide-ranging ones could reach such big sample sizes, and whether they were granted by universities or ministry was not stated in the publications. Thus, it is not possible to conclude that whether these 50 studies were carried out as a part of curriculum development or evaluation project.

Researcher has also distilled the data sources and data collection methods of program evaluation studies. The results indicate that most of the studies (40.43%) obtained their data from both teachers and

learners in the program. Furthermore, 25.53% of the studies only included teachers as a stakeholder in the evaluation process. 14.89% of the evaluation studies collected their data from multiple sources like teachers, learners, curriculum developers, subject area specialist, ministry workers, and inspectors. The rest of the studies (9.43%) employed only document analysis, and not included opinions and evaluations of multiple stakeholders. Finally, in this study, it is also shown that most of the curriculum evaluation studies ($n=34$) were carried out at different departments or disciplines from Educational Sciences and Curriculum and Instruction departments.

In this study, the discussions of the results and general conclusions are stated. During the paper, validity and reliability of the studies, reasons of not employing evaluation models, and the possible conclusions of this type of studies are discussed. In sum, empirical evidence considered here indicates that most of the evaluation studies seem to be carried out by researchers' publication motivation. It seems clear from the current study that these studies have not employed program evaluation models which are generally used for not only identification of problems but also for decision making about program future. As has been mentioned, how the results of these studies affect the related curricula are not discussed at the publications. Thus, whether the results of these curriculum evaluation studies are noticed by the authorities and referenced by authorities in the curriculum development process or not is not clear. Since the millions of Turkish Liras have been flowed to evaluate programs to accomplish variety of national goals, it is important to benefit from the results of recent research. As a policy recommendation to journal editors and reviewers, researchers should be subjected to point the necessity of adding the compulsory "significance" and "common effect" discussion at the publications of program evaluation studies.