

# ارزیابی سیاست و پژوهش ارزیابی<sup>۱</sup>

تألیف هلموت ولمن مترجم: امیراتابک چاپاری<sup>۲</sup>

## • تعاریف، مفاهیم، و انواع ارزیابی

ارزیابی در حوزه سیاست عمومی ممکن است به طور کلی به عنوان یک ابزار و رویه ای تحلیلی برای انجام دو کار تعریف شود، اول، تحقیق ارزیابی به عنوان یک ابزار تحلیلی، شامل بررسی‌های یک برنامه سیاست برای به دست آوردن تمام اطلاعات مربوط به ارزیابی عملکرد خود، یعنی هم فرایند و هم نتیجه‌اش؛ و دوم، ارزیابی به عنوان مرحله‌ای از چرخه سیاست که به طور کلی‌تر به گزارش اطلاعاتی که به فرایند سیاست‌گذاری مرتبط است، اشاره دارد. Wollmann 2003b, 4.

با این حال، مجموعه گیج‌کننده ای از مفاهیم و اصطلاحات در این زمینه، به ویژه با توجه به توسعه اخیر "موج سوم" واژگان جدید (مانند بازرسی مدیریت، بازرسی سیاست و نظارت بر عملکرد) ظاهر شده است. در پرتو تعریفی که بر کارکرد ارزیابی تمرکز دارد و بدین سان، زیر سطح اصطلاحات متنوع به نظر می‌رسد که اصطلاحات مختلف "کم و بیش یک زمینه را پوشش می‌دهد" (Bemelmans-Videc 2002, 94) بنابراین، روش‌های تحلیلی که «بازرسی عملکرد» نامیده می‌شوند، در تعریف ما گنجانده می‌شوند. به جز البته، "بازرسی مالی" که مطابقت هزینه‌های عمومی را با مقررات بودجه بررسی می‌کند و به عنوان یک ارزیابی به حساب نمی‌آید. (Sandahl 1992, 115)

<sup>۱</sup> این متن ترجمه اولیه فصل بیست و ششم کتاب (Handbook of Public Policy Analysis (2007) است. طبیعتاً کاستی‌های متن بر گردن مترجم است.

<sup>۲</sup> کارشناسی ارشد سیاست‌گذاری عمومی دانشگاه تهران (amir.atabak.chap@ut.ac.ir)

## • انواع ارزیابی: عملکردها و زمان‌بندی

از نظر پیوندهای زمانی و عملکردی مختلف با "چرخه سیاست"، اغلب تمایزات زیر انجام می‌شود. (Wollmann 2003).

ارزیابی پیشین (قبل از واقعه) تصمیم‌گیری از قبل به معنای پیش‌بینی و پیش‌ارزیابی اثرات و پیامدهای (فرضی) سیاست‌ها و اقدامات برنامه‌ریزی شده یا تعریف شده به منظور تغذیه کردن اطلاعات در فرآیند تصمیم‌گیری آینده یا در حال انجام است. اگر در مسیرهای جایگزین از سیاست‌ها و اقدامات انجام شود، ارزیابی پیشین ابزاری برای انتخاب بین گزینه‌های سیاست جایگزین (به طور ایده آل) از نظر تحلیلی شفاف‌تر، قابل پیش‌بینی‌تر و از نظر سیاسی قابل بحث‌تر است. این شامل اجرای ارزیابی آغازین است که به منظور پیش‌بینی تحلیلی روند اجرای سیاست با تمرکز بر فرآیند و همچنین ارزیابی اثرات محیطی آن و پیش‌بینی پیامدهایی که سیاست‌ها و اقدامات در نظر گرفته شده ممکن است بر محیط داشته باشد طراحی شده است.

ارزیابی توأمان وظیفه‌اش شناسایی اثرات و نتایج (موقتی) برنامه‌ها و اقدامات سیاست‌ها است در حالی که در چرخه سیاست، اجرا و ادراک همچنان در حال انجام است. کارکرد اساسی ارزیابی "توأمان" این است که برای بازگرداندن اطلاعات مرتبط به فرایند پیاده‌سازی در یک نقطه و مرحله زمانی که می‌توان از اطلاعات مربوط به ترتیب برای تنظیم، تصحیح یا تغییر جهت فرآیند اجرا یا حتی کلید اصلی تصمیمات سیاستی استفاده کرد. در یک کاربرد تقریباً مشابه برخی از *ارزیابی توأمان* که به موازات فرآیند اجرای سیاست انجام می‌شود صحبت می‌کنند. در ارزیابی "مستمر" یا "توأمان" می‌توان بین روشی عمدتاً «تحلیلی» که از فرآیند اجرا «منفصل» و یا «دورادور (نیمه متصل)» باقی می‌ماند، به منظور تعیین عینیت تمیز داد. به علاوه اصطلاح مداخله‌گرانه در ارزیابی توأمان زمانی به کار می‌رود که علاوه بر دستور تحلیلی، از ارزیابان نیز؛ به منظور رفع نواقص موظف به مداخله فعال در روند اجرا و ایراد در فرآیند اجرا که دستیابی به اهداف سیاستی از

پیش تعیین شده را به خطر می‌اندازد نباشند، انتظار داریم. در چنین جهت گیری «مداخله گرایانه»، ارزیابی «توامان» به مفهوم اقدام پژوهی نزدیک می‌شود.

در نهایت، نظارت را می‌توان به عنوان یک روش ارزیابی (مستمر - در حال انجام) که با هدف شناسایی (توصیفی) و در صورت امکان با کمک به جا و به کارگیری شاخص‌ها در «اندازه‌گیری» اثرات فعالیت‌های جاری دانست.

ارزیابی مبتنی بر نتایج واقعی (بعد از وقوع) نوع کلاسیک ارزیابی را برای ارزیابی دستیابی به هدف و اثرات سیاست‌ها و اقدامات تشکیل می‌دهد، هنگامی که آنها به طور کامل به این ترتیب، به برنامه‌های سیاستی هدایت شده خلاصه شده‌اند (به عنوان یک فرم اقدام سیاستی ترکیبی از اهداف سیاسی و منابع مالی، سازمانی و همچنین پرسنلی) نمونه‌ای از سیاست‌های اصلاحات اولیه در ایالات متحده، و همچنین در کشورهای اروپایی. ارزیابی سیاسی بعد از وقوع اغلب با ارزیابی برنامه شناسایی شده است. مشخصاً، به ارزیابی سیاست (یا برنامه) در درجه اول دو وظیفه داده شده است.

اولاً این به معنای تولید یک ارزیابی در درجه دستیابی به اهداف سیاست مورد نظر ("دستیابی به هدف") است. مشکلات مفهومی ناشی از این کار حول مفهوم سازی مناسب می‌چرخد، در صورت امکان شاخص‌هایی قابل اندازه‌گیری برای انجام چنین ارزیابی‌هایی از دستیابی به هدف باشد. اما، علاوه بر شناسایی پیامدهای «در نظر گرفته شده»، ارزیابی تأثیرات سیاست‌ها و برنامه‌ها به پیامدهای غیرمنتظره نیز مربوط می‌شود.

دوماً، ارزیابی سیاست‌ها و برنامه‌ها برای پاسخ به این سؤال (علّی) در مورد اینکه آیا اثرات و تغییرات مشاهده شده واقعاً (به طور علی) به سیاست یا برنامه مورد نظر مرتبط بوده است یا خیر نیز انتظار و التزام دارد. از این جا مسئله روش شناختی به کارگیری ابزارها و مهارت‌های روش شناختی (احتمالاً و امیدوارانه) قادر به حل «معمای علی» است.

فرا ارزیابی به معنای تجزیه و تحلیل یک ارزیابی از قبل تکمیل شده (اولیه) با استفاده از نوعی تحلیل ثانویه است. اول، فرا ارزیابی ممکن است بخش از قبل تکمیل شده ارزیابی (اولیه) را بررسی کند که آیا مطابق با معیارها و استانداردهای روش شناختی است یا خیر. ممکن است از فرا ارزیابی با بازنگری روش شناسی صحبت شود. دوم، فرا ارزیابی باید یافته‌های اساسی ارزیابی (اولیه) تمام شده در حال حاضر را جمع‌آوری کند و نتایج را ترکیب کند. این را می‌توان یک فرا ارزیابی «ترکیبی» نامید.

در حالی که ارزیابی (دقیق) ارائه تصویری جامع از آنچه در زمینه سیاست و پروژه تحت بررسی اتفاق افتاده است را شامل دوره‌های موفق و همچنین ناموفق رویدادها نشانه گرفته است، بهترین رویکرد عملی تمایل دارد تا داستان‌های موفقیت آمیز سیاست‌ها و پروژه‌های اصلاحی را با هدف تحلیلی شناسایی عوامل تبیین کننده موفقیت، و با هدف کاربردی (یادگیری و آموزشی) برای پرورش عبرت گیری از چنین تجربیاتی در زمینه‌های بین‌المللی و همچنین در زمینه ملی و فراملی بازگو کند. از یک طرف، چنین داستان‌های عملی خوب مملو از تهدید (مفهومی و روش شناختی) مغالطه بوم شناختی است که این ترجمه و انتقال سریع و گمراه‌کننده استراتژی‌های (به ظاهر مثبت) از یک محل و یک کشور به کشور دیگر است. از سوی دیگر، اگر به گونه‌ای انجام شود که با دقت به شرایط خاص و مشروط بودن این نمونه‌های عملی خوب توجه شود تجزیه و تحلیل، گفتن و انتشار چنین مواردی می‌تواند مسیر سریع مفیدی برای دانش ارزیابی و یادگیری درون ملی و همچنین فراملی ارائه دهد.

در مقابل این موانع مفهومی و روش شناختی متعدد، ارزیابی کامل اصلاحات بخش عمومی با نوعی شبه ارزیابی پیشنهاد شده است که کمتر از یک ارزیابی تمام عیار مملو از مخاطرات مفهومی و روش شناختی خواهد بود و تمایل بیشتری به تمرکز و محدود کردن خود به اطلاعات و داده‌ها و کارکردهای توصیفی ارزیابی به جای توضیحی دارد. یک دارایی اصلی ممکن است یک نوع مفهومی و روش شناختی تقلیل یافته از شبه ارزیابی باشد که ممکن است منجر به برقراری ارتباط مطمئن‌تر بین سیاست‌گذار و

ارزیاب شود که فرایند یادگیری تدریجی را ترویج می‌کند که فرهنگ اطلاعاتی را تقویت می‌کند. (Thoenig 2003).

در نهایت، میزان (قابلیت) ارزیابی ارزیابی را می‌توان انجام داد. این قبل از ارزیابی اتفاق می‌افتد، چه از نوع ارزیابی قبل از واقعه، چه از نوع ارزیابی بعد از وقوع و در همچنین ارزیابی در حال انجام. از آن استفاده می‌شود تا مشخص شود که کدام رویکرد و نوع ارزیابی باید بر اساس معیارهای امکان‌سنجی فنی، امکان‌پذیری اقتصادی، و شایستگی‌های عملی مورد توجه قرار گیرد.

ارزیابی کلاسیک، اول از همه، معطوف به ارزیابی (پس از وقوع) دستیابی یا عدم دستیابی به اهداف سیاست و برنامه یا برآورد (پیش از وقوع) دستیابی به اهداف است. که اساساً با اثربخشی سیاست‌ها و میزان منابع به کار گرفته شده (یا سرمایه‌گذاری شده) برای رسیدن به آن هدف سروکار دارد. این در تضاد با یک تحلیل هزینه - فایده است که نتایج را با منابع اختصاص داده شده برای دستیابی به آنها مقایسه می‌کند. بنابراین تأکید بر تحلیل هزینه و فایده کارآمدی ممکن است یک جهت‌گیری پس از وقوع نیز داشته باشد.

### • انواع ارزیابی: داخلی و خارجی

در موردی، ارزیابی ممکن است به عنوان یک ارزیابی داخلی انجام شود. چنین ارزیابی در داخل توسط خود آژانس عامل انجام می‌شود. در این مورد خودارزیابی انجام می‌گیرد. در واقع، ممکن است استدلال شود که شیوه‌های غیررسمی و غیرسیستماتیک خودارزیابی از زمانی که (در مدل بوروکراسی وبری) نظارت سلسله‌مراتبی بر اساس اشکال گزارش دهی منظم داخلی صورت گرفته شکل گرفته است. اما تحقیقات ارزیابی شامل رویکردهای رسمی‌تری است. تحقیقات ارزیابی به یک جزء کلیدی از نظریه‌های مختلف مدیریت دولتی تبدیل شده است. مدیریت دولتی نوین در سال‌های اخیر بر مفهوم نظارت و کنترل بر اساس شاخص‌های عملکرد ارزیابی تأکید کرده است. برای مثال،

چنین شاخص‌هایی نقشی محوری در سیستم‌های عامل حسابداری جامع دستیابی به هزینه داخلی دارند. (see Wollmann 2003b).

در مقابل، ارزیابی خارجی توسط منابع خارجی (بستن قرارداد با یک آژانس یا بازیگر خارج از واحد اداری عملیاتی) آغاز یا تأمین می‌شود. چنین جایگاه خارجی ارزیابی عملکرد توسط نهادها و بازیگرانی که خارج و فراتر از مدیریت هستند، ممکن است ایجاد شود که دارای منافع سیاسی یا ساختاری باشند و از ارزیابی به عنوان وسیله‌ای برای نظارت بر اجرای سیاست‌ها توسط دولت استفاده کنند. پارلمان‌ها نشان داده‌اند که نامزدهای طبیعی برای آغاز و اجرای ارزیابی سیاست‌ها و برنامه‌های افتتاح شده توسط دیگران هستند. در همین راستا، دادگاه‌های حسابرسی از ارزیابی به عنوان یک روش تحلیلی اضافی برای روشن کردن اثربخشی و کارایی عملیات اداری استفاده کرده‌اند.

اما همچنین سایر بازیگران در هسته دولت، مانند دفتر نخست وزیری یا وزارت دارایی، ممکن است به ارزیابی به عنوان ابزاری برای نظارت بر عملیات وزارتخانه‌ها و آژانس‌های بخشی روی آورند. در نهایت، باید به این نکته اشاره کرد که نهادها و کمیسیون‌های موقت (یعنی کمیسیون‌های تحقیق) موظف به بررسی دقیق موضوعات پیچیده و زمینه‌های سیاست هستند. چنین کمیسیون‌هایی ممکن است قبل از توصیه به اجرای سیاست‌ها توسط دولت و وزارتخانه‌ها، از ارزیابی به عنوان یک ابزار مهم حقیقت‌یابی استفاده کنند.

هرچه سیاست‌ها و برنامه‌های مورد بررسی پیچیده‌تر باشند و مشکلات مفهومی و روش‌شناختی انجام این گونه ارزیابی‌ها سخت‌تر شود، مؤسسات کمتری شروع کننده و قادر به انجام ارزیابی چنین تحلیل‌های مفهومی و روش‌شناختی پیچیده و پیچیده‌ای هستند. با توجه به چنین پیچیدگی‌هایی، تحقیقات ارزیابی به طور ایده آل مبتنی بر کاربرد روش‌شناسی و تخصص علوم اجتماعی است. بنابراین، در فقدان پرسنل آموزش‌دیده کافی و زمان سیاست، ارگان‌های اجرایی، اداری و سایر مؤسسات اغلب به

مؤسسات و سازمان‌های تحقیقاتی خارج (علوم اجتماعی) مراجعه می‌کنند تا آنها را برای انجام کار ارزیابی به صورت قراردادی مأمور کنند. (see Wollmann 2002).

- در واقع، توسعه ارزیابی، از اواسط دهه ۱۹۶۰، با گسترش (گاهی افسارگسیخته) "بازار پول قراردادی" همراه بوده است که از منابع وزارتخانه‌ها، مجلس، کمیسیون‌های موقت و غیره تغذیه و تحقیقات ارزیابی را عملاً به "صنعت جدید با بعد قابل توجهی" تبدیل کرده است. (Freeman and Solomon 1981, 13) که حول تحقیقات قراردادی می‌چرخد و چشم انداز تحقیقات سنتی را در یک تغییر مهم از تحقیقات «آکادمیک به کارآفرینی» عمیقاً تغییر داده است، موضوعی که به آن باز می‌گردیم.

#### • سه موج ارزیابی

سه مرحله را می‌توان در توسعه ارزیابی طی چهار سال گذشته متمایز کرد: اولین موج ارزیابی در دهه‌های ۱۹۶۰ و ۱۹۷۰ بود، موج دوم در اواسط دهه ۱۹۷۰ آغاز شد و موج سوم از دهه ۱۹۹۰ شروع شد. در طول دهه‌های ۱۹۶۰ و ۱۹۷۰، ظهور دولت رفاه پیشرفته با مفهوم افزایش توانایی دولت برای ارائه سیاست‌گذاری فعالانه از طریق نوسازی ساختارهای سیاسی و اداری خود همراه بود که در پی آن نهادینه‌سازی و به‌کارگیری برنامه‌ریزی ظرفیت‌های اطلاعاتی و ارزیابی به عنوان ابزاری در نظر گرفته شد. همان‌طور که قبلاً ذکر شد، مفهوم "چرخه سیاست" حول محور سه‌گانه شکل‌گیری، اجرا و خاتمه سیاست می‌چرخید، که به موجب آن ارزیابی به عنوان یک حلقه "سایبرنتیک" در جمع‌آوری و بازخورد اطلاعات مربوط به سیاست، حیاتی تلقی می‌شد. منطق علمی زیربنایی و چشم انداز یک مدل سیاست محور علمی با فراخوان معروف دونالد کمپبل برای جامعه آزمایشی تجسم شد. (Campbell 1969, "reforms as experiments")

در ایالات متحده، افزایش ارزیابی با آغاز به کار برنامه‌های اقدام اجتماعی فدرال مانند جنگ علیه فقر در اواسط دهه ۱۹۶۰ در زمان پرزیدنت جانسون اتفاق افتاد که ارزیابی تقریباً به طور معمول توسط قانون اصلاحات مربوطه اجباری شد و ارزیابی سیاست و

برنامه را عملاً به یک صنعت در حال رشد تبدیل کرد. آزمایش اجتماعی در مقیاس بزرگ همراه با ارزیابی عمده از این روش پیروی کرد. در اروپا، سوئد، آلمان و بریتانیا به پیشتازان این «موج اول» شناخته شدند. ( see Levine 1981; Wagner, and Wollmann 1986 Derlien 1990)؛ در آلمان آزمایش اجتماعی (experimentelle Politik) در مقیاسی بی نظیر در خارج از ایالات متحده انجام شد. (see Wagner, and Wollmann 1991, 74).

اجماع اصلاح طلبان، در آن زمان به طور گسترده و منعکس کننده توسط بازیگران سیاسی و اداری اصلاح طلب و همچنین توسط دانشمندان علوم اجتماعی مشترک بود. پروژه‌های ارزیابی که از طریق شکل‌های کاملاً ناشناخته تحقیقات قراردادی و مشاوره سیاستی انجام می‌شد، به طور هنجاری با سیاست‌های اصلاح‌طلبی که تحت نظارت بود موافقت و از آن‌ها حمایت می‌کردند و از این رو، هدفشان بهبود نتایج سیاست‌گذاری و به حداکثر رساندن اثربخشی خروجی بود. (Wittrock, Wagner, and Wollmann.1991, 52)

دوران شکوفایی سیاست‌های مداخله‌گرایانه دولت رفاه کوتاه مدت بود، زمانی که پس از اولین افزایش قیمت نفت در سال ۱۹۷۳، اقتصاد جهانی به رکود عمیق‌تری فرو رفت و بودجه‌های ملی با فشار مالی رو به وخامت روبه‌رو شدند که بسیاری از سیاست‌های اصلاحی هزینه‌بردار را متوقف کرد. این منجر به "موج دوم" شد. از آنجایی که سیاست به سمت درخواست‌ها برای کاهش بودجه و صرفه‌جویی در هزینه‌ها تعیین شد، وظیفه ارزیابی سیاست‌ها بر این اساس با هدف کاهش هزینه‌های سیاست‌ها و برنامه‌ها، اگر نه حذف تدریجی آن‌ها، دوباره تعریف شد. (see Wagner, and Wollmann 1986; Derlien, 1990)

در موج دوم ارزیابی با تمرکز بر کارایی هزینه سیاست‌ها و برنامه‌ها، ارزیابی در کشورهای دیگر، به عنوان مثال، در هلند، شاهد گسترش قابل توجهی بود. ( see Leeuw 2004, ) (60).

"موج سوم ارزیابی" تحت تأثیر جریان‌های مختلف عمل می‌کند. یکی، مفاهیم و الزامات مدیریت دولتی جدید که بر گفتمان نوسازی بین‌المللی تسلط پیدا کرده است و در نوع دیگر اصلاحات بخش عمومی در بسیاری از کشورها با «ارزیابی داخلی» (از طریق ایجاد

و به کارگیری شاخص‌ها یا - کنترل مبتنی بر حسابداری و دستیابی به هزینه و غیره) بخشی جدایی ناپذیر از "بسته مدیریت عمومی" و دادن شتاب جدید به رویه‌های ارزیابی را تشکیل داد. (see Wollmann 2003b.) علاوه بر این، در تعدادی از زمینه‌های سیاست، ارزیابی در آشکار ساختن کاستی‌های سیاست موجود و شناسایی پتانسیل اصلاحات و بهبودها برجسته شده است. توجه (و هیجان) زیادی که اخیراً تمرین ارزیابی بین‌المللی بزرگ در مورد سیستم‌های آموزشی ملی، توسط مطالعه "PISA" در گستره اروپا، مطرح شده است، نقش و پتانسیل ارزیابی را به عنوان ابزاری برای سیاست‌گذاری برجسته کرده و بدون شک آن را به سمت جلو پیش برده است. سوم، باید به این نکته اشاره کرد که در اتحادیه اروپا، زمانی که کمیسیون اروپا تصمیم گرفت هزینه‌های کلان صندوق ساختاری اروپا را به طور سیستماتیک ارزیابی کند، به ارزیابی ترویج زیادی داده شد.

از آنجایی که صندوق‌های ساختاری اتحادیه اروپا اکنون در چرخه برنامه پنج‌ساله خود به شیوه‌ای تقریباً شبیه اصول کتاب ارزیابی می‌شوند (با یک چرخه ارزیابی از ارزیابی قبل از وقوع تا ارزیابی پس از وقوع)، ارزیابی سیاست‌ها و برنامه‌های اتحادیه اروپا به طور کلی و قابل توجهی بر توسعه ارزیابی تأثیر گذاشته و آن را به پیش برده است. در برخی کشورها، به عنوان مثال، در ایتالیا، که دستور ارزیابی برنامه‌های اتحادیه اروپا، همان‌طور که بود را داشت، قوانین تحقیقات ارزیابی کشور قبلاً وجود نداشتند.

در یک چشم انداز تطبیقی بین‌المللی در آغاز هزاره جدید، ارزیابی سیاست در بسیاری از کشورها به عنوان ابزاری به طور گسترده پذیرفته شده و مورد استفاده برای به دست آوردن (و "بازخورد") اطلاعات مربوط به سیاست معرفی و نصب شده است. این امر در یک مطالعه اخیر بر اساس گزارش‌های بیست و دو کشور و بر اساس مجموعه‌ای پیچیده از معیارها به طرز چشمگیری تحلیل و مستند شده است. در حالی که ایالات متحده همچنان در «فرهنگ ارزیابی» پیشتاز است، شش رتبه برتر در میان کشورهای اروپایی در اختیار سوئد، هلند، بریتانیا، آلمان، دانمارک و فنلاند قرار دارد (see Furubo, and Sandahl 2002; Leeuw 2004, 63).

## • مسائل روش‌شناختی ارزیابی

پژوهش ارزیابی با دو وظیفه مفهومی و روش‌شناختی اصلی مواجه است: (۱) مفهوم‌سازی تغییرات دنیای واقعی قابل مشاهده بر حسب پیامدهای مورد نظر (یا غیر عمد) که ارزیابی سیاست به منظور شناسایی و ارزیابی آن است (از نظر روش‌شناسی، "متغیرهای وابسته")؛ و (۲) برای یافتن اینکه آیا و چگونه تغییرات مشاهده شده به طور علی با سیاست و معیار مورد بررسی مرتبط هستند (به عنوان متغیر "مستقل"). در رویارویی با این سؤالات کلیدی، تحقیقات ارزیابی بخشی جدایی ناپذیر از تحقیقات علوم اجتماعی به طور کلی است. به این ترتیب، بیشتر مباحث و مجادلات مفهومی و روش‌شناختی علوم اجتماعی را شامل می‌شود. در واقع، به نظر می‌رسد که بحث‌های روش‌شناختی که در جامعه علوم اجتماعی به طور کلی رخ داده است (مثلاً در نزاع بین مکاتب فکری «کمی» و «کیفی») یکی از بارزترین (و گاهی شدیدترین) مبارزات در جامعه تحقیقاتی ارزیابی بوده است.

در این مناقشه دو مرحله قابل تشخیص است. اولین مورد، که به دهه ۱۹۶۰ تا اوایل دهه ۱۹۸۰ باز می‌گردد، که با تسلط مدل علم نئوپوزیتیویستی - نومولوژیک (با برتری متعاقب روش‌های کمی و شبه تجربی) مشخص شده است. دوره دوم و جدیدتر ناشی از پیشرفت در رویکرد سازنده و تفسیری (با اولویت متناظر برای روش‌های اکتشافی کیفی) است.

بر این اساس، از دیدگاه نئوپوزیتیویستی، ارزیابی با دو پیش‌فرض مشخص شده است. اولین فرض این است که برای ارزیابی معتبر اینکه آیا و تا چه حد اهداف سیاست (به عنوان پیامدهای مورد نظر) با تغییرات قابل مشاهده در دنیای واقعی محقق شده است، لازم است از قبل مشخص شود که مقاصد و اهداف سیاسی برنامه چیست. در این دیدگاه، نیت «یک» نهاد یا کنشگر مرتبط در رأس قرار دارد.

دوم، به منظور شناسایی روابط علی بین تغییرات مشاهده شده و سیاست/برنامه مورد بررسی، گزاره‌های معتبر تنها از طریق کاربرد پوزیتیویستی طرح‌های تحقیقاتی کمی،

(شبه - تجربی) قابل دستیابی است. با این حال، با وجود تسلط طولانی این الگوی پژوهشی، مشکل تبدیل این مقدمات به عمل ارزیابی برای بسیاری از ناظران آشکار بود. به عنوان مثال، در شناسایی اهداف مرتبط، مسائل جدی مطرح می شود: (۱) شناسایی اهداف و مقاصدی که به عنوان میله اندازه گیری عمل می کنند، دشوار است، زیرا اغلب به عنوان اهداف «مجموعه‌ای» به سختی به شاخص‌های عملیاتی و قابل اندازه گیری ترجمه می شوند. (۲) داده‌های تجربی خوب برای پر کردن شاخص‌ها به سختی به دست می آیند، و هر چه یک شاخص معنی دارتر باشد، به دست آوردن داده‌های قابل اجرا دشوارتر است. (۳) هر چه بعد هدف دورتر (و اغلب مرتبط‌تر) باشد، عملیاتی کردن و اثبات تجربی آن دشوارتر می شود. (۴) ردیابی عوارض جانبی و عواقب ناخواسته دشوار است.

علاوه بر این طرح‌های تحقیقاتی قوی از نظر روش شناختی (نیمه تجربی، کنترل شده، سری‌های زمانی، و غیره) اغلب قابل اجرا نیستند، و حداقل از نظر روش شناختی رضایت بخش نیستند. در اینجا فرد باید شرایط ثابت بودن تمام عوامل دیگر را رعایت کند. (که از محورهای طراحی شبه تجربی بر روی آن‌ها استفاده می شود) اگر غیرممکن نباشد، دشوار است. در حالی که استفاده از روش‌های کمی بر اساس نیاز روش شناختی «موارد زیاد (N بزرگ)، متغیرهای کم»، در موقعیت پژوهشی در دنیای واقعی، اغلب به صورت دسته برعکس است: موارد کم (N کوچک)، متغیرهای زیاد (احتمالاً تأثیرگذار). این مشکلات استفاده از روش‌های کمی را رد می کند و در عوض به صورت کیفی پیش می رود. و در نهایت، استفاده از روش‌های سری زمانی (قبل/پس از طراحی) اغلب محدودیت‌های محدودی دارد، زیرا داده‌های «قبل» اغلب در دسترس نیستند و یا قابل تهیه نیستند. در مرحله دوم، پارادایم تحقیق غالباً رایج به مدت طولانی در دو مورد مرتبط به هم مورد انتقاد قرار گرفته است. به عنوان مثال، این فرض استاندارد که ارزیابی باید چارچوب مرجع خود را قبل از هر چیز در نیت سیاست نهاد(ها) یا بازیگران سیاسی مربوطه جستجو کند، با پیشرفت‌های مکتب فکری سازه انگاری/ تفسیری - اگر نگوئیم درهم شکسته - متزلزل شده است. این از پرسش بر اساس مبانی معرفت‌شناختی

تشخیص معتبر "یک" قصد یا هدف مرتبط حمایت می‌کند و به جای آن (اغلب) خواستار شناسایی کثرتی از دیدگاه‌ها، علایق و ارزش‌ها است. به عنوان مثال، دنیل استوفلیم (۱۹۸۳) در پیشبرد مفهومی از ارزیابی به نام "مدل CIPP" که در آن  $C = I + P = P$  فرایند،  $P =$  محصول است تأثیرگذار بوده، در میان چهار مؤلفه، عنصر «زمینه» (با تمرکز بر سؤالاتی مانند: اهداف برنامه چیست؟ آیا آنها نیازهای شرکت کنندگان را منعکس می‌کنند؟) هدفش این است که توجه ارزیاب را از ابتدا به نیازها (و علایق) شرکت کنندگان در برنامه مورد بررسی (و پیامدهای هنجاری زیربنایی آن) معطوف کند. این استدلال کلی در فرمول‌بندی‌های مختلفی مانند «ارزیابی پاسخگویی»، «ارزیابی مشارکت» یا «ارزیابی ذی‌نفعان» بیان شده است. از نظر روش شناختی، بحث نظریه یادگیری ساخت‌گرایی با (دوباره) به دست آوردن زمینه برای روش‌های کیفی - هرمنوتیکی در ارزیابی ارتباط تنگاتنگی داشته است. گوبا و لینکلن (۱۹۸۹) این پیشرفت را "ارزیابی نسل چهارم" نامیده‌اند.

در حالی که حدود بیست سال پیش خطوط نبرد بین اردوگاه‌های فکری به شدت ترسیم شده بود، اما اکنون نبرد از آن زمان کمتر شده است. از یک سو، بینش‌های معرفت‌شناختی، مفهومی و روش‌شناختی ایجاد شده در بحث برساخت‌گرایی پذیرفته شده و جدی گرفته می‌شود، دستور برای هرچه بیشتر نزدیک‌تر شدن به «عینی» شدن ارزیابی همچنان یک هدف اصلی باقی مانده است. مفهوم "ارزیابی واقع‌بینانه" که توسط پائوسون و تیلی (۱۹۹۷) فرموله شده است، خود را در خدمت این هدف قرار می‌دهد. علاوه بر این، به طور گسترده توافق شده است که هیچ «جاده پادشاهی (مسیر ویژه)» در طراحی روش شناختی تحقیق ارزیابی وجود ندارد. در عوض، باید به کثرت روش‌ها اذعان کرد. انتخاب و ترکیب مجموعه خاص و ترکیبی از روش‌ها به سؤال ارزیابی‌ای که باید به آن پاسخ داده شود و همچنین چارچوب زمانی و منابع مالی و پرسنلی موجود بستگی دارد.

## • تحقیق ارزیابی: تحقیقات پایه، تحقیقات کاربردی و تحقیقات قراردادی

ظهور و گسترش تحقیقات ارزیابی از اواسط دهه ۱۹۶۰ تأثیر قابل توجهی بر چشم انداز و جامعه تحقیقاتی علوم اجتماعی داشته است. در اصل عرصه تحقیقات علوم اجتماعی تحت تسلط تحقیقات آکادمیک (پایه) بود که عمدتاً در دانشگاه‌ها قرار داشت و توسط آژانس‌های مستقل تأمین مالی می‌شد. حتی زمانی که به یک جهت گیری سیاست کاربردی نیاز داشت، تحقیقات علوم اجتماعی اساساً به فرمول آکادمیک/پایه متعهد باقی ماند. در مقابل، تحقیقات ارزیابی، تا آنجا که به عنوان «تحقیقات قراردادی» انجام می‌شود و توسط یک نهاد سیاسی یا اداری تأمین می‌شود، شامل تغییر از «محیط‌های آکادمیک به محیط‌های کارآفرینی» است. (Freeman, and Solomon 1981)

تحقیقات علوم اجتماعی دانشگاهی، معمولاً مبتنی بر دانشگاه، بر چهار امر ضروری استوار شده است. اول تعهد به جستجوی حقیقت به عنوان هدف محوری و معیار پژوهش علمی بوده است. دومی مربوط به استقلال درون علمی در انتخاب موضوع و روش‌های تحقیق آن است. سومین مورد، تأمین مالی مستقل بوده است، چه از طریق منابع دانشگاهی و چه از طریق تأمین مالی مبتنی بر بررسی هم‌تا توسط بنیادهای تحقیقاتی مانند بنیاد ملی علوم. مؤلفه نهایی، آزمون کیفیت یافته‌های پژوهشی در یک بحث علمی باز و بررسی همتایان بوده است.

در حالی که علوم اجتماعی کاربردی هنوز استقلال و خودمختاری تحقیقات علوم اجتماعی را حفظ کرده است، تحقیقات قراردادی، که در حال حاضر وسیله اصلی تحقیقات ارزیابی را تشکیل می‌دهد، به فرمول کاملاً متفاوتی وابسته است. با یک سفارش دهنده/تولیدکننده یا مصرف کننده/پیمانکار تشخیص داده می‌شود: «مصرف کننده آنچه را که می‌خواهد می‌گوید، پیمانکار آن را انجام می‌دهد (اگر بتواند) و مصرف کننده هزینه‌اش را می‌پردازد» (به نقل قول از لرد روچیلد). از این رو، «درخواست پیشنهاد» (RFP) که از طریق آن آژانس راه‌اندازی به پیمانکاران احتمالی (در مناقصه عمومی، مناقصه انتخابی یا مستقیم) مورد خطاب قرار می‌دهد، به طور کلی سؤالاتی را که باید به آنها پاسخ داده شود و چارچوب زمانی که در دسترس قرار گیرد، تعریف و مشخص

می‌کند. در طرح پیشنهادی پروژه، پیمانکار برنامه تحقیقاتی خود را در چارچوب پارامترهای تعیین شده توسط مشتری توضیح می‌دهد و پیشنهاد مالی خود را ارائه می‌دهد که معمولاً بر اساس هزینه‌های پرسنل به اضافه فرمول هزینه‌های بالاسری محاسبه می‌شود.

بنابراین، هنگامی که تحقیقات ارزیابی توسط دولت سفارش و تأمین مالی می‌شود، با سه چالش اساسی مرتبط با موضوع، سؤالات اصلی و روش‌های تحقیق روبرو می‌شود. در تحقیقات قراردادی، برخلاف تحقیقات ارزیابی سنتی، این ملاحظات توسط آژانسی که ارزیابی را سفارش می‌دهد، تعیین می‌شود. همچنین، آژانس با تأمین بودجه، استقلال محققین را نیز به خطر می‌اندازد («کسی که دستمزد فلوت زن را می‌پردازد، آهنگ و صدا را تعیین می‌کند»). و در نهایت، یافته‌های تحقیقات سفارشی اغلب به صورت مخفیانه نگهداری می‌شوند، یا حداقل منتشر نمی‌شوند، بنابراین از بحث عمومی و همتایان دور می‌زنند. در واقع، تحقیقات قراردادی در معرض یک رانش معرفتی و یک فرایند استعماری است که در آن ارزیابان ممکن است «چارچوب مفهومی و چشم‌انداز (دلخواه)» نهادها و بازیگران سیاسی و اداری را که مأمور ارزیابی آنها هستند، اتخاذ کنند. (Elzinga 1983, 89).

در مواجهه با چالش‌های یکپارچگی فکری و صداقت تحقیقات قراردادی، ابتکاراتی توسط ارزیاب‌های حرفه‌ای برای تدوین استانداردهایی صورت گرفته است که می‌تواند آنها را در کار قراردادی‌شان، به‌ویژه در مذاکراتشان با «مشتریان»‌شان راهنمایی کند. به‌عنوان مثال، می‌توان در اینجا به اصول راهنمای ارزیابی، که در سال ۱۹۹۵ توسط انجمن ارزیابی آمریکا در سال ۱۹۹۵ تصویب شد، اشاره کرد. در میان پنج اصل آن، اصول قواعد درستی و صداقت تحقیق بسیار بزرگ نوشته شده است.

## • حرفه‌ای‌سازی

در این میان، ارزیابی در بسیاری از کشورها به فعالیت و شغل گروه و جامعه‌ای متشکل از پژوهشگران و تحلیلگران متخصص تبدیل شده است که حرفه‌ای شدن فزاینده آنها در شکل‌گیری انجمن‌های حرفه‌ای، ظهور نشریات حرفه‌ای و در ورود ارزیابی به عنوان موضوع در دانشگاه و آموزش حرفه‌ای سازی دیده می‌شود.

در مورد پایه‌گذاری انجمن‌های حرفه‌ای، انجمن ارزیابی آمریکا (AES) که در سال ۱۹۸۶ از طریق ادغام دو سازمان ارزیابی کوچک‌تر شبکه ارزیابی و جامعه تحقیقات ارزیابی تشکیل شد، نقشی پیشرو و نمونه را بر عهده گرفت. AES تا سال ۲۰۰۳، بیش از سه هزار عضو داشت. یک محصول مهم، تدوین منشور اخلاق حرفه‌ای فوق‌الذکر بود که در «اصول راهنمایی برای ارزیابان» که توسط AES در سال ۱۹۹۵ به تصویب رسید، لحاظ شد. در اروپا، انجمن ارزیابی اروپا در سال ۱۹۸۷ تأسیس شد و انجمن‌های ملی ارزیابی نیز به دنبال آن تأسیس گردید که انجمن ارزیابی بریتانیا اولین انجمن از آنها بود. در این میان، اکثر آنها قوانین حرفه‌ای اخلاقی را نیز تصویب و اتخاذ کرده‌اند که بیانگر قصد و عزم برای تثبیت و تضمین ارزیابی به عنوان یک شغل و حرفه جدید است. یکی دیگر از شاخص‌های مهم نهادینه‌سازی حرفه‌ای ارزیابی، میزان تبدیل شدن ارزیابی به موضوع بازار نشریات سریع‌الانتشار است. این بخصوص، شامل انتشار مجلات حرفه‌ای است که اغلب در ارتباط نزدیک با انجمن ملی مربوطه هستند. بدین ترتیب انجمن ارزیابی آمریکا دو نشریه دارد: مجله آمریکایی ارزیابی و سری تک‌نگاری دستورات عمل‌های جدید برای ارزیابی. در اروپا، مجله Evaluation با همکاری انجمن ارزیابی اروپا منتشر می‌شود. علاوه بر این، تعدادی از مجلات ملی ارزیابی (به زبان‌های ملی مربوطه) در چندین کشور اروپایی راه‌اندازی شده است. همه اینها به عنوان منابع مفید اطلاعات در مورد موضوع تحقیق ارزیابی عمل می‌کنند.