

ارزیابی سیاست؛ نابسامانی مفهومی؟

راضیه امامی میبیدی^۱

دکترای سیاستگذاری عمومی دانشگاه تهران

(تاریخ دریافت: ۹۱/۴/۵ - تاریخ تصویب: ۹۱/۶/۲۳)

چکیده

در این مقاله نشان خواهیم داد در ورای آشفتگی مفهومی مطالعات ارزیابی، جریان پیوسته و مرتبطی از دانش وجود دارد. این دانش، چندرشته‌ای بوده و همسو با تکامل مباحث معرفت‌شناسی و روش‌شناسی در علوم اجتماعی رشد کرده و فراز و نشیب‌های آن را پیموده است. تحولات معرفت‌شناسانه خود را در تعریف ارزیابی منعکس کرده و به ظهور مدل‌های تازه منجر شده‌اند. بدینسان ارزیابی قبل از آنکه بر مفهوم تخصصی مشخصی دلالت کند یک روش بوده و به مرور زمان برای انجام هر چه بهتر و پاسخ به نیازهای ذی‌نفعان توسط پژوهشگرانی از رشته‌های مختلف تکامل یافته است. اهمیت ثانویه بُعد مفهومی آن، موجب شده ادبیات ارزیابی بیش از هر چیز با قرائت‌های متنوع از مفاهیم پایه مواجه باشد لذا انواع مفاهیمی که در ارزیابی مطرح می‌شوند در چارچوب مفهومی پژوهشگر معنا می‌یابند و قراردادی هستند.

واژگان کلیدی: ارزیابی سیاست، مدل‌های ارزیابی، روش‌های ارزیابی، ادبیات ارزیابی، مفاهیم ارزیابی.

مقدمه

به‌رغم رشد فزاینده تحقیقات ارزیابی از دهه ۱۹۶۰، یک سنت مطالعاتی قوی درباره ارزیابی سیاست در ایران شکل نگرفته است. مباحث نظری و فنی ارزیابی عمدتاً به یک فصل یا کمتر در میان آثار ترجمه‌ای و تألیفی محدود شده‌اند (مولر، ۱۳۷۸؛ هاوالت و رامش، ۱۳۸۰؛ لستر و استوارت، ۱۳۸۱؛ وحید، ۱۳۸۳؛ الوانی، ۱۳۸۵؛ پارسونز، ۱۳۸۵؛ مانهایم و ریچ، ۱۳۸۵؛ اشتریان، ۱۳۹۰؛ الوانی و شریف‌زاده، ۱۳۹۰؛ شافریتز و بریک، ۱۳۹۰) این مطالعات که جنبه معرفی سیاستگذاری عمومی را دارند، ارزیابی سیاست را به عنوان یکی از مباحث سیاستگذاری عمومی در کنار سایر مباحث آن مورد توجه قرار داده‌اند و در نتیجه از یک معرفی مختصر فراتر نمی‌روند. اختصار و قلت این آثار موجب شده ارزیابی سیاست در ایران حوزه شناخته شده‌ای نباشد. لذا برخی، آن‌را با ارزیابی عملکرد سازمان‌ها یکسان می‌دانند. برخی، درباره معنای «سیاست» ابهام دارند و گروهی دیگر معتقدند میان «ارزیابی» و «ارزشیابی» تفاوت‌های بنیادی وجود دارد. برخی بر تفکیک ارزیابی پروژه و برنامه با ارزیابی سیاست اصرار می‌ورزند و عده‌ای دیگر هر نوع تحقیق و بررسی را شکلی از ارزیابی می‌پندارند.^۱ ابهامات موجود در کنار کمبود آثار تحلیلی-تحقیقی در این زمینه، موجب شده مطالعات ارزیابی در کشور توسعه نیابد و مفاهیم و ابزارهای آن مبهم باقی بماند. این در حالی است که ارزیابی سیاست نقش مهمی در یادگیری، پاسخگویی و تفکر بومی دارد و تقویت آن، نه تنها به بهبود سیاستگذاری و اجرا کمک می‌کند بلکه به رشد پژوهش‌های مسئله محور و کاربردی نیز منجر می‌شود.

با توجه به کمبود مطالعات ارزیابی سیاست در کشور، این مقاله با هدف ابهام‌زدایی از این مطالعات تنظیم شده است. این امر در چهار سطح تعریف ارزیابی^۲، مفاهیم پایه ارزیابی، معیارهای ارزیابی، مدل‌ها و روش‌های ارزیابی دنبال می‌شود. بدین‌سان، ضمن مرور مطالعات ارزیابی نشان داده می‌شود به‌رغم آشفتگی مفهومی، ارزیابی سیاست جریان مطالعاتی پیوسته و منظمی دارد و حول چگونگی طراحی و اجرای ارزیابی پرورش می‌یابد. این ادبیات با ارزیابی برنامه^۳ مشترک بوده و به مثابه یک فن توسط پژوهشگران مختلف توسعه یافته است.

۱- نگارنده حین اجرای پژوهش دو ساله‌ای درباره ارزیابی سیاست به طور گسترده‌ای با چنین اظهارنظرهایی در محیط‌های دانشگاهی و اجرایی مواجه بوده است.

۲- این مقاله معطوف به ارزیابی طرح‌ها، برنامه‌ها و سیاست‌ها است و به سایر انواع ارزیابی نمی‌پردازد.

3- Program Evaluation

مفاهیم پایه

گرچه قراردادی و نسبی بودن تعریف مفاهیم در علوم اجتماعی رایج است، اما این نسبت در ارزیابی تا آشفتگی مفهومی پیش رفته است. مطالعات ارزیابی سرشار از وضع اسامی متعدد برای مفاهیم کم و بیش یکسان است. این امر، حتی در محوری‌ترین مفهوم این حوزه نیز مشاهده می‌شود. گرچه ارزیابی عمدتاً در برابر Evaluation به کار می‌رود اما استفاده از Assessment به عنوان مترادف آن رواج دارد. کاربرد اصطلاحات مترادف در کشور ما در قالب مفاهیم «ارزیابی» و «ارزشیابی» متداول است. برخی آن‌ها را دو اصطلاح کاملاً متفاوت می‌پندارند درحالی‌که آن‌ها خواه در تعاریف علمی (Rossi, Freeman and Lipsey, 1999) خواه در تعاریف لغوی (لغت‌نامه معین) با یکدیگر تفاوتی ندارند.

نمونه مهم دیگر، سردرگمی در کاربرد ارزیابی پروژه، برنامه و سیاست است. گرچه در مطالعات ارزیابی، معمولاً اصطلاح «ارزیابی برنامه» به کار گرفته می‌شود اما در کنار آن از «ارزیابی پروژه»^۱ و «ارزیابی سیاست»^۲ نیز استفاده می‌شود. ممکن است آن‌ها، از نظر دامنه ارزیابی با یکدیگر تفاوت داشته باشند اما از روش‌ها و مدل‌های مشابهی استفاده می‌کنند.

اصولاً تمایز میان سیاست، برنامه و پروژه چندان شفاف نیست. این امر یک ویژگی ایرانی نیست و عمومیت دارد. لانگبین و فلبینگر با ذکر مثال‌هایی نشان می‌دهند آنچه برای یک فرد سیاست می‌نماید، ممکن است برای دیگری برنامه یا پروژه به نظر برسد و لذا دیدگاه و موقعیت سازمانی افراد در این زمینه تأثیر دارد (Langbein and Felbinger, 2006: 4). البته باید تأکید کرد تفکیک آن‌ها برای نظم و انسجام سیاستگذاری ضرورت دارد، اما این یک تفکیک دامنه‌ای است و نه محتوایی.

اگر «سیاست» را صرفاً به معنای کلان آن در نظر بگیریم، بر قواعد عامی دلالت می‌کند که برنامه‌ها یا پروژه‌های معینی را شکل می‌دهند. غالباً هر «سیاست» برای اجرا شامل چند «برنامه» و هر «برنامه» برای اجرا شامل چند «پروژه» می‌شود. بنابراین، ارزیابی سیاست ترکیبی و تجمیعی است و از طریق ارزیابی برنامه‌ها و پروژه‌ها انجام می‌شود (Langbein and Felbinger, 2006: 3).

ارزیابی پروژه در ایران ذیل مباحث مدیریت پروژه شناخته شده است. در ارزیابی برنامه، مجموعه‌ای از پروژه‌های مشابه که در یک برنامه یا خوشه طبقه‌بندی شده‌اند ارزیابی می‌شود. ارزیابی برنامه، به شناسایی رگه‌ها و زمینه‌های مشترک میان پروژه‌ها می‌پردازد و بازخورد برنامه و روش‌های تازه به کار گرفته شده در پروژه‌ها را بر مردم تعیین می‌کند. ارزیابی برنامه

1- Project Evaluation

2- Policy Evaluation

نشان می‌دهد تا چه حدی مجموعه پروژه‌های تعریف شده در یک برنامه یا خوشه اهداف تغییر سیستمی یک ابتکار سیاستی برنامه‌ریزی شده را تحقق بخشیده‌اند. مهم‌ترین جنبه ارزیابی برنامه، ایجاد یادگیری متقابل میان پروژه‌های مختلف مرتبط با یکدیگر است تا از این طریق از تجارب یکدیگر بیاموزند و به پیشرفت برنامه کمک کنند (W. K. Kellogg Foundation, 1998: 17).

تعریف ارزیابی

هاولت و رامش با اشاره به اینکه ارزیابی سیاست فراتر از آن چیزی است که اغلب در متون از آن یاد می‌شود، تعاریف تخصصی و روزمره ارزیابی را با یکدیگر ترکیب کرده‌اند. آن‌ها ارزیابی سیاست را به سه مقوله ارزیابی اداری، ارزیابی قضایی و ارزیابی سیاسی تقسیم می‌کنند. ارزیابی اداری، جریان مسلط ادبیات ارزیابی است که به مسائلی چون کارایی، اثربخشی، فرایند و عملکرد سیاست می‌پردازد. ارزیابی قضایی به انطباق سیاست‌ها با قوانین و مقررات عام و در رأس آن قانون اساسی توجه دارد. این امر از طریق بازرسی و رسیدگی به شکایات انجام می‌شود. ارزیابی سیاسی از قضاوت شهروندان درباره سیاست‌ها تا مشورت با اعضای زیرنظام سیاسی درباره مزایا و معایب آن نوسان دارد (هاولت و رامش، ۱۳۸۰: ۲۸۰-۲۷۳). بدینسان، در یک قرائت موسع مفهوم ارزیابی دامنه وسیعی از سنجش، قضاوت، کنترل، نظارت و انطباق را در بر می‌گیرد.

پژوهشگرانی که درباره ارزیابی سیاست قلم زده‌اند به ارزیابی اداری توجه داشته‌اند. حتی تحولی که در نسل چهارم ارزیابی صورت گرفته و آن را از سیطره اثبات‌گرایی خارج کرده با هدف رفع کاستی‌ها و بازسازی ارزیابی اداری صورت گرفته است. در خلال تکامل این مطالعات، مفهوم ارزیابی به گونه‌های مختلفی تعریف شده که از ارزیابی به عنوان مجموعه‌ای از روش‌های تحقیق (Rossi, Freeman and Lipsey, 1999) تا ارزیابی به مثابه فرایندی اجتماعی (Braybrooke and Lindblom, 1963) را در بر می‌گیرد. در بعضی از تعاریف، ارزیابی به مفاهیم سنجش و پایش نزدیک می‌شود. گرچه بسیاری آن را شبه‌ارزیابی می‌خوانند و معتقدند «ارزش» عنصر اصلی ارزیابی است (Dunn, 1994: 403). در تعاریف مشهور دیگری، ارزیابی قضاوتی سیستماتیک (نظام‌مند) پنداشته می‌شود (Mark, Greene and Shaw, 2006: 6). با وجود تنوع بسیار، تعاریف ویژگی‌های مشترکی دارند که در قرائت‌های مختلف از ارزیابی ملاحظه می‌شود:

- ارزیابی فرایندی نظام‌مند است یعنی فعالیتی برنامه‌ریزی شده و هدفمند است؛

- ارزیابی شامل جمع‌آوری داده‌هایی درباره پرسش‌ها یا مسائلی راجع به جامعه به طور اعم و سیاست‌ها و برنامه‌ها به طور اخص می‌شود؛
- ارزیابی فرایندی برای تقویت دانش و تصمیم‌گیری است، خواه این تصمیمات برای بهبود یا پالایش یک برنامه یا سیاست باشد یا برای تداوم و گستراندن آن، و در هر کدام از این تصمیمات ابعادی از قضاوت درباره شایستگی، ارزش و بهای موضوع تحت ارزیابی مشاهده می‌شود؛
- مفهوم کاربرد ارزیابی خواه تلویحی یا صریح در بیشتر تعاریف وجود دارد؛
- ارزیابی مربوط به سوالاتی درباره جریان واقعی امور می‌شود (Preskill and Russ-Eft, 2005: 1-2).

ویژگی‌های فوق نشان می‌دهند به رغم تنوع تعاریف، دست کم در جریان مسلط مطالعات ارزیابی ابعاد مشترکی وجود دارد. در واقع، آشفتگی مفهومی ارزیابی از ترکیب کاربرد روزمره و عام آن با کاربرد علمی- تخصصی ناشی می‌شود. بنابراین، تحقیقات ارزیابی سیاست سیستماتیک، تجربی، هنجاری و کاربردی هستند. تحقیق ارزیابی معمولاً معطوف به گذشته است و حین یا پس از اجرا انجام می‌شود. آنچه به عنوان ارزیابی پیش از اجرا صورت می‌گیرد در چارچوب پیش‌بینی و تصمیم‌گیری انجام می‌شود و گرچه از مفاهیم، روش‌ها و تکنیک‌های مشابه استفاده می‌کند، اما جایگاه متمایزی در چرخه سیاستی دارد.

معیارهای ارزیابی

معیارهای ارزیابی محور قضاوت درباره موفقیت یا ناکامی سیاست‌ها است، لذا تعیین آن‌ها یکی از چالش برانگیزترین مراحل ارزیابی به شمار می‌رود. این مسئله که معیار موفقیت یک سیاست دستیابی آن به اهداف اعلامی باشد یا صرفه‌جویی در هزینه‌ها یا توزیع عادلانه سود و زیان آن میان مردم وغیره، مناقشات زیادی برمی‌انگیزد و نظریه‌های مختلف پاسخ‌های مختلفی به آن می‌دهند. ارزیاب گاه به تنهایی و گاه در مشارکت با ذی‌نفعان معیارهای ارزیابی را معین می‌کند. گاهی نیز معیارها از قبل تعیین شده و ارزیابی الزاماً حول آن انجام می‌شود. معیارهای ارزیابی سیاست قراردادی است اما عمدتاً پیرامون اثربخشی^۱ و کارایی^۲ انجام می‌شود.

از مشهورترین معیارهای ارزیابی، می‌توان به معیارهای پنج‌گانه‌ای اشاره کرد که در نهادهای وابسته به سازمان ملل استفاده می‌شود. آن‌ها عبارتند از:

1- Effectiveness
2- Efficiency

- اثربخشی: تحقق یا احتمال تحقق اهداف یک مداخله معین؛
- کارایی: اندازه‌گیری خروجی یک مداخله معین با توجه به هزینه‌های آن؛
- تأثیر ۱: نتایج و آثار معلول یا قابل نسبت‌دهی به یک مداخله معین؛
- دوام‌پذیری ۲: قابلیت مالی، نهادی و محیطی تداوم خدمات و مزایای یک مداخله معین؛
- ارتباط ۳: ارتباط دائمی رویکرد و اهداف مداخله با مشکلات، سیاست‌ها و اولویت‌های متغیر (-UNDP, 2009: 168; Internal Oversight Service, 2007: 8; OECD, 2010: 170).

گرچه معیارهای فوق به شکل گسترده مورد استفاده قرار می‌گیرند اما تنها معیارهای ارزیابی نیستند. مثلاً در آن‌ها به معیارهای برابری^۴ به معنای توزیع برابر هزینه‌ها و مزایای مداخله میان گروه‌های مختلف و پاسخگویی^۵ به معنای رفع نیازها، ترجیحات یا ارزش‌های گروه‌های معین (Dunn, 1994: 405) پرداخته نشده است. تعیین معیارهای ارزیابی را می‌توان سیاسی‌ترین قسمت ارزیابی و محل برخورد منافع مختلف خواند. آنچه در نظر یک گروه سیاسی اهمیت دارد می‌تواند از منظر گروه دیگری ناچیز قلمداد شود. در ارزیابی‌های دولتی، ارزش‌ها و منافع تصمیم‌گیران، نوع نیازها و مشکلات آنان و شرایط و مقتضیات سیاسی، اجتماعی و اقتصادی بیشترین تأثیر را بر تعیین معیار ارزیابی می‌گذارد. افزایش شمول ذی‌نفعان در نسل چهارم ارزیابی با هدف انعکاس صداهای بیشتر در تعیین معیارهای ارزیابی و قضاوت درباره موفقیت یا شکست آن صورت می‌گیرد.

مدل‌ها و روش‌های ارزیابی

طراحی و توسعه مدل‌های ارزیابی، محور اصلی مطالعات ارزیابی سیاست‌ها را تشکیل می‌دهد. این مدل‌ها معمولاً روش طراحی و اجرای ارزیابی را مشخص کرده و مشخصات یک ارزیابی خوب را بر می‌شمارند. برخی آن‌ها را «نظریه» می‌نامند در حالی که آن‌ها کاملاً ویژگی‌های یک نظریه را ندارند (Alkin, 2004: 5) برخی دیگر نیز از «رهیافت» استفاده می‌کنند تا قالب وسیع‌تری نسبت به مدل برای بیان ایده‌های ارزیابی داشته باشند (Stufflebeam, 2001: 9).

1- Impact

2- Sustainability

3- Relevance

4- Equity

5- Responsiveness

مدل‌های ارزیابی به اشکال مختلف، نام‌گذاری و طبقه‌بندی شده‌اند. این امر موجب آشفتگی مفهومی در ارزیابی شده و شناخت مدل‌های متنوع آن را دشوار کرده است. همان‌گونه که بالارت می‌گوید تنوع عظیم ارزیابی، عملاً ناشی از پاسخ به تقاضاها و موقعیت‌های گوناگون برای انجام آن بوده است. فعالیت‌های ذیل آن نیز به خاطر موقعیت حرفه‌ای مختلف ارزیابان، شرکت‌های خصوصی، اساتید دانشگاه، بوروکرات‌ها و ابزارهای تحلیلی متنوعی که آن‌ها بر-اساس زمینه اصلی آموزشی‌شان به کار می‌گیرند انواع متعددی دارد (Ballart, 1998: 151). مرور برخی از قرائت‌های ارزیابی این گوناگونی را به خوبی منعکس می‌کند:

در یکی از مشهورترین آثار ارزیابی، گوبا و لینکلن از چهار «نسل» کمی، کیفی، قضاوتی و سازه‌گرای ارزیابی نام می‌برند (Guba and Lincoln, 1989). سیمونز، ارزیابی را به سه دسته بوروکراتیک (معطوف به گسترش قدرت مدیریتی)، اتوکراتیک (معطوف به تداوم مشروعیت علمی) و دموکراتیک (معطوف به خدمت به اجتماع) تقسیم می‌کند. پاول در ارزیابی سیاست دو نوع ذاتی ۱ و عرضی ۲ را از یکدیگر متمایز می‌کند. مورتاگ، ارزیابی را در دو طبقه کلی ابزاری ۳ و تفسیری ۴ قرار می‌دهد (Taylor, 2006). شادیش، کوک و لویتن با مرور تاریخچه ارزیابی، نظریه‌های آن را به سه مرحله تقسیم می‌کنند. مرحله اول بر حقیقت‌جویی و مرحله دوم بر کاربرد و عمل‌گرایی تأکید دارد. مرحله سوم آمیزه‌ای از نظریه‌های مراحل قبلی است و به یکپارچگی آن‌ها توجه دارد (Shadish, Cook and Leviton, 1991). روسی، فریمن و لیپسی از اصطلاح چارچوب‌های مفهومی و روش‌شناختی استفاده کرده‌اند. آن‌ها این چارچوب‌ها را به پنج دسته ارزیابی نیازها، ۵، ارزیابی نظریه برنامه، ۶، ارزیابی فرایند برنامه، ۷، ارزیابی تأثیر (یا ارزیابی نتیجه) ۸ و ارزیابی کارایی ۹ تقسیم می‌کنند. دان رهیافت‌های ارزیابی سیاست‌ها را به سه دسته شبه‌ارزیابی ۱۰، ارزیابی رسمی ۱۱ و ارزیابی تصمیم-تئوریک ۱۲ تقسیم می‌کند (Dunn, 1994: 405). وی در ویرایش‌های بعدی کتابش، در عناوین دسته‌بندی رهیافت‌های ارزیابی تجدیدنظر کرده و آن‌ها را به علی ۱۳، اداری ۱۴ و مشارکتی ۱۵ تغییر داده اما در ویرایش سال

-
- 1- Intrinsic
 - 2- Extrinsic
 - 3- Instrumental
 - 4- Interpretive
 - 5- Needs Assessment
 - 6- Assessment of Program Theory
 - 7- Assessment of Program Process
 - 8- Impact Assessment (Outcome Evaluation)
 - 9- Efficiency Assessment
 - 10- Pseudo-evaluation
 - 11- Formal Evaluation
 - 12- Decision-theoretic Evaluation
 - 13- Casual
 - 14- Official
 - 15- Participative

۲۰۱۲ مجدداً به دسته‌بندی قبلی بازگشت. استافلیم در تکراری ارزیابی بر پایه چگونگی طراحی و اجرای آن، ۲۲ «رهیافت» مختلف برای ارزیابی شناسایی کرده است. وی رهیافت‌های مذکور را در قالب چهار طبقه عمده ارزیابی ساختگی ۱، رهیافت‌های پرسش و روش‌گرا ۲، رهیافت‌های متمایل به اصلاح و پاسخگویی ۳ و رهیافت‌های طرفدارانه و معطوف به دستورکار اجتماعی ۴ قرار می‌دهد (Stufflebeam, 2001). هانسن، مدل‌های ارزیابی را در شش طبقه قرار می‌دهد: مدل‌های نتیجه‌ای ۵، مدل فرایندی ۶، مدل سیستمی ۷، مدل اقتصادی ۸، مدل کنشگر ۹ و مدل نظریه برنامه ۱۰ (Hansen, 2005). این لیست را می‌توان تا ده‌ها دسته‌بندی دیگر ادامه داد درحالی‌که حتی پیرامون مدل، رهیافت یا روش بودن آن‌ها اختلاف نظر بوده و هر نویسنده‌ای قرائت خاص خود را از آن دارد.

تنوع مدل‌های ارزیابی با تمرکز بر جزئیات، دوجندان می‌شود. مثلاً هانبرگر، سه نوع «رهیافت» دموکراتیک ارزیابی را از یکدیگر متمایز می‌کند: ارزیابی دموکراتیک مشارکتی ۱۱، ارزیابی دموکراتیک نخبه‌گرا ۱۲ و ارزیابی دموکراتیک نامنظم ۱۳ (Hanberger, 2006).

برای فهم مدل‌های ارزیابی، باید توجه داشت که اسامی متنوع آن‌ها ضرورتاً به معنای تفاوت در محتوای آن‌ها نیست. تفاوت در نام‌ها بیش از آنکه تفاوت در ماهیت را نشان دهد، قرائت‌های متنوع محققان عرصه ارزیابی را منعکس می‌کند.

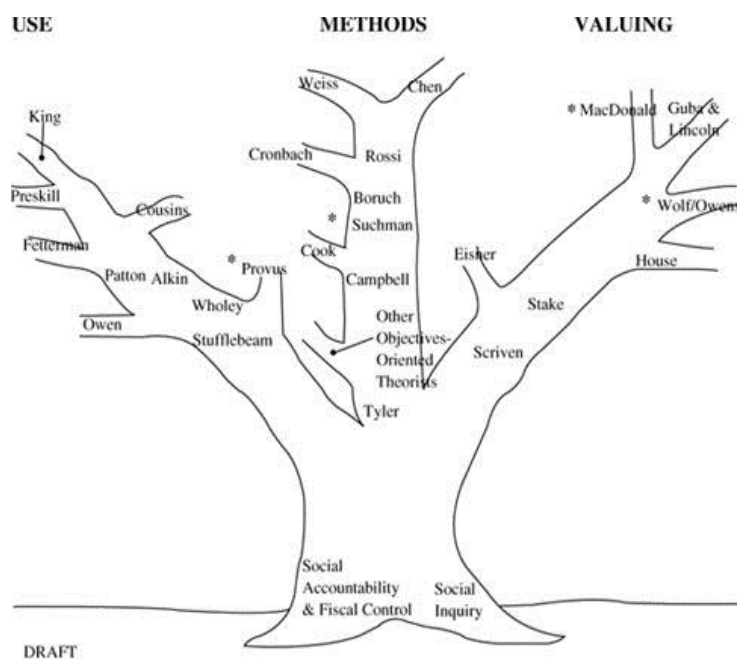
از دسته‌بندی آلکین و کریستی درباره محققان حوزه ارزیابی، می‌توان به عنوان الگویی برای دنبال کردن مطالعات ارزیابی استفاده کرد:

- کسانی که سهمی بسیار بنیادی در روش‌شناسی تحقیق اولیه که شالوده اساسی بیشتر کارهای ارزیابی را شکل می‌دهد دارند. این دسته، روش‌شناسانی مانند کمپل هستند که بر ارزیابی نفوذ داشته‌اند؛
- کسانی که لزوماً مدل ارزیابی تازه‌ای مطرح نکرده‌اند، اما کار آن‌ها به درک جنبه‌های متنوع ارزیابی کمک بزرگی کرده است. این دسته تحلیلگران ارزیابی مانند چلیمسکی و شادیش هستند؛

-
- 1- Pseudoevaluations
 - 2- Questions/Methods-oriented Approaches
 - 3- Improvement/Accountability-oriented Approaches
 - 4- Social Agenda/Advocacy Approaches
 - 5- Result Models
 - 6- Process Models
 - 7- System Models
 - 8- Economic Models
 - 9- Actor Models
 - 10- Programme Theory Models
 - 11- Participatory democratic evaluation
 - 12- Elitist democratic evaluation
 - 13- Discursive democratic evaluation

- کسانی که به نگارش کتاب‌های درسی ارزیابی پرداخته و به تفسیر اختلافات ظریف آن کمک کرده‌اند. این دسته معلمان و مفسران ارزیابی مانند ساندرز و شادیش هستند که برخی از آن‌ها در گروه تحلیلگران زبده ارزیابی نیز قرار می‌گیرند؛
- کسانی که جهت‌گیری خاصی در ارزیابی مطرح کرده‌اند هرچند که تفسیر نظری کاملی ارائه نداده باشند. این دسته نظریه‌پردازان ارزیابی مانند استافلیم و گوبا و لینکلن هستند (Alkin and Christie in Alkin, 2004: 5-6).

با وجود چندرشته‌ای بودن ارزیابی، تنوع مدل‌ها و نیز وضع اسامی متعدد برای آن‌ها، مطالعات ارزیابی جریان پیوسته‌ای دارد. آلکین با نظریه خواندن مدل‌های ارزیابی استدلال می‌کند بین نظریه‌پردازان ارزیابی ارتباط وجود دارد، یعنی نظریه‌ها بر پایه نظریه‌های دیگر ساخته شده‌اند و همه ریشه در روش‌شناسی دو رشته مسئولیت‌پذیری اجتماعی و بررسی‌های اجتماعی دارند. وی معتقد است مشاهده مسیری که ریشه‌های دوگانه بررسی‌های اجتماعی و پاسخگویی اجتماعی در طول زمان به شکل درختی با سه شاخه روش‌ها، ارزش‌گذاری و کاربرد در آمده به درک بهتر نظریه ارزیابی کمک می‌کند. در این درخت، شماری از مشهورترین «نظریه‌پردازان» ارزیابی بر اساس تاکید‌ی که بر کاربرد، روش‌ها و ارزش‌گذاری داشته‌اند بر روی شاخه‌های سه‌گانه جایگذاری شده‌اند (شکل ۱). بدین سان، موقعیت هرکدام از آن‌ها در یکی از شاخه‌ها و همچنین، ارتباط وی با سایر نظریه‌پردازان در گرایش مشابه و سایر گرایش‌ها نشان داده می‌شود (Alkin, 2004: 12-385).



* Not included in chapters

شکل ۱: درخت ارزیابی - منبع: Alkin and Christie in Alkin, 2004: 389

همچنان که در شکل ۱ ملاحظه می‌شود، شاخه میانی درخت ارزیابی، امتداد تنه بررسی‌های اجتماعی است. آلکین و کریستی این شاخه را روش‌ها می‌نامند زیرا نظریه‌هایی که بر روش تاکید دارند به تعمیم‌پذیری و خلق دانش اهمیت می‌دهند. آن‌ها ارزیابی را به مثابه تحقیق می‌بینند و رویکرد مسلط آن‌ها استفاده از روش‌های تحقیق برای انجام ارزیابی است. شاخه دیگر ارزش‌گذاری است که اثر اسکریون الهام‌بخش آن است. نظریه‌هایی که بر ارزش‌گذاری تاکید دارند تفاوت ارزیابی با سایر تحقیقات را در قضاوت و بعد ارزشی آن می‌دانند. آن‌ها نقش اصلی ارزیاب را ارزش‌گذاری می‌دانند. برخی از نظریه‌پردازان این شاخه، نقش ارزیاب را به تسهیل‌گری ارزش‌گذاری توسط دیگران نیز گسترش داده‌اند. شاخه سوم کاربرد است که آثار استفالیم و هولی، پیش‌تاز آن است. نظریه‌هایی که بر کاربرد تاکید دارند به استفاده‌کنندگان از ارزیابی و اطلاعات موردنیاز آن‌ها برای تصمیم‌گیری توجه دارند. رشته‌های مختلفی این شاخه‌ها را پرورش داده‌اند. روان‌شناسی و جامعه‌شناسی کمی در شاخه روش‌ها نفوذ داشته‌اند. تاثیر انسان‌شناسی در آثار برخی از ارزیابی‌پژوهان مانند استیک و آیزنر مشهود است و توجه به دغدغه‌های برخاسته از فلسفه علم خود را در آثار اسکریون نشان می‌دهد (Alkin, 2004: 13).

درخت ارزیابی به خوبی نشان می‌دهد مدل‌های ارزیابی بر پایه تعریف ارزیابی شکل گرفته و تکامل یافته‌اند. معمولاً در نظریه پردازی، هسته اصلی نظریه حفظ می‌شود و رویکردهای تازه به جرح و تعدیل آن می‌پردازند، درحالی‌که در ارزیابی، تحولات در هسته اصلی یعنی تعریف رخ می‌دهد.

تحول در تعریف ارزیابی متأثر از ارزش‌های ارزیابی‌پژوهان و نگاه معرفت‌شناسانه آن‌هاست. براین اساس، مدل‌های ارزیابی در چگونگی طراحی ارزیابی و ذی‌نفعان کلیدی آن، با هم اختلاف دارند. ارزش‌های متفاوت به تفاوت در هدف و کارکرد ارزیابی می‌انجامد و اختلافات معرفت‌شناسانه اعتبار یافته‌های ارزیابی را به یکی از مناقشه‌برانگیزترین مباحث ارزیابی تبدیل می‌کند.

همگام با افزایش نسبی‌گرایی در مطالعات ارزیابی، نقش و شمول ذی‌نفعان به کانون تغییر در مدل‌های ارزیابی تبدیل شده است. مدل‌های نسبی‌گرای انتقادی برای این شکل گرفته‌اند که ارزیابی از تسلط ارزیاب و سیاستگذاران خارج شود و نیازها، مطالبات و دیدگاه‌های دامنه وسیع‌تری از ذی‌نفعان را پوشش دهد. تاکید انتقادات روشنگرانه مدل‌های نسبی‌گرای انتقادی آن است که تمام صداها باید در تحقیقات ارزیابی شنیده شود. برخی از این مدل‌ها در میدان دادن به ذی‌نفعان تا جایی پیش می‌روند که در آن‌ها نقش ارزیاب به تسهیل‌گر تقلیل می‌یابد و کنترل خود را بر فرایند و یافته‌های ارزیابی از دست می‌دهد. این امر، رعایت روش‌مندی و انسجام تحقیقاتی ارزیابی را دشوار می‌کند. لذا باید توجه کرد به رغم جذابیت گشودن عرصه برای ذی‌نفعان، بایستی مدیریت ارزیاب بر فرایند ارزیابی حفظ شود، در غیر این صورت با یافته‌هایی نامنسجم مواجه خواهیم شد که صرفاً مهارت در بحث و جدل تعیین‌کننده تسلط آن‌ها بر دیدگاه‌های بدیل خواهد بود. مثلاً راهنمای عملی که گوبا و لینکلن برای انجام ارزیابی سازه‌گرا تهیه کرده‌اند بر گفتگو و مذاکره تاکید دارد که خود موجب می‌شود ذی‌نفعانی بر نتایج ارزیابی مسلط شوند که مهارت‌های جدلی قوی‌تری دارند نه اینکه لزوماً ادعای موجه‌تری دارا هستند (Guba and Lincoln, 2001).

هرچند که یکی از اصول نظریه‌تغییر، درگیرکردن افراد در فرایند تغییر برای افزایش اثربخشی مداخله است اما گاه این تلاش برای فایده‌مندی چنان از مشخصات یک تحقیق استاندارد علمی فاصله می‌گیرد که رویکرد فایزبندی در یکسان شمردن علم با سایر معرفت‌ها را تداعی می‌کند (ر.ک. چالمرز، ۱۳۸۴: ۱۷۱-۱۵۷) علاوه براین، انبوهی از تأویل‌ها، تفسیرها و روشنگری‌ها تا هنگامی که به یک نتیجه مشخص و توصیه عملیاتی منجر نشود مزیتی برای سیاستگذاری و تصمیم‌گیری برای بهبود عملکرد ندارد.

هر کدام از مدل‌های ارزیابی مستلزم به‌کارگیری مجموعه‌ای از روش‌ها و تکنیک‌های تحقیق است که از رشته‌های مختلف اقتصاد، علوم اجتماعی، علوم رفتاری و غیره گرفته شده است. آن‌ها معمولاً با نام روش‌های ارزیابی شناخته می‌شوند و در گردآوری و تحلیل داده‌ها به‌کار می‌روند. برخی از آن‌ها مانند «آزمایش» و «هزینه-فایده» به عنوان مدل ارزیابی نیز شناخته می‌شوند. بدیهی است کاربرد این روش‌ها و تکنیک‌ها در ارزیابی فرقی با وقتی که در اشکال دیگر پژوهش به‌کار می‌روند، ندارد. می‌توان آن‌ها را از طریق مراجعه به انبوه کتاب‌های روش تحقیق فراگرفت یا از کتاب‌ها و راهنماهای ارزیابی استفاده کرد (Treasury Board of Canada, 1998; Purdon et al, 2001; Government Chief Social Researcher's Office, 2003). در جدول ۱ روش‌های مورد استفاده در هر کدام از مدل‌های ارزیابی بر اساس قرائت استفالیم نشان داده شده است.^۱ در این جدول، مدل‌های کاذب ارزیابی و مدل‌هایی که تنها در حوزه آموزش، کاربرد دارند منعکس نشده است.

جدول ۱: مدل‌ها و روش‌های ارزیابی - منبع: Stufflebeam, 2001 با تصرف، اضافات و تلخیص

روش	کاربرد	ایده اصلی	پیشگام	رهیافت ارزیابی
آزمایش و کنترل تصادفی ۲. مطالعات مقایسه‌ای ۳. مدل‌سازی آماری	آگاهی از تحقق اهداف سیاست	پایش تحقق اهداف از پیش تعیین شده سیاست	تایلر	هدف‌محور
رویه‌های تنظیم استانداردهای موفقیت/ شکست ۲. پرداخت نتیجه‌گرا ۳. قرارداد عملکرد ۴. مدیریت هدف‌گرا ۵. پایگاه داده‌های ورودی، فرایند و خروجی برنامه ۶. نظام‌های برنامه‌ریزی و بودجه‌ریزی برنامه (PPBS) ۷. کارت‌ها/ پروفایل‌های	پیشگیری از اتلاف بودجه و ملزم کردن مجریان به تحقق اهداف و برنامه‌هایی که اعتبار آن‌را دریافت کرده‌اند	به‌کارگیری استانداردهای موفقیت/ شکست، پرداخت برای نتایج خوب و تنبیه برای عملکرد غیر قابل قبول	لسینگر	پاسخگویی با تاکید بر پرداخت نتیجه‌گرا

۱- تک‌نگاری استفالیم به فارسی ترجمه و توسط انتشارات یادواره کتاب منتشر شده است.

<p>گزارش نهادی ۸ حسابرسی‌های رعایت رویه‌ای و دستیابی به هدف ۹.خویش مطالعه ۱۰. بازبینی‌های همتای متمرکز بر معیار ثابت ۱۱. پرداخت شایستگی برای سازمان‌ها، برنامه‌ها و/ یا افراد</p>				
<p>۱. آزمایش ۲. شبه‌آزمایش</p>	<p>شناخت اثر یک سیاست معین</p>	<p>تعریف گروه آزمایش و کنترل و تعیین رابطه علی میان یک سیاست معین و اثر آن</p>	<p>کمپل</p>	<p>آزمایش</p>
<p>۲. خودارزیابی خودگزارشگری افراد/ نهادها ۳. بازدید از نهادها</p>	<p>رتبه‌بندی، استانداردگذاری و ارزیابی عملکرد نهادها، برنامه‌ها و کارکنانشان برای اجرای سیاست‌ها</p>	<p>بررسی صلاحیت نهادها، برنامه‌های نهادی، و/ یا کارکنان برای ارائه خدمات عمومی مشخص</p>	<p>هیأت‌های ممتحن</p>	<p>تصدیق/ اعتبارگذار ی</p>
<p>۱. تحلیل بایگانی‌ها ۲. جمع‌آوری مصنوعات مانند نمونه‌های کار ۳. تحلیل محتوای مستندات برنامه ۴. مشاهدات مستقل و مشارکتی ۵. مصاحبه ۶. تحلیل منطقی عملیات ۷. گروه‌های کانونی</p>	<p>توضیح موثق، عمیق و مستند سیاست برای ذی‌نفعان آن</p>	<p>توصیف، تحلیل و سنتز عمیق و متمرکز یک سیاست</p>	<p>کمپل، لینکلن و گوبا</p>	<p>مورد پژوهی</p>

۸. آزمون پرسشنامه ۹. مقیاس‌های درجه‌بندی ۱۰. تحقیق عمومی ۱۱. فروم‌ها ۱۲. پایگاه‌داده				
۱. تحلیل سیستم ۲. تکنیک بازبینی و ارزیابی برنامه ۳. روش مسیر بحرانی ۴. نظام برنامه‌ریزی و بودجه‌ریزی برنامه (PPBS) ۵. مدیریت هدف‌گرا ۶. نظام اطلاعات کامپیوترمحور ۷. گزارش‌های متناوب پیشرفت پرسنلی ۸. گزارشگری منظم بودجه	ارائه اطلاعات به مدیران برای برنامه‌ریزی، پایش، کنترل و گزارشگری	طراحی یک پایگاه داده برای سیاست و جمع‌آوری منظم و مستمر اطلاعات آن	دمینگ	نظام‌های اطلاعات مدیریت
۳. چک لیست ۲. نیازسنجی ۳. ارزیابی فارغ از اهداف ۴. آزمایش و شبه‌آزمایش ۵. تحلیل روش کار ۶. تحلیل هزینه	انتخاب میان گزینه‌های سیاستی با توجه به هزینه‌های مختلف، نیازهای گروه‌های مصرف‌کننده، ارزش‌های عمومی جامعه و شواهد نتایج مثبت و منفی	محوریت رفاه مصرف‌کننده و خدمت به وی در تعیین شایستگی سیاست	اسکریون	مصرف کننده‌گرا
۱. تحلیل هزینه ورودی‌ها ۲. تحلیل هزینه-اثربخشی	قضاوت درباره بهره‌وری سیاست	مقایسه هزینه و فایده سیاست	تحلیل‌گرا ن اقتصادی	هزینه- فایده
متناسب با نظریه پایه	توسعه و تکامل	ارزیابی سیاست	گلیزر و	نظریه‌محور

	استراوس	بر مبنای یک نظریه	مفهومی سیاست	
ترکیب روش‌ها	تایلر	استفاده از هر دو روش کمی و کیفی در ارزیابی سیاست	ارزیابی تکوینی و تلخیصی سیاست	۱. پیمایش با استفاده از نمونه‌های معرف و نمونه‌های تداومی - مقطعی و جرگه‌ای ۲. آزمون‌های قاعده مرجع ۳. مقیاس‌های درجه‌بندی ۴. شبه آزمایش ۵. آزمون‌های معناداری آثار اصلی ۶. آزمون‌های آماری استقرایی ۷. قوم‌نگاری ۷. تحلیل اسنادی ۸. تحلیل روایتی ۹. نمونه‌های هدفمند ۱۰. مشاهده‌گری مشارکتی ۱۱. مشاهده‌گری مستقل ۱۲. مطلعان کلیدی ۱۳. کمیته‌های مشورتی ۱۴. مصاحبه‌های ساختمند و غیر ساختمند ۱۵. گروه‌های کانونی ۱۶. موردپژوهی‌های افراد و گروه‌ها ۱۷. مطالعه برونی‌ها ۱۸. خاطره‌نویسی ۱۹. مدل‌های منطقی ۲۰. نظریه برخاسته از داده‌ها ۲۱. نمودارهای گردش کار ۲۲. درخت‌های تصمیم ۲۳. ماتریس‌ها ۲۴. ارزیابی‌های عملکرد
مراجع	استیک	تعامل مستمر	فراهم کردن	موردپژوهی ۲. اهداف گویا

محور (پاسخگو)	ارزیاب با متقاضیان، خودگردانی محلی و کمک به افراد درگیر سیاست برای انجام ارزیابی و استفاده از آن برای بهبود سیاست	اطلاعات مفید در دسترس برای متقاضیان و پاسخ به نیازهای متنوع آن‌ها و سایر ذی‌نفعان	۳. نمونه‌گیری هدفمند ۴. مشاهده ۵. گزارش‌های مخالف ۶. داستان‌گویی ۷. نقش‌گذاری اجتماعی ۸ گزارش‌های روایتی
محکمه‌ای (قضایی)	قضاوت نهایی درباره مطلوبیت سیاست بر مبنای ملاحظه توأمان نظرات موافقان و مخالفان آن	قضاوت در مجادلات سیاسی درباره سیاست	انواع روش‌های کمی و کیفی
تصمیم/ پاسخگویی گرا	کمک به مجریان سیاست برای بهبود تصمیم- گیری و پاسخگویی در مقابل اقداماتشان و درگیر کردن آن‌ها در ارزیابی	ارائه هر دو نوع اطلاعات ارزیابی تکوینی و تلخیصی به ذی‌نفعان برای بهبود سیاست	۱. پیمایش ۲. برآورد نیازها ۳. موردپژوهی ۴. تیم‌های مدافع ۵. مشاهده ۶. مصاحبه ۷. ارزیابان مقیم ۸ طراحی‌های آزمایش و شبه آزمایش
سازه‌گرا	قراردادن ذی‌نفعان در مرکزیت ارزیابی؛ انکار وجود هر- گونه واقعیت نهایی	یادگیری اجتماعی ۲. توانمند کردن ذی‌نفعان	هرمنوتیک ۲. دیالکتیک ۳. مذاکره

دموکراتیک مشورتی	هاوس و هاو	مشارکت دموکراتیک ذی‌نفعان فرایند ارزیابی	شنیدن صدای همه ذی‌نفعان در ارزیابی سیاست	بحث با ذی‌نفعان ۲. پیمایش ۳. مناظره
سودمندی- مدار	پاتون	تاکید بر به‌کار گرفته‌شدن یافته‌های ارزیابی؛ استفاده از پنل نماینده کاربران ارزیابی برای پیشبرد ارزیابی متناسب با نیاز کاربر	هر نوع نیاز و مأموریت مورد نظر	تحلیل ذی‌نفعان ۲. کار با پنل نماینده کاربران ارزیابی ۳. انواع روش‌های کمی و کیفی
نقد و خبرگی	آیزنر	توصیف، ارزیابی انتقادی و روشن- کردن شایستگی یک سیاست خاص	عدم امکان ارزیابی عمیق سیاست از طرق دیگر	نقد و ارزیابی فنی- تخصصی خبرگان

قرائت استفلیم از مدل‌های ارزیابی به چهار دلیل انتخاب شده است: اولاً، وی از صاحب‌نظران و معماران اصلی ارزیابی است؛ ثانیاً، رهیافت‌هایی که وی از یکدیگر تفکیک کرده از تفصیل لازم برخوردار است و تفاوت میان مدل‌های ارزیابی را منعکس می‌کند؛ ثالثاً، آن‌ها قابلیت پوشش‌دادن سایر مدل‌های ارزیابی را دارند و می‌توان انواع مدل‌های نوظهور ارزیابی را در آن‌ها جایگذاری کرد؛ رابعاً، وی معرفی جامعی از مدل‌های ارزیابی ارائه کرده که پیشتان، روش‌ها و تکنیک‌ها، و کاربرد آن‌ها را دربرمی‌گیرد.

نتیجه‌گیری

در این مقاله، ضمن مرور مطالعات ارزیابی، نشان داده شد در ورای آشفتگی مفهومی آن، جریان پیوسته و مرتبطی از دانش وجود دارد. ادبیات ارزیابی در سطح عملیاتی پرورش یافته و

معطوف به چگونگی طراحی و اجرای ارزیابی است. به عبارت دیگر، زمانی می‌توان از مباحث آن استفاده کرد که تصمیم به ارزیابی گرفته شده و در مرحله طراحی ارزیابی باشیم. تکامل جریان مسلط ارزیابی همسو با تکامل مباحث معرفت‌شناسی و روش‌شناسی در علوم اجتماعی صورت گرفته و فراز و نشیب‌های آن را پیموده است. ارزیابی قبل از آنکه بر مفهوم تخصصی مشخصی دلالت کند یک روش بوده و به مرور زمان برای انجام هر چه بهتر و پاسخ به نیازهای ذی‌نفعان تکامل یافته است. اهمیت ثانویه بُعد مفهومی موجب شده ادبیات ارزیابی بیش از هر چیز با قرائت‌های متنوع از مفاهیم پایه مواجه باشد. این امر از یک سو در درک چستی ارزیابی مشاهده می‌شود که موجب شده مفاهیم متعددی از قبیل انطباق، سنجش، ممیزی، پایش، حسابرسی، ارزشیابی، نظارت، کنترل، قضاوت، نظرسنجی و بررسی هر یک نوعی «ارزیابی» قلمداد شوند و از سوی دیگر در وضع اسامی متعدد برای دلالت‌های مفهومی مشابه و تلقی‌های مختلف از مفاهیم کم و بیش یکسان در جریان مسلط ارزیابی دیده می‌شود. انواع مفاهیمی که در ارزیابی مطرح می‌شوند در چارچوب مفهومی پژوهشگر معنا می‌یابند و قراردادی هستند.

از آنجا که تحقیق ارزیابی در یک زمینه کاربردی و نیازمحور رشد کرده وجه کاربردی بر وجه نظری و مفهومی آن غلبه دارد. علاوه بر این، ارزیابی یک موضوع مطالعاتی چندرشته‌ای است. مدل‌ها و روش‌های ارزیابی سیاست اختصاص به یک رشته خاص ندارد و پژوهشگرانی از رشته‌های مختلف آن‌ها را توسعه داده‌اند. بسیاری از ارزیابی‌پژوهانی که آثارشان در مبحث ارزیابی سیاست در کتاب‌های سیاستگذاری عمومی مورد استناد قرار می‌گیرد در آموزش، علوم اجتماعی و روانشناسی تحصیل کرده‌اند. غفلت از این نکته در فضای دانشگاهی و اجرایی ایران، گاه موجب متفاوت دانستن مدل‌های ارزیابی سیاست با مدل‌های ارزیابی طرح‌ها و برنامه‌ها در حوزه‌های مختلف صنعت، آموزش، فرهنگ، سلامت و غیره شده است. در واقع، برخلاف تصور رایج - و نه مطلق - در فضای اجرایی و دانشگاهی ایران روش‌ها و مدل‌های ارزیابی پروژه/برنامه/سیاست در زمینه‌های مختلف موضوعی تفاوتی با یکدیگر ندارند. آن‌ها صرفاً از نظر سطح تحلیل، میزان پیچیدگی و ویژگی‌های زمینه‌ای از هم متمایز می‌شوند و از ابزارهای تحلیلی مشابهی استفاده می‌کنند. هم‌چنان که یک پژوهشگر اجتماعی با توجه به موضوع و مسئله پژوهش از میان انواع روش‌های علوم اجتماعی دست به انتخاب می‌زند، ارزیاب نیز با توجه به ویژگی‌های طرح/برنامه/سیاست مورد نظر استراتژی ارزیابی را طراحی می‌کند، مدل ارزیابی مناسب را بر می‌گزیند و متناسب با آن از روش‌ها و تکنیک‌های علوم-اجتماعی، علوم رفتاری، اقتصاد و غیره برای گردآوری و تحلیل داده‌ها استفاده می‌کند.

نگارنده در این مقاله، تلاش کرده با اتکا به مطالعات ارزیابی پاره‌ای از ابهامات موجود در شناخت ارزیابی سیاست در ایران را روشن نماید. سیاست‌پژوهان و کارشناسان نهادهای سیاستگذار و برنامه‌ریز ایران در حالی با این حوزه آشنایی ندارند که روش‌ها و تکنیک‌های آن سال‌هاست در دانشکده‌های رشته‌های مختلف علوم انسانی کشور تدریس می‌شود و دست‌کم در حوزه آموزش تجارب قابل توجهی انباشته شده است. دانش، یک‌کل به‌هم‌پیوسته است و تقسیمات رشته‌ای نباید به تنگ‌نظری در استفاده از آن بینجامد. پژوهشگران رشته‌های مختلف سیاستگذاری، اقتصاد، علوم اجتماعی، روانشناسی، مدیریت و آموزش می‌توانند از یکدیگر بیاموزند و دانش ارزیابی سیاست در ایران را توسعه دهند^۱.

منابع و مآخذ

الف) فارسی

- ۱- اشترینان، کیومرث. ۱۳۹۰. مقدمه‌ای بر روش سیاستگذاری فرهنگی. اداره کل مطالعات اجتماعی و فرهنگی شهرداری تهران. تهران: جامعه‌شناسان.
- ۲- الوانی، مهدی. ۱۳۸۵. تصمیم‌گیری و تعیین خطمشی دولتی. چاپ یازدهم. تهران: سمت.
- ۳- الوانی، مهدی و فتاح شریف زاده. ۱۳۹۰. فرایند سیاستگذاری عمومی. چاپ هشتم. تهران: دانشگاه علامه طباطبائی.
- ۴- پارسونز، واین. ۱۳۸۵. مبانی سیاستگذاری عمومی و تحلیل سیاست‌ها. ترجمه حمیدرضا ملک محمدی. تهران: پژوهشکده مطالعات راهبردی.
- ۵- چالمرز، آلن اف. ۱۳۸۴. چستی علم: درآمدی بر مکاتب علم شناسی فلسفی. ترجمه سعید زیباکلام. تهران: سمت.
- ۶- شافریتز، جی.ام و کریستوفر پی. بریک. ۱۳۹۰. سیاستگذاری عمومی در ایالات متحده آمریکا. ترجمه حمیدرضا ملک محمدی. تهران: دانشگاه امام صادق.۷.
- ۷- لستر، جیمز پی. و جوزف استوارت جی. آر. ۱۳۸۱. فرایند خطمشی گذاری عمومی: رویکردی تکاملی. ترجمه مجتبی طبری، تورج مجیبی میکلائی و سهیل خورشیدی. تهران: نشر ساوالان.
- ۸- مانهایم، یارول و ریچارد ریچ. ۱۳۸۵. روش‌های تحقیق در علوم سیاسی: تحلیل تجربی. ترجمه لی‌لا سازگار. چاپ دوم. تهران: مرکز نشر دانشگاهی.
- ۹- مولر، پیر. ۱۳۷۸. سیاستگذاری عمومی. ترجمه حمیدرضا ملک محمدی. تهران: دادگستر.
- ۱۰- هاولت، مایکل و ام. رامش. ۱۳۸۰. مطالعه خطمشی عمومی. ترجمه عباس منوریان و ابراهیم گلشن. تهران: مرکز آموزش مدیریت دولتی.
- ۱۱- وحید، مجید. ۱۳۸۳. سیاستگذاری عمومی. تهران: میزان.

ب) انگلیسی

۱- نکته قابل تامل آن است که بسیاری از دانش‌آموختگان این رشته‌ها یا «ارزیابی سیاست» را حوزه مطلق خود می‌پندارند یا نسبت به توانمندی‌های خود در این عرصه آگاه نیستند و از روش‌ها و تکنیک‌های آن اظهار بی‌اطلاعی می‌کنند.

- 12- Alkin, Marvin C., ed. 2004. *Evaluation Roots*. London, Thousand Oaks and New Delhi: Sage Publications. doi:10.4135/9781412984157.
- 13- Ballart, Xavier. 1998. "Spanish Evaluation Practice Versus Program Evaluation Theory: Cases From Five Policy Areas." *American Journal of Evaluation* 4 (2): 149-70. doi:10.1177/13563899822208509.
- 14- Braybrooke, David, and Charlez E. Lindblom. 1963. *A Strategy of Decision, Policy Evaluation as a Social Process*. London: The Free Press of Glencoe, Collier-Macmillan Limited.
- 15- Chouinard, Jill Anne, and J. Bradley Cousins. 2009. "A Review and Synthesis of Current Research on Cross-Cultural Evaluation." *American Journal of Evaluation* 30 (4): 457-94. doi:10.1177/1098214009349865.
- 16- Dunn, Willaim N. 1994-2012. *Public Policy Analysis: An Introduction*. Englewood Cliffs: Prentice Hall.
- 17- Government Chief Social Researcher's Office. 2003. *The Magenta Book: Guidance Notes for Policy Evaluation and Analysis*. London: Prime Minister's Strategy Unit, Cabinet Office, Admiralty Arch, The Mall, SWIA2WH.
- 18- Guba, Egon G., and Yvonna S. Lincoln. 1989. *Fourth Generation Evaluation*. London, Thousand Oaks and New Delhi: Sage Publications.
- 19- Guba, Egon G., and Yvonna S. Lincoln. 2001. *Guidelines and Checklists for Constructivist (a.k.a. Fourth Generation) Evaluation*. Evaluation Checklists Project. <http://www.wmich.edu/evalctr/checklists>.
- 20- Hanberger, Anders. 2006. "Evaluation of and for Democracy." *Evaluation* 12 (1): 17-37. doi:10.1177/1356389006064194.
- 21- Hansen, Hanne Foss. 2005. "Choosing Evaluation Models: A Discussion on Evaluation Design." *Evaluation* 11 (4): 447-62. doi:10.1177/1356389005060265.
- 22- Internal Oversight Service. 2007. *Evaluation Handbook*. Internal Oversight Service, Evaluation Section, IOS/EVS/PI/63, December.
- 23- Langbein, Laura and Claire L. Felbinger. 2006. *Public Program Evaluation: a Statistical Guide*. New York: M.E. Sharpe, Inc.
- 24- Mark, Melvin M., Jennifer C. Greene, and Ian F. Shaw. 2006. "Introduction: The Evaluation of Policies, Programs, and Practices." In *The Sage Handbook of Evaluation*, edited by Ian F. Shaw, Jennifer C. Greene, and Melvin M. Mark, 1-29. London, Thousand Oaks and New Delhi: Sage Publications. doi:10.4135/9781848608078.
- 25- OECD (Organisation for Economic Co-operation and Development). 2010. *DAC Criteria for Evaluating Development Assistance*. Development Assistance Committee. http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html.
- 26- Preskill, Hallie, and Darlene Russ-Eft. 2005. *Building Evaluation Capacity*. London, Thousand Oaks and New Delhi: Sage Publications. doi: 10.4135/9781412983549.
- 27- Purdon, Susan, Carli Lessof, Kandy Woodfield, and Caroline Bryson. 2001. *Research Methods for Policy Evaluation*. Research Working Paper. no. 2. Department for Work and Pensions, National Centre for Social Research.
- 28- Rossi, Peter Henry, Howard E. Freeman, and Mark W. Lipsey. 1999. *Evaluation: A Systematic Approach*. London, Thousand Oaks and New Delhi: Sage Publications, Inc.
- 29- Seigart, Denise. 2005. "Feminist Evaluation." In *Encyclopedia of Evaluation*, edited by Sandra Mathison, 154-57. London, Thousand Oaks and New Delhi: Sage Publications. doi:10.4135/9781412950558.
- 30- Shadish, William R., Thomas D. Cook, and Laura C. Leviton. 1991. *Foundations of Program Evaluation, Theories of Practice*. London, Thousand Oaks and New Delhi: Sage Publications, Inc.
- 31- Stufflebeam, Daniel L. 2001. "Foundational Models for 21st Century Program Evaluation" *New Directions for Evaluation*. no. 89. Jossey-Bass, A Publishing Unit of John Wiley & Sons, Inc.
- 32- Treasury Board of Canada. 1998. *Program Evaluation Methods: Measurement and Attribution of Program Results*. Third Edition, Minister of Public Works and Government Services, Public Affairs Branch, Treasury Board of Canada, Secretariat, March.
- 33- UNDP (United Nations Development Programme). 2009. *Handbook on Planning, Monitoring and Evaluating for Development Results*. New York: NY 10017.
- 34- W. K. Kellogg Foundation. 1998. *Evaluation Handbook*. W. K. Kellogg Foundation publication.