



طراحی سیستم ارزیابی کمی

پایان نامه ها و رساله ها

(بررسی نوشتجات)

طراحی سیستم ارزیابی کمی

پایان نامه ها و رساله ها

(بررسی نوشتجات)

مجری

یعقوب حسین زاده

ناظر

دکتر سید امید فاطمی



وزارت علوم، تحقیقات و فناوری
پژوهشگاه اطلاعات و مدارک علمی ایران
پاییز 1387



وزارت علوم، تحقیقات و فناوری
پژوهشگاه اطلاعات و مدارک علمی ایران
تهران - صندوق پستی 1371-13185

www.irandoc.ac.ir

طراحی سیستم ارزیابی کمی پایان نامه ها و رساله ها (بررسی نوشتجات)

یعقوب حسین زاده

Hosseinzadeh@irandoc.ac.ir

فهرست مطالب

1 پژوهش.....	1
1-1- مقدمه.....	2
2-1- نقش پژوهش در توسعه.....	2
3-1- عملکرد پژوهش.....	3
4-1- روندهای سرمایه گذاری در پژوهش.....	3
2 ارزیابی پژوهش.....	5
1-2- تاریخچه.....	6
2-2- ارزیابی عملکرد پژوهش.....	9
1-2-2- تعریف.....	10
3-2- انواع ارزیابی.....	12
1-3-2- خودارزیابی.....	14
4-2- الگو برداری.....	15
5-2- اهداف و ضرورت ها.....	15
6-2- نقش ارزیابی در دانشگاه ها و پژوهشگاه ها.....	16
7-2- نقش ارزیابی در سیاست گذاری.....	17
8-2- ترکیب ارزشیابی آموزشی و پژوهشی.....	18
9-2- ارزیابی در کشورهای جهان سوم.....	19
10-2- اصول زیربنایی ارزیابی.....	23
11-2- ارزیابی به عنوان یک سیستم.....	24
1-11-2- چهار سیستم ارزیابی.....	25
12-2- واحدهای مورد ارزیابی.....	26
1-12-2- افراد.....	26
2-12-2- سازمان ها.....	27
3-12-2- پرونده های پژوهشی.....	28
13-2- انواع ارزیابی.....	29
1-13-2- ارزیابی کمی.....	29
1-1-13-2- چگونگی کاربرد شاخص های کمی.....	31
2-1-13-2- نقش شاخص های کمی.....	31
2-13-2- ارزیابی کیفی.....	34
1-2-13-2- اساس چارچوب کیفی پژوهش.....	34
14-2- روش ها و سبک های ارزیابی.....	37
15-2- ابعاد ارزیابی.....	42
1-15-2- ارزیابی فرآیند.....	43
2-15-2- ارزیابی نتایج.....	44
3-15-2- ارزیابی هزینه.....	45

46.....	2-15-4- ارزیابی تطبیقی
46.....	2-15-5- ارزیابی تعمیم پذیری
47.....	2-16- حوزه های ارزیابی
48.....	2-16-1- ارزیابی برنامه
48.....	2-16-2- ارزیابی افراد
49.....	2-16-3- ارزیابی عملکرد
49.....	2-16-4- ارزیابی محصول
50.....	2-16-5- ارزیابی خط مشی
50.....	2-16-6- ارزیابی دارایی
51.....	2-16-7- ارزیابی طرح پیشنهادی
52.....	2-17- مدل های ارزیابی
52.....	2-17-1- مدل Transdisciplinary ترا رشته ای
56.....	2-17-2- مدل RAE
56.....	2-17-3- مدل ATP
57.....	2-17-4- بررسی انواع مدل های جهانی و مقایسه آنها با یکدیگر
59.....	2-17-4-1- ابزار
59.....	2-17-4-2- سنجها
59.....	2-17-4-3- روش تحلیل
60.....	2-18- چالشهای ارزیابی پژوهش
64	3 ارزیابی پژوهش در حوزه ها و کشورهای جهان
65.....	3-1- ارزیابی پژوهش در حوزه های علمی
65.....	3-1-1- ارزیابی در حوزه های علوم انسانی و اجتماعی و هنر (HASS)
67.....	3-1-2- ارزیابی در حوزه های انسانی و هنرهای خلاق
68.....	3-1-3- ارزیابی در حوزه اقتصاد
69.....	3-1-4- ارزیابی در حوزه علوم سیاسی
70.....	3-1-5- ارزیابی در حوزه موزه و باستان شناسی
70.....	3-2- روشهای ارزیابی پایان نامه ها و رساله ها در دانشگاه ها
71.....	3-2-1- بعد فرهنگی
71.....	3-2-2- بعد علمی
72.....	3-2-2-1- ارزیابی ساختاری
72.....	3-2-2-2- ارزیابی روشی
73.....	3-2-2-3- ارزیابی محتوایی
73.....	3-2-2-4- ارزیابی نگارشی
77.....	3-3- جمع بندی
78	منابع و مأخذ

1

پژوهش

۱-۱- مقدمه

این گزارش روشهای ارزیابی کمی و کیفی پژوهش توضیح داده شده است. همچنین برنامه‌ها و نهادهایی که در کشورهای مختلف به ارزیابی مراکز دانشگاهی و تحقیقاتی می‌پردازند معرفی می‌گردند. هدف این است که چشم‌اندازی از وضعیت موجود در جهان به دست آید و بتوان با الگوبرداری از آنها سیستمی برای کشور طراحی نمود.

ارزیابی می‌تواند از طریق افزایش بهره‌وری و اثربخشی پژوهش به بهبود عملکرد آن کمک کند. ابزارها و روش‌های کمی و کیفی بسیاری نیز برای ارزیابی میزان اثربخشی پژوهش وجود دارد که در این گزارش بدان‌ها اشاره خواهد شد. (Davis and Carden 1998).

۱-۲- نقش پژوهش در توسعه

یکی از جدی‌ترین مشکلات بر سر راه ارزیابی پژوهش یا پژوهشگر، مسأله تعریف پژوهش است. (Davis and Carden 1998)

پژوهش فعالیتی حقیقت‌جو^۱ است که در پیشرفت دانش مشارکت می‌کند. هدف آن توصیف، پیش‌بینی، یا توضیح پدیده‌های جهانی است و توسط افرادی با تخصص و مهارت بالای اجرا و مدیریت می‌شود. هر چند که برخی مخالف این تعریف بوده‌اند، و ادعا می‌کنند که پژوهش باید در دو دسته

¹truth-seeking activity

بنیادی و کاربردی به طور مجزا تعریف شود. با این حال هر دو نوع پژوهش، در پیشرفت دانش نقش دارند و تفکیک قائل شدن بین این دو، سؤال برانگیز خواهد بود. بنابراین این تعریف پیشنهادی به سادگی و بدون پیچیدگی، ماهیت اصلی پژوهش را در برمی گیرد (Coryn 2007).

۱-۳- عملکرد پژوهش

عملکرد به بهره‌وری^۱، خلاقیت^۲، و پیشرفت^۳ مربوط است. بررسی روند ارزیابی‌هایی که از عملکردهای پژوهشی صورت گرفته است، نشان می‌دهد سه موضوع یا فاکتور اصلی در آنها وجود دارد: اعتبار^۴، بازدهی^۵، و تأثیر^۶. هرچند بیشتر الگوهای ارزیابی سعی دارند نسبت به الگوهای قبلی جهت‌گیری کمتری داشته باشند، اما هیچ کدام از آنها در میان پژوهشگران و مدیران به عنوان یک استاندارد برای ارزیابی عملکرد پژوهشی شناخته نشده‌اند (Avital and Collopy 2001).

۱-۴- روندهای سرمایه‌گذاری در پژوهش

روندهای سرمایه‌گذاری در تحقیق و توسعه در دهه ۱۹۹۰ و همزمان با گسترش هزینه‌های تحقیق و توسعه به حد متوسطی رسید. پژوهشهای دانشگاهی افزایش یافت و بدین ترتیب دانشگاه‌ها با موفقیت از جایگاه خود در بافت سیستم‌های ملی نوآوری که در حوزه گسترده‌تر جهانی‌سازی قرار دارد، دفاع می‌کردند. مهم‌تر این که، دانشگاه‌ها، پژوهش‌های بنیادی‌ای را ارائه می‌کردند، که سرمایه‌گذاری آن توسط عموم جامعه صورت می‌گرفت. اما در همین زمان، چالش‌های این پژوهش‌ها بیشتر و متحول شدند.

¹ productivity

² creativity

³ achievement

⁴ reputation

⁵ yield

⁶ influence

برای نمونه، دانشگاه‌ها و پژوهش‌های آنها بیشتر و بیشتر با محیط اجتماعی ارتباط پیدا می‌کردند (Campbell 2005).

2

ارزیابی پژوهش

۲-۱- تاریخچه

ادعا می‌شود که ارزیابی نظام‌مند و صریح از پژوهش، در طفولیت خود به سر می‌برد، و تنها طی دهه‌های اخیر به عنوان یک حوزه مطرح شده است. با این وجود، ارزیابی مفهومی^۸ در پژوهش امری است که قدمتی طولانی دارد و با اطمینان می‌توان گفت ارزیابی مفهومی اساس ارزیابی صریح است. شواهد وجود ارزیابی مفهومی را می‌توان در پیشینه‌های ثبت شده قبل از قرن دوم پیش از میلاد یافت. از جمله آنها می‌توان به نوشته‌های دانشمندان یونانی اشاره کرد. ارشمیدس ریاضی‌دان در یکی از این نوشته‌ها، مطلبی را آورده که در آن آشکارا است کوششی برای ارزیابی مفهومی و درون‌نگرانه مشهود است (Corryn 2007):

«برخی چیزها در ابتدا برای من با روشی مکانیکی روشن می‌شدند؛ اما بعد از آن، با روش هندسی قابل حل بود. زیرا بررسی با روش مذکور راه حل واقعی نبود. البته برای ما آسان‌تر آن بود که به جای وارد عمل شدن بدون دانش، ابتدا؛ با روش‌های موجود آشنا شویم و دانش مسائل خود را بیابیم».

ارزیابی ویژگی اساسی انسان است و شاید مهم‌ترین و پیشرفته‌ترین فرآیند شناختی در سیستم منطقی انسان باشد که نزدیک به دو هزار سال درباره آن بحث شده است (Coryn 2007).

⁸ implicit evaluation

استفاده از شاخص‌های علم‌سنجی در ارزیابی پژوهش، از دهه 1960 و 1970 ابتدا در ایالات متحده و سپس در کشورهای مختلف اروپایی ظهور کرد. قبل از این زمان، ارزیابی پژوهش، جز از طریق سیستم همترانخوانی و از طریق شاخص‌های اقتصادی مورد استفاده در سیستم‌های ملی سطح بالا، شکل نگرفته بود. شاخص‌های اقتصادی (مثل درصد تولید ناخالص داخلی صرف شده در تحقیق و توسعه) به طور جهانی توسط سازمان توسعه و همکاری اقتصادی (او.ای.سی.دی.)⁹ در پاریس ایجاد شد. برای مثال، دستورالعمل فراسکاتی برای سنجش فعالیت‌های علمی و فنی می‌تواند پاسخی به اهمیت اقتصادی افزاینده علم و فناوری محسوب شود که در آمارهای اقتصادی طی دهه پنجاه آشکار شد (Leydesdorff 2005).

ایده اصلی استفاده از تحقیقاتی علمی، در نتیجه جنگ جهانی دوم پدید آمد. تأثیری که جنگ بر علم و فناوری از طریق توسعه دانش مدار و پژوهش هدف گرا در طی جنگ جهانی دوم ایجاد کرد، اتخاذ سیاستهای نوین علم و فناوری را در شرایط صلح ضروری ساخت (Leydesdorff 2005).

راه‌اندازی ماهواره «اسپانیتیک» توسط اتحاد جماهیر شوروی در سال 1957 علت اصلی تغییر سیاستگذاری در علم بود. به نظر می‌رسید شوروی‌ها در زمینه پژوهش هدف گرا موفق‌تر از غرب بودند. اما در ادامه غربی‌ها نیز عقب نماندند. که سازماندهی گسترده سیستم پژوهش امریکا یکی از نتایج آن بود. ناسا و آژانس طرح‌های پژوهشی پیشرفته (آرپا)¹⁰ نیز در واکنش به طرح اسپانیتیک و تهدید نظامی آن ایجاد شدند. (Leydesdorff 2005)

در همین دوران، رشد اقتصادی کشورهای غربی رفته رفته با فاکتورهای سنتی اقتصادی چون زمین، نیروی کار، و سرمایه قابل توجیه نبود. همراه با همکاری نظامی ناتو، سازمان توسعه و همکاری اقتصادی در سال 1961 به منظور سازماندهی و هماهنگ نمودن سیاست‌های علمی و فنی در بین کشورهای عضو

⁹ OECD

¹⁰ Advanced Research Projects Agency (ARPA)

یعنی کشورهای صنعتی پیشرفته ایجاد شد. در نتیجه این امر، در سال 1963 دستورالعمل فراسکاتی ایجاد شد که در آن پارامترهایی برای نظارت آماری علم و فناوری بر مبنای مقایسه تعریف شدند. اما مقایسه بین کشورها، سؤلهایی در مورد نقاط قوت و ضعف در اصل این امر مطرح کرد. در نیمه دوم دهه شصت، سیاستهای علم و فناوری در کشورهای عضو «او.ای.سی.دی.» نمایان شدند. (Leydesdorff 2005)

برای نمونه، آمارها نشان می‌داد که در علم فیزیک بعد از جنگ جهانی دوم در زمینه سازماندهی علایق؛ هم در بین کشورهای مختلف و هم در سطح همکاری‌های جهانی موفقیت‌هایی حاصل شده است. جوامع علمی دیگری چون میکروبیولوژی ادعا کردند که میزان بودجه اختصاص داده شده به پیشرفت‌های نوین و در کل میزان بودجه علمی افزایش یافته است. در فاصله بین سال‌های 1965 – 1975 ابزار برتر برای رویارویی با چنین مسائلی در سطح ملی، تمایز قایل شدن در افزودن میزان بودجه‌ها در سطح رشته‌ها بود. به طور خلاصه، تمرکز بر شاخص‌های دروندادی (مالی) باقی ماند، و این در حالی بود که سیستم تکیه بر همترازخوانی برای تصمیم‌گیری کم اهمیت‌تر در سطوح پایین‌تر بود (Leydesdorff 2005).

ریشه توجه به سنجش ارتباطات علمی، ارزیابی پژوهشی نیست. در فاصله دهه‌های 1950 تا 1960، جامعه علمی بیش از پیش دریافت که منابع و اطلاعات علمی پس از جنگ جهانی دوم، به طرز کنترل‌ناپذیری رو به افزایش است. «نمایه استنادی علوم» یا آی.اس.آی. که توسط مؤسسه گارفیلد تولید می‌شد، به زودی ابزاری برای مشخص نمودن استانداردها شناخته شد. معرفی تدریجی شاخص‌های بروندادی مثل شمارش اسنادها و انتشارات، می‌توانست هم در سطح اجتماعی قانونی شود، و هم در سطح داخلی؛ زیرا با کمک آن سیاست‌گذاران و مدیران علمی می‌توانستند از استدلال‌های کارایی اقتصادی استفاده کنند، از طرفی کنترل کیفیت در بین رشته‌ها برای قانونی شدن، مشکل بود، مگر اینکه استانداردهای مشخصی علاوه بر فرآیند همترازخوانی قابل دسترس باشند.

هنری اسمال روشی را برای تهیه نقشه علمی مبتنی بر استندهای گروهی به مقالات علمی پیشنهاد کرد. در حالی که رویکرد اسمال، سعی داشت متخصصان را در ساختارهای قانونی گرد هم آورد، اما نارین بر ساختارهای سلسله‌مراتبی که به روش بالا-پایین (عمودی) عمل می‌کرد، تمرکز داشت. برای رده‌بندی علوم به لحاظ خوشه‌های نشریات و وزن دهی انتشارات طبق استاندارد، طرح پژوهشی علم‌سنجی در سطح وسیع به اجرا در می‌آمد. بررسی‌های آمریکایی‌ها در دهه 1970 بر سازماندهی منابع علمی متمرکز بود، اما کاربرد شاخص‌های گفته شده برای ارزیابی پژوهش در سطح سازمانی طی دهه 1980 در اروپا گسترش یافت.

در پی انتشار نتیجه بررسی مارتین و آبروین از عملکرد پژوهش‌های ایستگاه‌های مختلف مراکز رادیویی ستاره‌شناسی، ایده ارزیابی واحدهای نهادی در بین سیاستگذاران پیاده شد. دانشگاه لیدن در هلند در این میان پیشرو بود و پس از آن، این ایده از سال 1992 برای سرمایه‌گذاری پژوهش‌های دانشگاهی در انگلستان نیز تعمیم داده شد.

دیگر کشورهای اروپایی از مدل انگلستان پیروی نکردند، اما طی دهه نود فشارها برای رتبه‌بندی استنادها و انتشارات در اقدامات ارزشیابی غالب شد. برای مثال، بعد از اتحاد دو آلمان در سال 1990 ارزیابی گسترده‌ای از آثار پژوهشی آلمان شرقی در سیاست علمی این کشور قرار گرفت (Leydesdorff, 2005).

۲-۲- ارزیابی عملکرد پژوهش

بررسی 67 پژوهش در حوزه بهره‌وری و عملکرد پژوهش‌های منتشر شده طی سال‌های 1974 و 1998 نشان داد در این حوزه دو روش ارزیابی و توصیفی¹¹ وجود دارد. بررسی‌های ارزیابی بر سنجش،

¹¹ Explanatory

مقایسه، و رتبه‌بندی عملکرد پژوهشگران متمرکز است و هر چند برخی اوقات بر اساس همترازخوانی انجام می‌شود، اما غالباً در آنها از روش‌های کتابسنجی ناشی از شمارش استنادها یا مدارک استفاده می‌شود. به عکس، مطالعات توصیفی به عوامل تقویت‌کننده بهره‌وری پژوهش و ارزیابی میزان اثرگذاری آنها بر عملکرد پژوهشگران می‌پردازند. در این روش، مشخصات بهره‌وری پژوهش بررسی، و به عوامل سازمانی، مالی، همکاری، حرفه‌ای، فردی و آماری دسته‌بندی می‌شود. بررسی متون نشان می‌دهد که بیشتر بررسی‌های ارزیابی در سطح سازمانی یا مؤسسه‌ای اجرا شده، در صورتی که بررسی‌های توصیفی در سطح فردی صورت گرفته است. این یافته‌ها خیلی منطقی به نظر نمی‌رسند، زیرا در کل افراد هستند که ارزیابی می‌شوند، در صورتی که عوامل تقویت‌کننده بهره‌وری قابل کنترل عمدتاً سازمان‌ها هستند. علاوه بر این، هر چند این دو روش در اصل، مکمل هم هستند، اما به ندرت در یک بررسی از هر دو روش استفاده می‌شود (Avital and Collopy 2001).

۲-۱-۲- تعریف

ارزیابی پژوهش، فرآیندی است که بر اساس تحلیل انتقادی داده‌ها و اطلاعات صورت می‌گیرد و در طی آن شایستگی پژوهش مورد قضاوت قرار می‌گیرد. این فرآیند با این دو فرض همراه است (CIVR 2006):

- روش‌شناسی‌ها و معیارهای از پیش تعیین شده؛
 - ارزیاب‌ها، که خارج از ساختارهای تحت ارزیابی قرار دارند و با محصول مورد ارزیابی نیز سر و کار ندارند.
- ارزیابی نظام‌مند و هدف‌مند یک محصول یا خدمت اطلاعاتی؛ به طراحی، یا نحوه طراحی و اجرا و بروندها و یا تأثیرات آن مربوط می‌شود. ارزیابی قضاوت، سنجش، یا تعیین ارزش یا کیفیت چیزی است. همچنین مستلزم فرآیندی از بررسی داده‌ها و استنتاج از آن است. در صدد است تا چستی،

چگونگی، و چرایی وقوع رخدادها را توضیح دهد. هدف آن نیز ارتقاء خودیادگیری، غنی سازی، افزایش ظرفیت و پایدارسازی محصول یا خدمت اطلاعاتی است.

نظارت: که در ارزیابی ها استفاده می شود، فرآیندی مستمر از گردآوری و تحلیل داده ها برای شاخص های عملکردی است، تا عملکرد خدمت یا محصول اطلاعاتی را با نتایج مورد انتظار، مقایسه کند. نظارت به تعریف مشکلات در طی اجرای ارزیابی محصول یا خدمات اطلاعاتی کمک می کند .

ممیزی/رسیدگی: در لفظ به معنای کنترل است و فرم رسمی و بیرونی ارزیابی محسوب می شود. تعریف کمیسیون اروپا از ممیزی ارزیابی: ممیزی ارزیابی عبارت است از قانونی و مقرراتی بودن هزینه ها و درآمدهای یک پروژه و مؤثر و به صرفه بودن آنها.

مسئله ارزیابی پژوهش، از ارزیابی کیفیت پژوهش فاصله گرفته است و طیفی از مسائل مالی عمومی در حوزه های علم و فناوری را در بر می گیرد، نظیر: منطق پژوهش، ربط، ساختار، پیوندها، و مدیریت آن. همچنین مسائل برون دادی مانند نیل به اهداف و ضریب نفوذ، و مسائل دیگر. بافت سازمانی مجموعه ای بحرانی از متغیرها برای خود در نظر می گیرد، زیرا هر نوع از سازمان منطق رفتاری خاص خود را دارد. (Davis and Carden 1998)

اصطلاح «میزان اثربخشی پژوهش» اشاره به این ویژگی توسعه پژوهش دارد که بین نتایج واقعی و دریافت شده پژوهش از یک سو و پیشرفت ناشی از آن در میان انتشاراتی پیوند ایجاد می کند که بر افرادی که مسائل پژوهشی دارند و راه حل هایی را از طریق سیاست ها و پژوهش های جدید طراحی می کنند، تأثیر می گذارد. بنابراین اثربخشی پژوهشی دربرگیرنده (و نه محدود به) مطالعه «هزینه بهره وری» تلاش ها، بررسی فرآیند، اثرات، میزان نفوذ انتقال تکنولوژی، بررسی اثرات متقابل بین تصمیم گیران، «پژوهشگران اصولی»، و مسائل سیاسی رایج که اغلب توسط تغییرات و پیشرفتهای علمی تأثیر می پذیرند، و بررسی میزان کاربرد پژوهش و توسعه در بخش های سودآور اقتصادی است.

اثربخشی به عنوان اصلی از طراحی پژوهش و اجرای آن، انعکاس دهنده هدف اولیه برای تحویل نتایج کاربردی به گروه‌های مخاطب است. در مواردی چون یادگیری درس‌های آموخته شده و پیشرفت‌های حاصل در انتخاب مسأله پژوهشی، طراحی برنامه کاری، نوآوری در ارائه محصول یا خدمات، و تحویل محصولات و خدمات به کاربران، ارزیابی سبب افزایش بهره‌وری می‌شود. ... اما، کاربرد استراتژی‌های پژوهشی مستلزم این است که نتایج آن‌ها به شکل محصول یا خدمت درآید و دیگر اینکه به صورت اطلاعات درباره کاربران، عوامل بالقوه بازاریابی، محیط‌های سیاسی و اقتصادی، مجراهای توزیع، پیش‌برنده‌ها، رفتار کاربر، و قیمت دسته بنده شوند. معمولاً، اهداف پژوهش‌های «فنی» نسبت به اهداف پژوهش‌های «نظام‌مند» یا «اجتماعی» قابل دسترس‌تر هستند، و کاربرد نتایج پژوهشی به درگیری کاربر نهایی و مؤسسه‌هایی که در طراحی و اجرای پروژه به لحاظ فنی حمایت می‌کنند، بستگی دارد. (Davis and Carden 1998)

۲-۳- انواع ارزیابی

رویکردهای مختلفی برای ارزیابی وجود دارد. بر اساس ارتباط بین چرخه پروژه و محصول یا خدمت اطلاعاتی، دو نوع عمده ارزیابی عبارتند از:

- ارزیابی تکوینی: در حین و بعد از مراحل برنامه ریزی و اجرای پروژه صورت می‌گیرد. این نوع از ارزیابی کمک می‌کند تا چگونگی بهبود و ارتقاء مدیریت و پیاده سازی محصول یا خدمات اطلاعاتی درک شود. آنهایی که درگیر تولید محصول یا خدمت اطلاعاتی هستند، از این نوع ارزیابی سود می‌برند.

- ارزیابی تلخیصی: در حین و بعد از اتمام مرحله پروژه اطلاعاتی صورت می گیرد. این نوع ارزیابی کمک می کند تا مشخص شود آیا پروژه اطلاعاتی به انتظارات و اهداف از پیش تعیین شده خود رسیده است یا خیر.

دسته بندی های دیگری نیز از ارزیابی وجود دارند:

ارزیابی هایی که در زمان های مختلف در چرخه پروژه انجام می شوند:

1. **ارزیابی پیش بینی شده:** مانند بررسی امکانسنجی این نوع ارزیابی در انتهای مرحله

برنامه ریزی پروژه انجام می شود. ارزیابی طرح پیشنهادی پروژه از نظر ربط، امکان پذیری،

اثرات بالقوه یا فواید مورد انتظار و همچنین میزان پایایی پروژه مورد تأکید قرار می گیرد. در

اینجا کنترل ها برای بررسی ارزیابی میزان برطرف شدن نیازهای سهام داران و تعیین

استراتژی ها و برنامه ها به کار می رود.

2. **ارزیابی میان دوره:** این ارزیابی در طی دوران یک محصول یا خدمت اطلاعاتی (یعنی در

مرحله اجرا) انجام می شود. نظارت و ارزیابی در میان دوره نقش عمده ای در بررسی روند و

عملکرد محصول یا خدمت اطلاعاتی و همچنین شناسایی تغییرات محیطی مؤثر بر اثر آن

دارد. ارزیابی میان دوره مستلزم گردآوری و تحلیل داده ها برای شاخص های عملکردی است

تا بتوان عملکرد را با نتایج مورد انتظار مقایسه کرد. گاهی تمرکز این ارزیابی بر تشریح

حوادث غیر عادی است. بسیاری از فعالیت های این مرحله باید به طریق زیر انجام شوند:

- تلفیق و مقایسه داده های دروندادی و بروندادی؛

- اطمینان از زیاد نبودن هزینه ها نسبت به بودجه، و استفاده آن در روشی بهینه ؛

- اطمینان از انجام امور طبق برنامه و نیل به نتایج مطلوب

3. **ارزیابی عملی یا ارزیابی به اعتبار گذشته:** که درانتهای مرحله انجام پروژه صورت می گیرد. برای انجام آن باید کل چرخه شامل ساختار، پیشینه، اهداف، نتایج، فعالیت ها و ابزار آن بررسی شوند. این ارزیابی بر بررسی تأثیر کوتاه مدت، همچنین ارزیابی میزان پایداری تأثیر و عوامل موفقیت یا شکست پروژه تأکید دارد.

4. **ارزیابی میزان اثرگذاری:** در هنگام اجرا یا بعد از انجام یک محصول یا خدمت اطلاعاتی صورت می گیرد. ارزیابی میزان اثرگذاری در صدد بررسی اثرات گسترده و بلند مدت است. در اینجا مسأله این است که آیا اطلاعات تولید شده واقعاً بر فعالیت های مخاطب تأثیر می گذارد و سبب تغییر در جامعه می شود. هر چند ثبت کردن میزان تأثیر بلندمدت در سطح اجتماعی اقتصادی به علت مشکلات نگرشی بسیار سخت است؛ اما تحت برخی شرایط و با برخی ویژگی ها فرض می شود که این تغییرات مشاهده شده و پروژه ارتباط وجود دارد. در این شرایط این ارتباط پذیرفتنی است و محصول یا خدمت دیگری هم وجود ندارد که بتواند چنین تأثیری را ایجاد کرده باشد (CTA/KIT/IICD Smart Toolkit 2005).

۲-۳-۱- خودارزیابی

در خود ارزیابی هدف، یک بررسی انتقادی از عملکرد و ساختار سازمانی در ارتباط با اهداف طرح ریزی شده در مقابل برنامه های عملیاتی و استراتژیک است. این فرآیند منجر به ارائه گزارشی خواهد شد که توسط خود متصدیان ساختارهای سازمانی تهیه شده است. عناصری که در کل مسئولیت مدیریتی را تشکیل می دهند و باید در گزارش خودارزیابی لحاظ شوند، عبارت اند از:

1. برنامه ریزی استراتژیک؛

2. برنامه عملیاتی؛

3. مدیریت منابع.

باید به خاطر داشت که مدیریت نقش اصلی را در سطوح سازمانی یا استراتژیک مختلف ساختارها

ایفا می کند، مانند:

- تعریف مأموریت؛
- برنامه ریزی اهداف مطابق با مأموریت؛
- تخصیص منابع کافی برای نیل به اهداف از پیش تعیین شده؛
- ارزیابی دوره ای از نتایج به دست آمده و میزان اثربخشی و کارایی فرآیندهایی طراحی شده برای نیل به اهداف؛
- تعریف و اتخاذ سنجه های صحیح طبق بازنگری های مستمر از فرآیندها (CIVR 2006).

۲-۴- الگوبرداری^{۱۲}

ترازیابی یا الگوبرداری؛ مقایسه بین فعالیت ها و فرآیندهای یک ساختار معین با بهترین نمونه موجود

در آن حوزه برای تعیین رتبه بندی و استراتژی های ممکن برای توسعه ساختار است.

این فرآیند مستلزم ساختارهای پژوهشی گوناگون برای طراحی یک فهرست رتبه بندی است. نتایج

این مقایسه کل یک کشور را فرا می گیرد و به ساختارهای دیگری که به نتایج خوبی در همان حوزه

علمی رسیده اند، هم مربوط می شود. اهمیت شناخت ترازیابی همچنین ناشی از دیدارهای میدانی است که

در طی آن تجارب موفقیت آمیز می تواند به راحتی گردآوری و منتشر شوند (CIVR 2006).

۲-۵- اهداف و ضرورتها

به نظر می رسد، تقاضای روز افزونی برای ارزیابی فعالیتهای پژوهشی در سطح عمومی وجود داشته

باشد. رقابت بر سر تصاحب منابع کمیاب پژوهشی، جستجو برای معیارهای پذیرفته شده را برمی انگیزاند و

¹²Benchmarking

آژانس‌های سرمایه‌گذار و مقام‌های مجاز می‌توانند گزینه‌های منطقی و هدفمندی را برای تخصیص این منابع داشته باشند.

جدا از چنین عوامل خارجی، عامل درونی که سبب استقبال پژوهشگران از ارزیابی می‌شود این است که بسته به قابلیت پایداری روش مورد استفاده، ارزیابی به رشد کیفیت پژوهش در آینده کمک می‌کند. برای مثال، در علوم طبیعی و پزشکی، شاخص‌های کتابسنجی و بخصوص سنجش میزان تأثیر مقالات نوشته شده توسط پژوهشگران، به طور مستقیم از طریق شمارش استنادها یا به طور مستقیم از طریق ضریب تأثیر رسانه‌ای، که در آن این مقالات منتشر می‌شوند، اقدام‌های استاندارد می‌هستند که در همه انواع ارزیابی‌های پژوهشی اعم از سطوح فردی یا سازمانی به کار گرفته می‌شوند (KNAW 2005).

۲-۶- نقش ارزیابی در دانشگاه‌ها و پژوهشگاه‌ها

استانداردهای جدید برای ارزیابی‌ها که اغلب در مسیر مدیریت عملکرد قرار دارند، طی ده سال گذشته در حوزه ارزیابی پژوهش در دانشگاه‌ها مطرح شده‌اند. سیستم همترازخوانی در فعالیت‌های متناوب پژوهشگران، به مثابه سیستمی برای کنترل کیفیت پژوهش تلقی می‌شود که به عنوان روشی برای کنترل کیفیت خودگردان و ایجادبخش مهمی از شناسایی علمی به کار گرفته می‌شود. سیستم همترازخوانی، نوع خاصی از نظارت سنتی را اجرا می‌کند که بر اساس شناخت کیفیت و ارتباط با قواعد سنتی علم است. در این مدل کلاسیک، دانشگاه از جامعه جدا انگاشته می‌شود و نتایج فرآیند همترازخوانی پژوهشگر تنها به خود او مربوط می‌شود. سیستم همترازخوانی به تدریج اما مستمر، تغییر شکل داده، یکپارچه شده و با سایر انواع ارزیابی‌های کمی که هدف آن سنجش عملکرد و محاسبه آن بود، ترکیب و اجرا شده است. این نوع ارزیابی‌ها، بر اساس ثبت میزان بهره‌وری با استفاده از شاخص‌های کمی مثل نمایه استنادی علوم و محاسبه ضریب تأثیر قرار دارند.

معرفی انواع جدید ارزیابی، این سؤال را مطرح می‌کند که چگونه این سیستم‌ها بر رفتار پژوهشگر تأثیر می‌گذارد و از نقطه نظر پژوهشگر چگونه کل یک سازمان پژوهشی را تغییر می‌دهد. طبق نظر برانسون و یاکوبسن، اجرای مجموعه‌ای از رویه‌های استاندارد در سازمان‌های حرفه‌ای، در ارتباط با قوانین و ارزش‌های حرفه‌ای موجود در سازمان‌ها، اغلب منجر به نتایج ناخواسته و مخرب می‌شود.

روش‌های سنتی کنترل کیفیت در جامعه علمی (سیستم همترازخوانی) هنوز هم حائز اهمیت هستند، اما کافی نیستند. این روش‌ها تغییرات ساختار سازمانی دانش‌کاری یا حوزه‌هایی چون مسئولیت‌پذیری و قابلیت محاسبه را در بر نمی‌گیرند. (Hansson 2005)

۲-۷- نقش ارزیابی در سیاستگذاری

در اصل، همه سازمان‌های متصدی وظایف ارزیابی ادعا می‌کنند که هدف از شناسایی عملکرد پژوهش در کج‌چگونگی فرآیند نوآوری در عمل است. بیانیه‌های دولتی و اسناد اداری تأکید می‌کنند که هدف از ارزیابی، سیاستگذاری پژوهشی و تخصیص بودجه در آینده است. اما سیاست‌ها اغلب فاش نشده باقی می‌مانند. کمبود حمایت فرهنگی در بسیاری از موارد مشکل واقعی است که مانع موفقیت هر چارچوب ساخت یافته برای ارزیابی پژوهش می‌شود

این نوشتار، در پی یافتن این مطلب است که آیا ارزیابی پژوهش به عنوان ابزار سیاست‌گذاری استفاده می‌شود. نتیجه مهم به دست آمده این است که نتایج ارزیابی در سطح سیاست‌گذاری به قدر کفایت به کار برده نمی‌شود. در سال 2002 آژانس ملی فناوری فنلاند تصدیق کرد که دولت‌ها به طور مستمر نیاز به تنظیم سیاست‌های خود در نتیجه شرایط متغیر و تجربه یادگیری دارند، اما اقدامات مهم ارزیابی هنوز فراگیر نشده است. (Papponetti and Bucchi 2007)

از نقطه نظر بین کشوری، اجرای مدل‌های ارزیابی با اصلاح سیستم علمی ملی مکمل هم هستند. کاربرد ناموفق ارزیابی‌ها اغلب به این علت است که تلاش‌ها هنوز به از سرگیری ساختارهای پژوهشی کشوری هدایت می‌شوند. برای نمونه، در ایتالیا اولین کمیته مسئول برای ارزیابی در سال 1998 تشکیل شد. در اسپانیا چهارمین طرح ملی برای سالهای 2003-2000 اجرا شد که استراتژی شفافی برای ارزیابی پژوهش ارائه می‌کند. با نگاهی نزدیک‌تر، شواهد نشان می‌دهند که هر مطالعه موردی با مشکلات خاصی روبرو بوده است. (Papponetti and Bucchi 2007)

مطالعه‌های موردی نشان می‌دهد که دانش بیشتری در مورد فرآیندهای گردآوری، تازه‌سازی، و اشاعه دانش علمی و فنی و مهم‌تر از آن، چگونگی کاربرد این دانش برای بهبود بهره‌وری و توسعه عمومی، اجتماعی و اقتصادی لازم است.

۲-۸- ترکیب ارزشیابی آموزشی و پژوهشی

یکی از اهداف ارزیابی پژوهش‌های دانشگاهی، بهبود و همچنین افزایش اثربخشی آنها، سنجش ابعادی از آنها مثل کیفیت، کارایی، ربط، و پایداری است. بر اساس چنین ابعادی، امکان طراحی و سنجش دوباره آنها پدید می‌آید. ارزیابی‌های اصولی می‌توانند از طریق کاربرد شاخص یا همترازخوانی صورت گیرند. به هر حال، افزایش جایگاه پژوهش‌های دانشگاه در بطن جامعه باید انعکاس یابد (Campbell 2005). به دلیل ارتباط پژوهش‌های دانشگاهی و آموزش دانشگاهی، ارزیابی پژوهش باید به طریقی به ارزیابی آموزشی ارتباط داده شود. در آخر این سؤال مطرح می‌شود که چه عملکردهایی در ارزیابی باید بررسی شود. بخش اعظمی از عملکردها به مسائلی چون بازخورد، نتایج کار برای سرمایه‌گذاری، و انگیزه‌ای برای یادگیری و سامان‌دهی برمی‌گردد. بحث روی این موارد پیچیده، چارچوب مفهومی برای ارزیابی پژوهش‌های دانشگاهی پدید می‌آورد. (Campbell 2005)

با توجه به مدل‌های موجود چند بعدی از کیفیت و همچنین ابعاد دیگری که می‌تواند در سطوح عالی و پیشرفته و برتر اضافه شوند، مثل اثربخشی که تمایل زیادی وجود دارد که در باره اثر پژوهش‌های دانشگاهی قضاوت شود. با چنین درکی از اثربخشی، پس این بعد نیز به ارزیابی‌های قبلی قابل افزودن است:

1. اثربخشی: تحلیل اثربخشی پژوهش‌های دانشگاهی می‌تواند بر اساس درک جامع و ارزیابی ابعاد سطح اول کیفیت، بهره‌وری، ربط و پایایی آنها باشد.
2. بهبود سازمانی: آیا ارزیابی پژوهش دانشگاهی منجر به توسعه سازمانی پایدار ساختارهای دانشگاهی می‌شود. این امر یعنی اینکه عملکرد مثبت از ارزیابی پژوهش به سایر حوزه‌ها نیز تسری می‌یابد.
3. افزایش کیفیت پویا: آیا انجام ارزیابی جامع از پژوهش‌های دانشگاهی سبب افزایش نظام‌مند کیفیت این پژوهش‌ها خواهد شد؟ و چنین کیفیتی در طول زمان قابل سنجش خواهد بود؟
4. تحول مشترک پژوهش دانشگاهی و ارزیابی پژوهش دانشگاهی: تا چه حد بهبود پژوهش دانشگاهی و ارزیابی پژوهش دانشگاهی با هم تقابل دارند؟ بهبود کیفی در پژوهش دانشگاهی می‌تواند سبب افزایش کیفیت در ارزیابی‌های پژوهشی شود. به همین ترتیب، بهبود ارزیابی پژوهش می‌تواند منجر به عملکرد بهتر پژوهش دانشگاهی شود (Campbell 2005).

۲-۹- ارزیابی در کشورهای جهان سوم

بسیاری از کشورهای در حال توسعه، حس می‌کنند که نمی‌توانند از پژوهش‌های محض بهره گیرند، و اعتنای زیادی برای کاربردهایی که پژوهش در آینده خواهد داشت، ندارند. این دیدگاه که صرف هزینه برای تحقیق و توسعه باید برگشت اقتصادی و اجتماعی داشته باشد، از خصیصه‌های همه عواملی

است که در صدد استفاده از دانش سازمانی برای تحقق اهداف هستند. در رابطه با اهداف سودگرایانه علم، معمولاً ادعاهای زیادی در بین دانشمندان و مدیران سیاست گذار عرصه تحقیق و توسعه مطرح است. اما در بسیاری از کشورهای در حال توسعه، این نوع ادعاها در بین طبقه‌های سیاسی و تجاری کم‌رنگ‌تر است و انجام پژوهش‌های پرهزینه به سیاست‌گذاری غیر رسمی توأم با برخی چشم‌پوشی‌ها منتج می‌شود. اغلب، روش ساده‌تر و ارزان‌تر برای عاملان صنعتی و تجاری محلی، به دست آوردن علمی است که در فناوری‌های وارداتی جای گرفته و هزینه، انرژی و خطر سر و کار داشتن با نیروهای علمی و فنی محلی را هم نداشته باشد. جهانی‌سازی و آزادسازی تجاری نیز سرعت انتقال فناوری مستقیم و غیرمستقیم را بسیار افزایش داده است، و استدلال‌هایی نیز برای الویت در سرمایه‌گذاری ملی در علم و فناوری نیز مطرح می‌شود که برای سرامدان سیاست و تجارت کمتر آشکار است. اما، فعالیت‌های اقتصادی با ارزش افزوده بسیار، بدون سرمایه‌گذاری محلی در پژوهش و آموزش، نمی‌تواند جلب نظر نماید یا اختصاص داده شود، و پیشرفتی هم نمی‌کند. اما حامیان صرف هزینه در علم و فناوری باید ترتیب و ترکیب مناسبی برای صرف هزینه عمومی در علم، پژوهش، آموزش، و توسعه فناوری برای برگشت اقتصادی و اجتماعی پس از آن را فراهم کنند.

سه عامل در کشورهای در حال توسعه، مانع پژوهش اثربخش می‌شود: اولین عامل محدودیت بودجه تحقیق و توسعه این کشورها است. روندهای تخصیص منابع تحقیق و توسعه تحت نفوذ افراد و اعتبارات و وابستگی‌های سیاسی برای در نظر گرفتن مسائل کارایی، اثربخشی، ربط، کاربرد، یا برتری آنها است. عامل دوم، وابستگی کشورهای در حال توسعه به کمک‌های علمی بین‌المللی، که دستورات آنها نیز دائماً در حال تغییر است، می‌باشد. عامل سوم و مهمترین عامل این است که تحقیق و توسعه برای مؤثر بودن باید جزئی از یک سیستم نوآوری یا جامعه نوآوری (ولو اندک) باشد که در آن پیوند و ارتباط فعالی

بین پژوهشگران و دیگر عوامل چون واسطه‌ها، سازمان‌های هماهنگ کننده، آموزش دهندگان، مقررات توسعه، و بخصوص کاربران برقرار است. (Davis and Carden 1998)

سازمان‌های علمی بسیاری در جنوب مجبور به تنظیم سریع خود با تغییرات سریع اقتصادی، اجتماعی، و شرایط علمی هستند. محدودیت‌های مالی در بخش عمومی، به طور نامتناسبی بر سیستم‌های پژوهشی کشورهای در حال توسعه تأثیر می‌گذارند که بیشتر آنها به میزان زیادی به حمایت عمومی وابسته هستند. در شرایط نوسازی و طراحی ساختار، پژوهش باید برای کسب بودجه عمومی، با تقاضاهای ساخت جاده، بیمارستان، و امنیت نظامی رقابت کند. پژوهش در این رقابت، تنها با ارائه سودمندی یا سازماندهی خود برای جلب حمایت سیاسی، می‌تواند پیروز باشد. زیرا جلب حمایت سیاسی در بیشتر مواقع، آسانتر از اثبات سودمندی پژوهش است این امر از آنجا ناشی می‌شود که رویه‌های اختصاص منابع در کشورهای در حال توسعه می‌تواند بسیار مبهم و شخصی باشند.

در حالیکه یک توافق عمومی در جنوب درباره پژوهش وجود دارد که پژوهش امری مهم است و فضای سازمانی کافی نیز همیشه برای اجرای آن فراهم نمی‌شود؛ ضعف سازمانی در این مورد کمتر دیده می‌شود زیرا وظایف جنبی مثل آموزش بر درصدی از کارکنان علاقه‌ای به آن ندارند، غلبه می‌کنند و پژوهشگران نیز اغلب نیاز به افزایش درآمد خود از طریق انجام امور غیرپژوهشی دارند. در بدترین موارد، سالها سرمایه‌گذاری در توسعه سازمانهای علمی و فنی و منابع انسانی می‌تواند به علت ناسازگاری اقتصادی و اجتماعی از بین رود. بررسی اخیر که از جوامع علمی 12 کشور در حال توسعه صورت گرفته است، حاکی از کاهش در توانایی و رسمیت جامعه علمی در چندین کشور آفریقایی، عربی، و آمریکای لاتین می‌باشد و تلاش‌های بسیار کشورهای دیگر برای بازسازی و سازگاری سیستم‌های علمی خود به روش‌هایی که پاسخ‌گوی امور توسعه اجتماعی و اقتصادی آنها باشد، ناکام مانده است.

ارزیابی عملکرد دانشگاه‌ها اغلب بار سیاسی دارد. در کشورهای در حال توسعه، دانشگاه‌ها معمولاً حساسترین عاملان علم و فناوری به سیاست هستند. در بسیاری از کشورهای افریقایی، سیستم‌های آموزش عالی با مسائل آماری روبرو است که در آن تعداد قابل افزایشی از فارغ‌التحصیلان تحصیلات متوسطه درخواست ادامه تحصیل در دانشگاه‌ها را دارند، اما دولت‌ها برای سرمایه‌گذاری نامتناسب در تولید نخبگان جوان با تحصیلات دانشگاهی که احتمالاً در صدد ورود به بازار شغلی کشور نیستند، بی‌میل هستند. در ضمن، بخش اقتصادی نیز قادر نیست تا نیازهای آموزشی آینده را بررسی کند یا بسیاری از محصولات علمی یا انسانی دانشگاه‌ها را جذب نماید. در کنیا، پژوهشگران در دانشگاه‌ها حمایت صنعتی را برای پژوهش‌های دانشگاهی به لحاظ کمک‌های مالی ارزیابی کردند، در حالیکه عاملان صنعتی آن را به لحاظ میزان دسترسی موفق به نوآوری‌ها ارزیابی نمودند. این امر شکل بسیار متفاوتی از انتظارات کشورهای شمال است، که در آنجا تولید منابع انسانی انگیزه اولیه برای همکاری دانشگاه و صنعت است.

در چندین کشور آمریکای لاتین، نمایندگان بلندمرتبه دولتی طی سال‌های اخیر اعلام کرده‌اند که هزینه‌های دانشگاه‌ها بیشتر از ارزش آنها است. طراحی مرحله‌ای برای ارزیابی‌هایی که پژوهشگران دانشگاهی از آن‌ها بهره می‌توانند موجب کاهش عملکردهای حیاتی دانشگاه‌ها برای افزایش بهره‌وری و کارایی گردد. در محیط دانشگاه‌های کشورهای جنوب، مشکلات عملیاتی برای انجام یک ارزیابی هدف‌مند و منصفانه اندک اما با اهمیت هستند؛ چگونه ارزیاب‌های شایسته در جامعه علمی ملی یافته می‌شود؟ چگونه شاخص‌های مناسب برای ارزیابی کیفیت و بهره‌وری تعیین می‌گردد و امثال آن. بدون این قوانین ارزیابی نمی‌تواند برای اجرا سرمایه‌ای جذب کند. آنچه که مورد تردید است این است که رویه‌های ارزیابی نمی‌توانند به اندازه کافی بی‌طرف باشند تا استفاده از آنها را به عنوان ابزاری برای امور حقوقی کاربران یا مسائل اداری سیاسی مانع شوند.

ارزیابی پژوهش برای رد یا قبول یک پژوهش یا اهداف سنجش توسط افراد متخصص همرد که معمولاً کیفیت را از نقطه نظر رشته‌ها یا حوزه‌های علمی خاصی ارزیابی می‌کنند، انجام نمی‌شود، بلکه باید به عنوان روش شناسایی شرایطی برای تحقق اثربخشی در تحقیق و توسعه توجه شوند. این شرایط از طریق بافت سازمانی تعیین شوند. به دلیل وابستگی اثربخشی پژوهش به بافت سازمانی، شرایط برای آگاهی از میزان اثربخشی پژوهشی به خصوص در کشورهای در حال توسعه بسیار سخت تر است؛ زیرا در آنجا تقاضا برای علم و فناوری و مخصوصاً برای علم و فناوری بومی کمتر از کشورهای شمالی است. ارزیابی به عنوان یک فرآیند یادگیری وابسته به توان سازمانی است و این امر در کشورهایی که سازمان‌های علمی ضعیفی دارند، یک مشکل محسوب می‌شود.

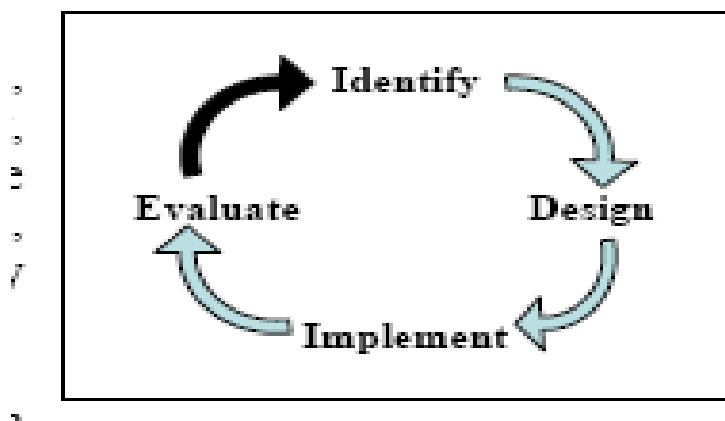
سازمان‌ها باید ظرفیت مدیریت مداخله‌های آژانس‌های سرمایه‌گذاری پژوهشی ملی و بین‌المللی و حفظ برنامه پژوهشی منسجم و همچنین توان گفتگو و همکاری با ذی‌نفعان داخلی و خارجی را داشته باشند. ارزیابی مطلوب و نظرسنجی کارشناسی می‌تواند برای برآورد موفقیت‌های کوتاه مدت و روند پیشرفت طرح‌های پژوهشی استفاده شوند. اشتراک گذاری اثربخشی برای فرآیندهای توسعه سازمانی و برای ارتباط بین توسعه و سازمانیهای پژوهشی بومی و فرآیندهای توسعه ملی مهم هستند. ارزیابی پژوهش می‌تواند به تعریف شرایطی برای پژوهش مؤثر و ایجاد آن شرایط کمک کند. (Davis and

(Carden 1998)

۲-۱۰- اصول زیربنایی ارزیابی

ارزیابی به این صورت می‌تواند تعریف شود:

1. فرآیندی نظام مند و بی طرفانه که ربط، عملکرد، کارایی و تأثیر (مورد انتظار یا غیر مورد انتظار) یک طرح یا سیاست را در ارتباط با اهداف تعیین شده بررسی می کند.
2. ارزیابی از نظارت و بازرسی متمایز است. نظارت یک روند مستمر است و بر فعالیت های روزمره متمرکز است، و کانون بازرسی بر اندازه گیری منابع است. ارزیابی اقدامی است قبل از راه اندازی یک طرح، برنامه یا سازمان، و نگرستن به آن به عنوان یک کل به لحاظ اهداف آن، میزان تحقق آن، و درس های آموخته شده آن انجام می شود.
3. ارزیابی با نگارش یک گزارش پایان نمی گیرد. برای اثربخشی آن، نتایج به دست آمده باید به سیاست گذاران، برنامه ریزان، و مجریان به صورت دقیق ارائه داده شوند. ارزیابی بخشی اساسی از شواهد، تجربه، و قضاوت های کارشناسانه است که برای سیاست گذاری و مدیریت طرح ها استفاده می شوند.



شکل ۲-۱ فرآیند ارزیابی

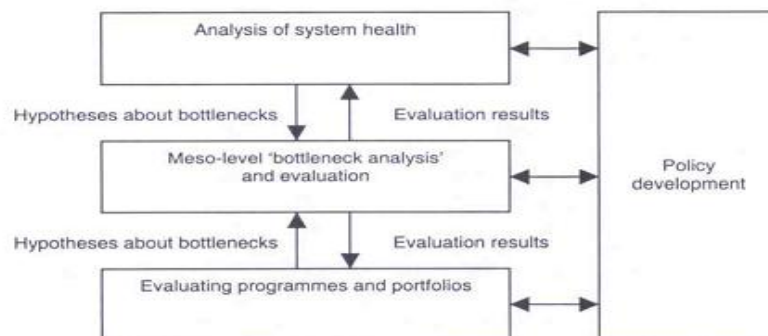
۲-۱۱- ارزیابی به عنوان یک سیستم

در همه سطوح، وظیفه ارزیابی پاسخ به این مسائل است:

1. آیا درست انجام می دهیم (شایستگی)؟
2. نتیجه تأثیر کار چه خواهد بود (تأثیر)؟

3. می‌توان آن را بهتر انجام داد (اثربخشی)؟

همان‌طور که در سلسله مراتب سیستم‌ها مشاهده می‌شود، تعادل بین تغییرات سؤال‌های سنتی ارزیابی مشاهده می‌شود. ارزیابی‌های خوب، همیشه به سوی سطوح بالاتر سیاست‌گذاری متمایل هستند. یعنی، این نوع ارزیابی‌ها مسائلی درباره شایستگی مطرح می‌کند. اما در سطوح پایین‌تر سیستم ارزیابی، با فعالیت‌های مجری سیاست‌ها سر و کار دارند و بنابراین بیشتر تلاش آنها به جای شایستگی، بر تأثیر و اثربخشی قرار دارد. برای نمونه، ارزیابی‌ای که از یک برنامه تحقیق و توسعه مربوط به مهندسی مکانیک صورت می‌گیرد، به احتمال زیاد، در صدد یافتن شکاف‌های موجود در جامعه کاربران خدمت‌گیرنده است تا به چالش کشیدن اصولی که چنین برنامه‌ای بر اساس آنها مورد نیاز است. ارزیابی در سطوح سیستمی بالاتر، چالش بیشتری درباره سیاست‌گذاری ایجاد می‌کند، زیرا ارزیاب‌ها مجبور هستند برخی از مشکلات سیستماتیک مشترک را مانند سیاست‌گذاران تحلیل کنند. برای مثال، ارزیابی‌ای از یک شورای پژوهشی می‌تواند به سرعت منجر به مورد سؤال قرار دادن سیاست ملی پژوهش شود. (Arnold 2004)



شکل ۲-۲/ارزیابی سیاست نوآوری و پژوهش در سیستم‌های جهانی

۲-۱۱-۱- چهار سیستم ارزیابی

- الگوریتم‌ها (بر اساس سنجه‌های کمی): سیستمی است که صرفاً بر اساس الگوریتمی برای ترکیب سنجه‌ها می‌باشد. چنین سیستمی خودکار است و تنها به ارزیابی عینی می‌پردازد. سنجه‌های مختلفی می‌تواند در این سیستم به کار گرفته شود. سنجه‌های اعتبار بر پایه

بررسی‌ها؛ میزان ورود پژوهش‌های خارجی؛ سنجه‌های کتابسنجی (استنادها یا انتشارها)؛ تعداد دانشجویان پژوهشی؛ و سنجه‌های مالی.

- بررسی کارشناسانه (از جمله همتراز خوانی): در این سیستم، کارشناسان عملکرد اشخاص یا گروه‌ها و عملکرد احتمالی آنها را در آینده، طی چرخه‌ای مشخص مورد قضاوت قرار می‌دهند. ارزیابی می‌تواند به طور کامل توسط داوران یا کارشناسان دیگری چون نمایندگان گروه‌های کاربری، افراد و کارشناسان مالی انجام شود.
- رتبه‌بندی سوابق: سیستمی است که در آن گروه‌ها، واحدها، یا دانشگاه‌ها کاملاً بر اساس عملکرد گذشته آنها رتبه‌بندی می‌شوند. در عمل پژوهشی در این سازمان‌ها بهتر محسوب می‌شود، که سابقه بهتری داشته باشد. برای تعیین عملکرد گذشته علاوه بر رتبه‌بندی آر.ای. ای. سنجه‌های متفاوتی چون میزان جذب سرمایه‌های مالی می‌تواند به کار گرفته شوند.
- خودارزیابی: در این سیستم، مؤسسه‌ها و سازمانها یا اشخاص خود را ارزیابی می‌کنند. بخشی از این ارزیابی با دقت بازنگری می‌شود. البته این ارزیابی داخلی است و با ارزیابان خارجی می‌تواند دچار چالش شود. (Wooding and Grant 2003)

۲-۱۲- واحدهای مورد ارزیابی

۲-۱۲-۱- افراد

مدل دیگری برای ارزیابی طرح‌های علمی و فنی، مدل «سرمایه انسانی علمی و فنی» است، که در این مدل به محصولات و بروندادها توجه کمتر می‌شود، بلکه تمرکز اصلی آن بر خط سیر حرفه دانشمندان و توانایی آنها در همکاری و ارتقاء قابلیت‌های خود است.

به همین دلیل، مدل پیچیده‌ای است که نیاز به دانشی از استعدادهای نیروهای انسانی در طرحها و بنا به عقیده دیگر، مشخص کردن این منابع است. این قابلیت‌ها هم می‌توانند در حوزه‌های خود آنها مورد بررسی قرار گیرند و هم می‌توانند به عنوان مجموعه‌ای از تأثیرات علمی و فنی قلمداد شوند که نه تنها ارزش علمی آن را نسبت به ارزش مالی بروندها بیان کند، بلکه ارزش آن را در سیاست‌گذاری علمی نسبت به سایر ارزش‌های نوین یا سنتی تعیین کند. پس، وظیفه سختی در این مدل دنبال می‌شود که به مراتب از شمارش محصولات یا حتی تعیین میزان ارزش مالی آنها مشکلتر است. با این وجود، این مدل اجرا می‌شود.

زیرا دست کم چهار مزیت می‌توان برای آن بیان نمود:

1. با «چرخه دوام» و تکامل حوزه‌های علمی و فنی ارتباط دارد،
2. تطابق نزدیک‌تری با درک دانشمندان از آثار خود دارد و از دانش ایجاد شده در جامعه شناسی علمی بهره‌برداری می‌کند،
3. می‌تواند پاسخ‌گوی نیازهای سیاست‌گذاران برای قضاوت و داوری باشد،
4. اساس این مدل، بر اموری غیر از ارزش مالی علم قرار دارد. (Bozeman, Dietz et al. 2001)

۲-۱۲-۲- سازمان‌ها

لیدسدورف به طور خلاصه از جانب کالینز بیان می‌کند که در زمینه مجموعه معین از مجلات، سازمان‌ها واحدهای مناسبی برای مقایسات علم‌سنجی نمی‌باشند زیرا این سازمانها دارای ماهیت همگن و متجانس نمی‌باشد. در یک حوزه تخصصی، شاید تمرکز بر مجلات و عدم کاوش روابط میان‌رشته‌ای استراتژی خوبی باشد. همچنین سنجش افراد، گروه‌ها یا طرح‌های پژوهشی با یکدیگر متفاوت هستند. کشورها می‌تواند به عنوان واحدهای سازمانی یک پارچه‌ای برای تحلیل محسوب شوند. بنابراین آنچه که برای تحلیل سازمان‌ها به کار می‌رود، در مورد آنها نیز صادق است. مثلاً میزان دارایی کشورهای

آسیای شری در علوم زیستی نسبت به ایالات متحده و بیشتر کشورهای اروپای غربی کمتر است. پایگاه آی.اس.آی. بیشتر شامل علوم مربوط به زندگی است تا علوم فیزیکی. بنابراین مقایسه کشورهای آسیای شرقی با این پایگاه نتایج ضعیف تری نسبت به مقایسه کشورهای غربی با آی.اس.آی به دنبال خواهد داشت.

در حالی که کشورها را می توان واحدهای بزرگ تحلیل سازمانی محسوب کرد، افراد می توانند واحدهای کوچک آن باشند. افراد نیز به لحاظ بروندهای علمی خود، ممکن است نامتجانس باشند. در برخی موارد، پژوهشگران خلاق، آثار متفاوت تری ارائه می کنند و ادعا می شود که دانش آنها مرزها را پشت سر می گذارد. بار-یان دریافت که این نامتجانسی در گوگل اسکالر بیشتر از آی.اس.آی. در مورد علوم اطلاع رسانی مشهود است اما نمی توان این آمار را قطعی تلقی نمود. زیرا قطعا در اغلب موارد، افراد از این میانگین فاصله دارند. (Leydesdorff 2008)

۲-۱۲-۳- بروندهای پژوهشی

در نمونه های بارز ارزیابی پژوهش در ایالات متحده، معمولاً ارزیابی می تواند به صورت پرسشهای

زیر دسته بندی شوند: (Roessner 2002)

- آیا این برنامه باید ادامه یابد یا متوقف شود؟ (ارزیابی تلخیصی)
- عواید اقتصادی این طرح چیست؟ (برگشت سرمایه، هزینه سودمندی)
- چگونه نتایج و تأثیرات ناشی از این طرح تعیین می شوند؟ (ارزیابی فرآیند یا تکوینی)
- آیا نتایج حاصل شده از طرح، با استانداردهای پذیرفته شده کیفیت علمی همخوانی دارد؟ (بازنگری صلاحیت، همتراز خوانی)
- بروندهای قابل سنجش این برنامه چیست؟ (نظارت، سنجش عملکرد)

در این میان، همترازخوانی روش غالب شناخته شده است و روشهای دیگر اعتبار خود را با آن می

سنجند.

روشهای متداول برای سنجش عملکرد پژوهش نیز به این شرح است:

- مطالعه شکل گیری، رشد، و بهره‌وری شبکه‌های علمی؛
- استفاده از توزیع چولگی ارزش برون‌دادهای پژوهشی برای تعیین ارزش طرح نسبت به هزینه های آن؛
- تأکید بر روش‌های ارزیابی تکوینی، بخصوص مطالعه‌های موردی.

۲-۱۳- انواع ارزیابی

۲-۱۳-۱- ارزیابی کمی

قضاوت کاری مشکل است و هر کسی که قضاوت می‌کند می‌خواهد مورد پشتیبانی سنجه‌هایی قرار گیرد که به طور عینی عملکرد را توصیف و به مقایسه و تعیین رتبه آنچه که مورد قضاوت قرار گرفته است، کمک می‌نماید.

نگرانی‌های بسیاری برای استفاده از سنجه‌های مختلف و مخصوصاً تحلیل استنادی وجود دارد. بسیاری از این سنجه‌ها در روش‌هایی به کار گرفته می‌شوند که برای آنها طراحی نشده بودند و قضاوت‌های مبتنی بر آنها نیز با داده‌های آنها تطابق ندارند. بسیاری، کاربرد آنها را با این استدلال که این سنجه‌ها بر پایه فرمول‌های ریاضی دقیقی به وجود آمده‌اند، مورد تأیید قرار می‌دهند. با این وجود، دقت فرمول‌ها و درستی داده‌ها ممکن است صحت فرآیند تصمیم‌گیری را تحت شعاع قرار دهد.

شورای بین‌المللی ریاضیات کاربردی و صنعتی، موسسه آمارهای ریاضی، و اتحادیه بین‌المللی ریاضی، سازمان‌هایی هستند که ارتباطات جهانی ریاضی‌دانان و آمارگران را ارائه می‌کنند و مشکلاتی

درباره استفاده نادرست مفاهیم ریاضی یا شاخص‌های آماری دارند. از آنچه که جوامع به ریاضی‌دانان و آمارگران نشان می‌دهند، آنها احساس می‌کنند برای تهیه درک شفاف‌تر درباره کاربرد صحیح داده‌های آماری در ارزیابی پژوهش و بخصوص پژوهشهای حوزه ریاضی مسئول هستند.

این سازمانها، کمیته مشترک «ارزیابی کمی پژوهش» را ایجاد کردند و وظایف زیر را برای آن در نظر گرفتند:

1. ارزیابی گسترده‌ای که ضریب تأثیر آی.اس. آی شاخص مهمی برای کیفیت یک پژوهشگر،

واحد سازمانی، یا سازمانهای مشابه در ریاضیات یا آمار دارند.

2. تعیین حد و گستره‌ای که ضریب تأثیر آی.اس. آی. می‌تواند برای مقایسه کیفیت پژوهش در

ریاضیات با سایر رشته‌های دیگر به کار گرفته شود.

3. تعیین اینکه آیا ضریب تأثیر آی.اس. آی. نسبت به زبان، منطقه، زمان، منبع یا حوزه مدرک

منتشره، کار بین رشته‌ای جهت‌گیری دارد یا نه.

4. بررسی این پرسش‌ها برای انتخاب سنجه‌های دیگر بر اساس داده‌های استنادی.

5. پیشنهاد جایگزین‌های مناسب برای این سنجه‌ها بر اساس داده‌های استنادی.

6. فهرست کردن مزیت‌ها و مضرات استفاده گسترده از ضریب تأثیرها و سنجه‌های مشابه که

می‌توانند بر رفتار، بکارگیری، تعادل میان رشته‌های علمی و .. داشته باشند.

از کمیته خواسته شد خلاصه‌ای از یافته‌های خود را برای ثبت و توزیع عمومی از طریق کمیته‌های

اجرایی سازمان‌های عضو ارائه کند و نتایج خود را همراه با نمونه‌ها و داده‌های آماری در اسناد تکمیلی

ارائه نماید تا شواهد و پایه‌های لازم برای یافته‌های خود را نشان دهد.

روش علمی، تلفیقی پیچیده از همکاری‌های پژوهشی و آموزشی و همچنین خود سازمان علمی برای

جامعه علمی است. طبق بررسی‌ها ارزیابی کیفیت علمی مشکل است و پاسخ‌های ساده به سؤالات پیچیده

داوری غیرممکن می‌نماید. اما از این کمیته خواسته شده است که بررسی نماید؛ آیا امکان به کارگیری سنج‌های دیگری که صرفاً برپایه استنادها بنا شده‌اند وجود دارد یا خیر؟ این امر می‌تواند به ارزیابی پژوهش و روش‌های آکادمیک کمک کند و کیفیت را در روشی عینی نشان دهد.

۲-۱۳-۱-۱- چگونگی کاربرد شاخص‌های کمی

به دلیل اینکه به کارگیری شاخص‌های عملکردی در عمل آسان نیست، لازم است برخی مسائل آن مشخص شود، مثلاً این شاخص‌ها در ارزیابی پژوهش چه نقشی را ایفا می‌کنند؟ کدام شاخص‌ها باید انتخاب شوند؟ و کاربرد آنها چه نتایج ناخواسته‌ای را می‌تواند در پی داشته باشد؟ در همه شاخص‌های کمی این مشکل وجود دارد که در حوزه‌های گوناگون پژوهشی، امور متفاوتی وجود دارند و لازم است سطح تراکم موجود و قالبی که نتایج در آن ارائه خواهند شد، مشخص شود. اکثریت قابل توجهی از متون مربوطه، این مشکلات را تنها مربوط به شاخص‌های کتابسنجی می‌دانند، اما این مشکلات بر همه شاخص‌های کمی تأثیرگذار هستند.

۲-۱۳-۱-۲- نقش شاخص‌های کمی

شاخص‌های کمی چه نقشی را باید در فرآیند ارزشیابی ایفا کنند؟ منابع موجود به این پرسش دو پاسخ متفاوت داده‌اند: جایگزینی کامل آنها به جای همترازخوانی، یا استفاده از همترازخوانی به عنوان کمکی برای ارزیابی. پاسخ به این پرسش، همانند بسیاری از موضوعات طرح شده در این بررسی دارای اهمیت است و رابطه مستقیمی با سنج‌های مورد استفاده دارد. (Donovan 2005)

• جایگزینی شاخص‌های کمی به جای همترازخوانی

آیا تحلیلهای کمی، می توانند جایگزین قضاوتهای داوری شوند؟ اپنهایم^{۱۳} (1997) اظهار می دارد تحلیل استنادی به دلیل هزینه های زیاد، باید در «بریتیش ریسرچ اسسمنت اکسرسایز»^{۱۴} جایگزین کار داوری شود. اما، مطالعات انجام شده درباره همبستگی بین امتیازدهی (آر.ای.ای.) و سنجه های کتابسنجی نشان می دهد که با وجود همبستگی زیاد، موارد انحرافی نیز وجود دارند (Warner 2000). وارنر هشدار می دهد در جایی که تصمیم های بنیادی منوط به نتیجه ارزیابی پژوهش باشد، درباره جایگزینی کامل کتابسنجی به جای همترازخوانی نگرانیهایی وجود دارد. وجود نمونه های خوبی از تطابق بین شاخص های کتابسنجی و قضاوتهای داوری، بعضاً منجر به این گرایش می شود که شاخص های کتابسنجی به عنوان سنجه های عینی در تقابل با خصیصه ذهنی همترازخوانی تلقی شوند. اما باید به خاطر داشت که شاخص های کتابسنجی خود تا حدی بر اساس تصمیمهای داوری بنا شده اند، به عنوان مثال، پذیرفته شدن مقالات مجلات برای انتشار، مستلزم طی کردن مراحل ارزشیابی های داوری است. (Weingart 2003)

از دیاد رتبه بندی های اخیر مؤسسه ها، که اغلب به عنوان ابزار روابط عمومی استفاده می شوند، فکر جایگزینی هر چه بیشتر روشهای کمی به جای همترازخوانی را مطرح می کنند. این روشها درگیری داوری کمی دارند و اغلب، شاخص های کمی را در روش نه چندان تخصصی به کار می برند (van Raan 2005) و نوان افزایش عمومی ارزشیابی های سریع و بخصوص کم ارزش با کمک ضریب تأثیرهای استاندارد مجلات را برای مدیران سازمانها و مؤسسه ها اغواکننده می داند.

• بهبود همترازخوانی

به نظر بیشتر پژوهشگران، شاخص های کتابسنجی مورد استفاده جایگزین همترازخوانی نشده اند، بلکه نتایج ارزیابی پژوهش را به بحث می کشانند و به متخصصین اطلاعات افزوده ای ارائه می دهند (van Raan)

¹³ Oppenheim

¹⁴ British Research Assessment Exercise (RAE)

(and van Leeuwen 2002). همترازخوانی می‌تواند با استفاده از شاخص‌های کتابسنجی شفافتر شود و کمبودهای آن رفع گردد. به عنوان مثال، با استفاده از شاخص‌های کتابسنجی می‌توان از اعمال نظر سلیقه‌ای داوران در رد یا پذیرش مقالات جلوگیری نمود. همین‌طور صاحب‌نظران شاخص‌های کتابسنجی را در مواردی که بین همترازخوانان تردید صورت می‌گیرد؛ منبعی مفید می‌دانند. علاوه بر این، از آنها می‌توان برای مشخص کردن شکاف‌های دانش همترازخوانان استفاده کرد. در جایی هم که این شاخص‌ها با نتایج ارزشیابی‌های همترازخوانی همراستا نیستند، دلایل آن باید جستجو شود. ممکن است به خاطر مشکلات این شاخص‌ها باشند، یا اینکه متخصصان، دربارهٔ پژوهشی که آن را ارزیابی می‌کنند دانش کاملی ندارند تناقض‌های میان داده‌های کمی و همترازخوانی احتمالاً تحلیل‌های بیشتر و عمیق‌تری از عملکرد واحدهایی را می‌طلبد که مورد ارزشیابی مجریان ارزیابی قرار می‌گیرند.

از این بحث روشن می‌شود که تحلیل‌های کمی عملکرد پژوهش به جای اینکه جایگزین همترازخوانی شود، باید آن را بهبود بخشد و در نوشته‌ها نیز توافق گسترده‌ای روی این نظریه وجود دارد. بسیاری از شاخص‌ها تنها آمارهایی هستند که اطلاعات زیادی نیز در آنها وجود ندارد. بنابراین داده‌هایی به دست آمده از چنین تحلیل‌هایی باید همیشه با پیش‌زمینه‌ای از دانش درباره روش‌های رایج در انتشارات و خصیصه‌های حوزه‌های پژوهشی و ... تفسیر شوند (van Raan and van Leeuwen 2002; Gläser, Spurling et al. 2004).

افزایش کاربرد کتابسنجی غیرحرفه‌ای، به تازگی سبب بروز برخی مسائل شده است. ون ران (2005) افزایش عمومی ارزشیابی‌های سریع و بخصوص کم‌ارزش با کمک ضریب تأثیرهای استاندارد مجلات را برای مدیران سازمانها و مؤسسه‌ها اغواکننده می‌داند. وی این مسأله را از مدیران به دیگران تسری می‌بخشد. وینگارت منشأ افزایش استفاده آنها را کمبود صلاحیت و مهارت در تفسیر شاخص‌ها می‌داند

که منجر به هر چه تجاری تر شدن تولیدکننده انحصاری اطلاعات یعنی آی. اس. آی. شده است (Weingart 2003)

ورود اتفاقی شاخص‌های کتابسنجی در فرآیند همترازخوانی، به جای آنکه مکمل آن باشند، هنگامی که مسائل غیرمتعارف شناسایی و درک نشوند، می‌تواند امر ارزشیابی را مختل سازد. گلاسر خطر اصلی این کار را کتابسنج‌های غیرحرفه‌ای می‌داند که همه عوامل پشت سر استفاده چنین سنجه‌هایی را درک نمی‌کنند، و احتمالاً کورکورانه به آنها اعتماد می‌شود. آنها می‌ترسند که چنین کارهایی منجر به وضعیتی شود که در آن همترازخوانان محتوا را قضاوت نکنند و کتابسنج‌ها در آن به طور غیرعمدی چیره شوند. فشار زمانی بر همترازخوانان احتمال این رویداد را افزایش می‌دهد.

۲-۱۳-۲- ارزیابی کیفی

دونوان (2005)، در بررسی خود نتیجه گرفت که هیچ سنجه‌ای برای ارزیابی کیفیت یا تأثیر پژوهش در دانشگاه‌های استرالیا وجود ندارد. وی همچنین دریافت که بر خلاف پیشرفت‌های بوجود آمده در علوم تجربی، مفهوم تأثیر در علوم انسانی و اجتماعی عملی نیست و شاخص‌های مبتنی بر شمارش مدارک و مبتنی بر تحلیل استنادی، روشهای غالب در ارزیابی کیفیت پژوهش در علوم انسانی و اجتماعی هستند (Donovan 2005).

۲-۱۳-۲-۱- اساس چارچوب کیفی پژوهش

مکانیسم یا رویه‌ای برای ارزیابی کیفیت پژوهش یا بررسی ضریب تأثیر و مزیت‌های پژوهش‌هایی که از بودجه عمومی تأمین می‌شوند، در استرالیا وجود نداشت. بنابراین دولت استرالیا تصمیم گرفت با ایجاد «ساختار کیفی پژوهش» (آر. کیو. اف.)^{۱۵}، رویکردی منسجم و جامع‌تر برای ارزیابی کیفیت یا میزان تأثیر پژوهش‌هایی فراهم کند که با بودجه عمومی تأمین شده‌اند. هدف از این طرح، توسعه بنیانی

¹⁵Research Quality Framework (RQF)

برای ارزیابی بهینه کیفیت و میزان تأثیر در دانشگاه‌ها و آژانس‌های عمومی و فرآیند مؤثری برای نیل به آن بود. هنگامی که این ساختار توسعه یابد، می‌تواند مبنایی برای تخصیص منابع پژوهشی در آینده باشد. تعریف پژوهش و پیشرفت کاربردی، که شامل اثری خلاقانه است یکی از اجزای نوآوری می‌باشد که بر اساسی نظام‌مند برای افزایش میزان دانش آن استفاده می‌شود. «اداره آمار استرالیا»^{۱۶} پژوهش و پیشرفت کاربردی را در چهار نوع از فعالیت‌ها دسته بندی می‌کند: پژوهش بنیادین محض، پژوهش بنیادین راهبردی، پژوهش کاربردی شامل روش‌های نوین دستیابی به اهداف علمی و از پیش تعیین شده مانند اقدامات کلینیکی، و پژوهش‌های مربوط به توسعه کاربردی، «آر. کیو. اف.» تا آن جا که به پژوهش کاربردی و بنیادی مربوط می‌شود، تمایل دارد همه ابعاد پژوهش بنیادی و کاربردی و توسعه کاربردی را در برگیرد.

قابلیت‌های ساختار کیفی پژوهش به طور وسیعی مورد توجه سازمان‌ها و نهادهای پژوهشی استرالیا قرار گرفته است و بسیاری از آنها بر اتخاذ رویکردی مرحله‌ای برای توسعه آن توافق کردند. گام اول این طرح، تمرکز بر دانشگاه‌ها و سازمان‌های پژوهشی عمده است. در مراحل بعدی، این طرح می‌تواند به سایر طرح‌های بزرگتر نیز بپیوندد. (Nelson 2005)

علوم انسانی و اجتماعی در استرالیا، در افزایش آگاهی‌رسانی مشکلات این رشته برای تعیین سنجه‌های مناسب ارزیابی کیفیت موفق بوده است. از میان سنجه‌های کیفی، آن دسته از آنها که بر اساس تحلیل استنادی مجلات ایجاد شده‌اند، بیشتر از سایرین پیشنهاد شده است. اما کمتر از یک چهارم بروندادهای علوم انسانی و اجتماعی در پایگاه استنادی آی. اس. آی. قرار دارند و این سنجه‌ها تنها برای همین یک چهارم مناسب هستند. بررسی‌های زیادی برای تعیین معیارهای معادل سنجه‌های کتابسنجی صورت گرفته است، اما برای رشته‌های علوم انسانی و اجتماعی مناسب نیستند. این بررسی‌ها دو مسیر را

¹⁶ Australian Bureau of Statistics (ABS)

پیش گرفته اند: تهیه فهرست‌های رتبه‌بندی شده از بروندهای انتشاراتی (مانند ناشران کتاب، کنفرانس‌ها، اجرا یا تجسم‌های هنرهای خلاقانه، و مجلات)، و بررسی سنجه‌های کتابسنجی جدید.

در مسیر سنجه‌های کتابسنجی جدید، از پایگاه تامسون (آی.اس.آی.) برای تعیین اسنادها و برای مهمترین بروندهای انتشاراتی در یک رشته استفاده شده است. در مورد علوم سیاسی، این امر به معنای استخراج اسنادها برای کتاب‌ها، فصلهایی از کتاب، و مجلاتی است که در این پایگاه‌ها نمایه نشده‌اند، همچنین اسنادهایی به مجلاتی که نمایه شده‌اند. پیش آزمون‌ها این روش را برای دو رشته علوم انسانی و اجتماعی یعنی تاریخ و علوم سیاسی به کار برده‌اند و دانشگاه‌های هر حوزه را برای ارزیابی نتایج کنار یکدیگر قرار داده‌اند. هر دو رشته از سازگاری با چنین سنجه‌هایی پشتیبانی زیادی می‌کنند، اما با این شرط که هر معیاری که بر اساس این داده‌ها است، باید توسط بازنگران و نه تنها از طریق فرمول تفسیر شوند تا توزیع مکانیزه شده‌ای از بودجه صورت گیرد.

با وجود پشتیبانی که از این رویکردهای جدید کتابسنجی در استرالیا می‌شود، برای به کارگیری آنها در بریتانیا و برای تعیین میزان اثربخشی آنها در آینده بررسی‌هایی صورت گرفته است.... در حالیکه کارشناسان استرالیایی باور دارند که سنجه‌های جدید کتابسنجی، تصویر درستی از نقاط قوت و ضعف مرتبط با رشته آنها را در استرالیا ارائه می‌دهند، اما رتبه‌بندی هدف‌مندی برای آزمون این موضوعات وجود ندارد. به عکس در بریتانیا، رتبه‌بندی موسوم به RAE 2001، را می‌توان برای ارزیابی کارایی این رویکرد مورد استفاده قرار داد. این که سنجه‌های مبتنی بر اسنادها، به تنهایی با رتبه‌بندی‌های آی.اس.آی. همبستگی زیادی دارند، در علوم سیاسی غیر قابل فرض است، هر چند که بر خلاف یافته‌هایی که از این موضوع در سایر رشته‌های دیگر پشتیبانی می‌کنند. (Butler and McAllister 2007)

فرصت‌های عمده‌ای که در آنها کیفیت پژوهش یک فرد یا یک گروه از افراد سنجیده می‌شوند،

عبارتند از: (Rudd 1988)

الف) فردی که متقاضی یک شغل یا ابقاء در یک سمت، در دانشگاه یا برخی سازمان‌های دیگر

آموزش عالی است.

ب) مقاله‌ای که برای انتشار ارسال می‌شود،

ج) درخواستی برای تأمین بودجه بیشتر ارسال می‌شود.

معیارهای اصلی مورد استفاده در ارزیابی کیفیت پژوهش عبارت‌اند از:

1. ارزیابی تلخیصی (همترازخوانی)

2. تعداد انتشارات؛

3. تحلیل استنادی؛

4. تعداد و ارزش سرمایه‌های پژوهشی و قراردادهای منعقد شده.

۲-۱۴- روش‌ها و سبکهای ارزیابی

دو رویکرد اصلی برای ارزیابی پژوهش وجود دارند: اولین رویکرد ارزیابی برآوردی^{۱۷} یا قانونی^{۱۸}

است که تمرکز آن بر این امر است که آیا پژوهشگران واقعاً هر آنچه که اظهار کرده‌اند انجام داده‌اند.

این رویکرد به اطمینان یافتن از قابلیت اندازه‌گیری کمک می‌کند اما به ندرت بازخورد مفیدی برای

بهبود طراحی و عملکرد برنامه‌های تحقیق و توسعه، طرح‌ها یا سازمان‌ها ارائه می‌نماید. به طور عمده این

نوع رویکرد با جنبه‌های کنترلی مدیریتی مدیریت پژوهش مرتبط است و سبب افزایش یادگیری سازمانی

یا بهینه‌سازی مسیر صرف هزینه در تحقیق و توسعه نمی‌شود.

دومین رویکرد رایج در ارزیابی پژوهش، همترازخوانی است که در آن افراد مطلع از روش‌های

مختلفی برای ارزیابی **تعالی** علمی، شایستگی، قابلیت بازتولید، اعتبار، صحت، یا نتایج ممکن از خطوط

¹⁷ compliance

¹⁸ forensic

واقعی یا پیشنهادی پژوهش در میزان کاربردها، استفاده از دروندادها، صحت، بیان دانش، و غیره استفاده می‌کنند. ارزیابی در این روش اغلب کیفی است و همترازخوانان روش‌شناسی‌های پژوهش را بر اساس دانش و تجربه خود ارزیابی و نتایج و کاربرد آنها را تحلیل می‌کنند. ارزیابی به این روش به ما می‌گوید که آیا یک پژوهش به درستی انجام یافته است، در حالی که رویکرد برآوردی یا قانونی به ما می‌گفت آیا آن پژوهش واقعاً انجام شده است یا نه. (Davis and Carden 1998)

هیچ یک از دو رویکرد ذکر شده، به عنوان روشی برای ارزیابی ربط یا تأثیر پژوهش محسوب نمی‌شوند. هر دو روش با گسترش پژوهش در زمان رشد اقتصادی ادغام شده‌اند یعنی زمانی که تحقیق و توسعه بر اساس این فرض که کسب فواید ناچیز مشکل کاربران است، به طور وسیعی حمایت شد. این امر پذیرفته شد که پژوهش عام‌المنفعه و نتایج نامطمئن (سیاری از طرح‌های پژوهشی رد می‌شدند) و یافته‌های غیرقابل پیش‌بینی داشتند. پرسش‌های ارزیابی باید با انسجام داخلی پژوهش و کنترل مدیریتی خارجی مطرح می‌شدند.

ارزیابی برآوردی و همترازخوانی محدودیت‌های آشکاری در روش‌های شناسایی ارزش از سرمایه‌گذاری در تحقیق و توسعه دارد. این رویکردها می‌توانند در تعیین اینکه ما کاری را درست انجام می‌دهیم، کمک کنند، اما در تصمیم‌گیری اینکه آیا ما کاری را درست انجام می‌دهیم، دارای محدودیت‌هایی هستند. افزایش درخواست‌ها برای نتایج پژوهشی مفید یا مرتبط با این امر را ضروری می‌سازد که ماهیت نتایج برای انواع مختلف سرمایه‌گذاری‌های تحقیق و توسعه باید شناخته و درک شود. بنابراین، عمل ارزیابی باید بسط داده شود تا بکارگیری ارزیاب‌های مستقل، استفاده از فرآیندهای تعاملی با کاربران و ذی‌نفعان، و پژوهش علمی در حوزه ارزیابی، شامل بررسی اقتصادی برای تخمین میزان برگشت سرمایه در تحقیق و توسعه را در حوزه خود داشته باشد.

اکنون در ارزیابی تحقیق و توسعه اطلاعات و روش‌های تخصصی وجود دارد تا جایی که ارزیابی میزان اثرگذاری تحقیق و توسعه باید بین منطق پژوهش و قضاوت‌های همترازخوانان، کاربران، یا ذی‌نفعان تعادل ایجاد کند، زیرا این قضاوت‌ها می‌توانند بسیار ناشی از علاقه‌های افراد باشند یا بسیار محدود به برداشت‌های عاملان بسیار مرتبط با پژوهش باشند. همچنین این ارزیابی‌ها باید طیفی از شاخص‌ها و مشاهدات ارائه شده توسط سازمان‌های ارزیابی شده، آژانس‌های حامی، دیدگاه‌ها، مشاوران، یا پژوهشگران دانشگاهی را در برداشته باشد. در ترکیب این روش‌ها می‌توان دیدگاهی مستقل، گذشته‌نگر بی‌طرف (و کم و بیش آینده‌نگرانه) از سیستم‌های پژوهشی یا بخش‌هایی از آن را ارائه داد و قضاوت‌های کارشناسان، همترازخوانان یا عوامل مرتبط را تکمیل کرد. در جهان انگلیسی زبان و در سازمان‌های منطقه‌ای و بین‌المللی بسیاری از آژانس‌هایی با مسئولیت سیاسی یا پژوهشی وجود دارند که واحدهایی از خود را همراه با کارکنان حرفه‌ای به ارزیابی اختصاص داده‌اند

اما، تمایل اندکی به استفاده از روش‌های پیشرفته در ارزیابی میزان اثرگذاری تحقیق و توسعه وجود دارد. حتی در ایالات متحده که در سال 1990، 22 میلیارد دلار صرف پژوهش شده است، این امر صادق است. در حالی که در کشورهای در حال توسعه تمایل و توانایی برای اجرای پژوهش‌های مرتبط با ارزیابی یا استفاده از روش‌های پیچیده ارزیابی کم می‌باشد. به طوریکه بسیاری از کشورهای در حال توسعه برای ایجاد شاخص‌های اساسی مناسب برای فعالیت‌های ملی در عرصه علم و فناوری با مشکل روبرو هستند و شاخص‌های ایجاد شده بسیار کلی‌تر از آن هستند که بتوانند برای ارزیابی‌های مفید پژوهش‌ها استفاده شوند.

برخی از مشکل‌ترین حوزه‌های ارزیابی پژوهش، مربوط به مسائل تعریف برون داده‌های پژوهشی است. به دلیل اینکه درون داده‌ها ملموس‌تر هستند، از برون داده‌ها شناخته شده تر هستند. کالون و همکارانش (1995) گونه شناسی از پنج بعد از برون داده‌های پژوهشی ارائه کرد: اسناد و دانش تأیید شده،

آموزش و مهارت، فواید رقابتی (نوآوری)، جامع یا عام‌المنفعه، و مهارت و شهرت. سازمان‌های تحقیق و توسعه صدها نوع مختلف از برون‌دادهای واسطه‌ای و نهایی قابل کاربرد برای نظارت عملکرد و ترابندیت تولید می‌کنند که شامل انواع خدمات ارائه شده به کاربران تا تولید الگوها می‌شود. برای ارزیابی عملکرد در حوزه انتقال فناوری بین سازمانی، سازمان‌ها نیاز دارند تا شاخص‌هایی برای درون‌دادها، فرآیندهای انتقال، و برون‌دادها داشته باشند. سازمان‌های بسیار کمی قادر به نظارت تولیدات خود در چنین طیفی از برون‌دادهای پژوهشی در یک روش علمی هستند.

ارزیابی پژوهشی نه تنها برون‌دادها بلکه نتایج (مثلاً نتایج و اهمیت نسبی مربوط به برون‌دادها) را باید در نظر گیرد، و باید آن را از لحاظ اهداف یا آرمان‌هایی که برای تحقق آنها برون‌دادها به وجود می‌آیند درک شوند. در مقایسه با اهداف پژوهشی شرکتها یا دانشگاه‌ها در کشورهای شمال که ایجاد ارزش یا ایجاد دانش است، پژوهش‌های توسعه‌ای بین‌المللی طیف گسترده‌ای از اهداف اجتماعی، اقتصادی، و محیطی را در بر می‌گیرد.

در میان مدیران تحقیق و توسعه در مرکز تحقیق و توسعه بین‌المللی (آژانسی که روی پروژه‌های تحقیق و توسعه در کشورهای جنوب سرمایه‌گذاری می‌کند)، «پژوهش اثربخش» یکی از چهار هدف کلی را در بردارد؛ می‌تواند بر قوانین اجرایی سازمان یا حکومت از طریق تولید دانش مرتبط یا وابسته به سیاست‌گذاری تأثیر گذارد؛ می‌تواند با تولیدات خود امکاناتی را برای کاربران فراهم کند که آنها بتوانند در خانه، شرکت یا در کشاورزی از آن بهره‌گیرند؛ می‌تواند با فراهم کردن دانش معتبر در کشورهای جنوب این امکان را برای پژوهشگران مهیا کند تا به راحتی صحبت کنند، می‌تواند گروه‌های حاشیه‌ای یا محروم را از طریق فراهم کردن حمایت فنی برای افرادی که از مزیت‌های علم و فناوری محروم هستند، منتفع سازد. این طیف از برون‌دادهای تحقیق و توسعه، فواید، و نتایج خصیصه بافت

پژوهشی، مشکلات استانداردسازی، و نیاز به استفاده از انواع شاخص‌های تأثیر، ربط، و کیفیت را اظهار می‌کنند.

دو عامل کلیدی در استفاده از ارزیابی‌های تحقیق و توسعه عبارتند از: 1- تنظیم زمانی ارزیابی به طور که نتایج در مرحله طراحی نسل بعدی طرحها قابل دستیابی باشند، 2- تمایل برای مورد توجه قرار دادن اهداف حامیان پژوهشی، مثل «ارزش برای پول»، یا «تأثیر بر رقبا». در قاره اروپا که برنامه‌های تحقیق و توسعه در بین چند کشور انجام می‌شوند، ارزیابی تحقیق و توسعه مستلزم ارزیابی تأثیرات ساختاری هزینه‌های عمومی صرف شده در علم و فناوری از طریق پانل‌هایی با حضور کارشناسان، کاربران، و ذی‌نفعان صورت می‌گیرد.

ارزیابی پژوهشی که برای اهداف نظارتی صورت نمی‌گیرد یکی از عناصر مکانیزم مدیریت و سیاست‌گذاری برای هدایت تولید دانش برای تحقق برخی اهداف است. به عبارتی دیگر، شکلی از ارزیابی آینده‌نگرانه است. فشار خارجی بر روی سازمانهای پژوهشی شرطی لازم برای عملکرد سطح بالای تحقیق و توسعه است، اما کافی نیست.

در بین گروه‌های ذی‌نفع، شامل حامیان و کاربران عقاید درباره چگونگی راهبری پژوهش، از مدل‌های انتخاب راهبردی که نیاز به حجم زیادی از اطلاعات و کنترل متمرکز خارجی دارد، به مسأله بازاریابی و مکانیزم قیمت‌گذاری و طراحی مکانیزم‌هایی برای اطمینان از تعامل پایدار گسترش یافته‌اند. در برخی کشورها، اژانس‌های عام‌المنفعه سرمایه‌های راهبردی خود را در مکانیزم‌های سازمانی تعاملی متمرکز می‌کنند تا به تولیدکنندگان دانش، کاربران، و ذی‌نفعان این امکان را فراهم کنند تا با یکدیگر فرآیندهای تولید دانش را در روش‌هایی که به شایستگی آنها ایمان دارند، راهبری کنند. (Davis and

(Carden 1998

برا (1990) سه فرآیندی را که بخش مدیریت باید بر آن نظارت داشته باشد، تعریف می‌کند که سیستم‌های علم و فناوری بایستی طبق آنها راهبری شوند: ارزیابی طرحها و سازمان‌های پژوهشی، که عملکرد عاملان علم و فناوری را مورد سنجش و ارزیابی قرار می‌دهند؛ تحلیل آینده‌نگرانه و راهبردی برای علم و فناوری، که مشکلات سیستم‌های علم و فناوری را هشدار می‌دهد؛ و ارزیابی فناوری، که نتایج اجتماعی نوآوری‌های حاصل شده در زمینه فناوری را بررسی و مورد قضاوت قرار می‌دهد. با برخی تفاوت‌ها، اگر انتخاب خارجی بایستی بر تولید دانش در منطقه‌ای معین صورت می‌گیرد، پس هر سه شرط باید در نظر گرفته شود: روندهای تخصیص منابع تحقیق و توسعه باید غیرشخصی و منسجم باشند؛ اطلاعات مناسب باید درباره سیستم‌های پژوهشی، محیط آن، و برون‌دادها و تأثیرات آنها قابل دستیابی باشند؛ و ارزیابی و دیگر اطلاعات باید دارای ارتباط زمانی مناسب و مفید با تصمیم‌گیری باشد. (Barré 1990)

۲-۱۵- ابعاد ارزیابی

پنج بعد اصلی برای ارزیابی در نظر گرفته شده است که اسکیروان آنها را ارزیابی‌های فرعی می‌نامد. این پنج بعد عبارتند از: فرآیند، نتایج، هزینه‌ها، تطبیق، و قابلیت تعمیم‌پذیری. هر یک از این ابعاد نیاز به ترکیب هر دو اطلاعات ارزش و واقعیت را برای طرح ادعاهای ضروری ارزیابی دارند. در برخی موارد، این ابعاد می‌توانند به صورت مجموعه‌ای از شایستگی، ارزش، و اهمیت ترکیب شوند.

در حالی که هدف اصلی ابعاد اصلی ارزیابی کاربرد در ارزیابی برنامه‌ها، و در سطحی پایین‌تر، سیاست‌ها است، با این وجود آنها در برخی موارد برای ارزیابی محصولات، عملکردها، افراد، طرحهای پیشنهادی و خط‌مش‌ها مفید هستند. برای ارزیابی پژوهشگران و پژوهش‌ها شاید لازم باشد همه این ابعاد به کار گرفته شوند. در ارزیابی یک اثر پژوهشی، ارزیابی فرآیند شاید در همه موارد مرتبط نباشد، مگر

آنکه این باور به وجود آمده باشد که ارزش قانونی یا اخلاقی آن نقض شده است، یا اگر فرآیند تحقیق برای تعیین شایستگی‌ها یا ارزش پژوهش حیاتی باشد (برای مثال، آیا اسناد مورد استفاده شده حائز ارزش هستند؟ آیا پیشینه تهیه شده کافی است؟)، در مواردی که هزینه‌ها به اندازه کافی کم هستند و برای بررسی چندان حائز اهمیت نیستند، نیازی به ارزیابی هزینه نخواهد بود. (Coryn 2007)

۲-۱۵-۱- ارزیابی فرآیند

ارزیابی فرآیند، به طور معمول ارزیابی هر چیزی است که پیش از ظهور نتایج واقعی رخ می‌دهد. در بیشتر موارد، ارزیابی فرآیند شامل ارزیابی مزایا، ارزش، یا اهمیت برون‌دادها، چشم‌انداز، طرح، برنامه‌ریزی، عملیات، و توجیه (مثلاً اهداف)، صحت، مدیریت، فعالیت‌ها، روندها، و غیره است. برون‌دادها، محصولات ملموس منتج از فعالیت‌های مسأله مورد ارزیابی هستند، که اغلب برای تولید نتایج واقعی اجرا می‌شوند. در حوزه پژوهش، این انواع برون‌دادها می‌توانند شامل انتشارات تولید شده توسط پژوهشگران یا گروه‌های پژوهشی، ثبت اختراعات تولید شده از طریق فرآیند تحقیق و توسعه، یا یک محصول یا فناوری تحقیق و توسعه باشند. برون‌دادها معمولاً در حوزه ارزیابی فرآیند قرار می‌گیرند، اما وقتی که اهمیت آنها زیاد باشد، باید در ارزیابی نتایج مورد بررسی قرار گیرند.

بعد صحت ارزیابی فرآیند در بافت پژوهش، برای نمونه، می‌تواند شامل نحوه اجرای پژوهش بعد از طراحی یا توافقات صورت گرفته باشد، در صورتی که روندهای مدیریت، اجرا، و دیگر روندها باید این عناصر را نه تنها در محدوده‌های قانونی و اخلاقی، بلکه به لحاظ اثربخشی و کارایی نیز مورد ارزیابی قرار دهد.

ارزیابی فرآیند پژوهش می‌تواند شامل طیف گسترده‌ای از مسائلی چون شایستگی پیشینه تحقیق، اعتبار ابزارهای به کار گرفته شده، و غیره باشد. همچنین، این ارزیابی به طور آرمانی به ارزیابی متغیرهای مستقیم فرآیند که می‌توانند با نتایج همراه باشند، اشاره دارد.

۲-۱۵-۲- ارزیابی نتایج

ارزیابی نتایج، سنجش اثرات مثبت و منفی مورد ارزیابی است. این تأثیرات شامل تأثیرات مستقیم و غیرمستقیم، خواسته و ناخواسته، و ابتدایی (فوری)، میانی (کوتاه مدت/میان مدت)، و انتهایی (بلندمدت) می باشند. نتایج یا تأثیرات می توانند انفرادی، چندگانه، یا سلسله مراتبی باشند. نتایج یا تأثیرات ابتدایی اغلب به برون داده‌ها اشاره دارند و معمولاً زیر پوشش ارزیابی فرآیند قرار می گیرند. اما گاهی هم زیر پوشش ارزیابی نتایج قرار می گیرند، مخصوصاً «اگر نقش آنها به عنوان واسطه یا علت اصلی نتایج اصلی باشند». اکتشاف مهم در یک پژوهش می تواند از این نوع نتایج محسوب شود. به عکس، نتایج میانی و انتهایی که معمولاً وابسته به زمان هستند، اغلب کند رخ می دهند، و گاهی مشاهده آنها بسیار مشکل است. برای نمونه، استنادها به پژوهش‌های منتشر شده اغلب به عنوان نتایج پژوهش شناخته می شوند، یعنی آنها به عنوان سنجه‌ای از کیفیت یا تأثیر علمی محسوب می شوند. اما استنادها معمولاً تا اولین سال پس از انتشار اثر ظاهر نمی شوند و طی دو یا سه سال به طور متوسط زمان به اوج رسیدن تعداد استنادها است. در مورد پژوهش‌ها و پژوهشگران بسیار مهم، ارجاع‌ها (استنادها) اغلب ادامه می یابد و در برخی موارد برای ده‌ها سال پس از انتشار اثر اولیه استمرار دارد. معمولاً و نه همیشه، نتایج و اثرات میانی معمولاً پیش شرطی برای نتایج یا اثرات انتهایی محسوب می شوند، مخصوصاً به اندازه‌ای که پژوهش اثر ماندگار تری بر دانش بگذارد.

اما، همه نتایج به راحتی ظاهر نمی شوند، و این امر برای اکثریت پژوهش‌ها صادق است. برای مثال، هدف پژوهش در فیزیک اغلب درک بهتر قوانین طبیعت است که رفتار موضوع و انرژی را اداره می کند. پژوهش درباره مواد ابرهادی در دمای پایین، می تواند نتایج نهایی دانش ترکیب موادی باشد که در دمای اتاق ابرهادی هستند. اما، این نتایج ممکن است در طی چند سال یا اصلاً ظاهر نشوند. همچنین،

این پژوهش می‌تواند اظهار کند که چنین موادی نمی‌توانند ایجاد شوند و به دنبال آن نتایجی در آینده گرفته شوند.

۲-۱۵-۳- ارزیابی هزینه

حتی اگر آنچه که مورد ارزیابی قرار می‌گیرد، نتایج بسیار ارزشمندی داشته است، باید از لحاظ اینکه این نتایج با چه هزینه‌ای حاصل شده‌اند (ارزش‌های اقتصادی منابع)، بررسی صورت گیرد و تصمیم‌های آگاهانه نیاز به اطلاعات در هر دو مورد نتایج و هزینه دارد. معمولاً، هزینه‌ها به عنوان بار منفی ظاهر می‌شوند که در ایجاد یا کسب چیزی وارد می‌شوند. ارزیابی هزینه‌ها شامل ارزیابی هزینه‌ها شامل ارزیابی هزینه‌های مالی و غیرمالی، هزینه‌های مستقیم و غیرمستقیم، و هزینه‌های تمام شده و هزینه کالای تولید شده^{۱۹} است.

این بعد از ارزیابی، نه تنها موارد کلاسیکی چون هزینه سودمندی، اثربخشی هزینه، سودمندی هزینه، امکان‌سنجی مالی، تحلیل سرمایه، و تحلیل سهام مالی، را در بر می‌گیرد، همچنین به ارزیابی هزینه‌هایی که به ندرت مالی هستند مانند استرس‌ها، سرمایه‌های فردی و سیاسی، اثرات محیطی، و هزینه‌های عدم مشارکت نیز می‌پردازد. این هزینه‌ها می‌توانند به طور معمول در سه بعد دسته‌بندی شوند: انواع هزینه‌ها، هزینه‌های مربوط به افراد، و هزینه‌های زمانی.

هزینه‌های فرصت (هزینه‌های تولید کالا)، اهمیت زیادی در ارزیابی پژوهش دارد. در تلاش‌هایی که برای بهینه‌سازی منابع کمیاب، صورت گرفته است، سرمایه‌گذاران پژوهشی باید تصمیم بگیرند که به چه طرح‌های پیشنهادی باید بودجه اختصاص دهند و به چه میزانی، همچنین طرح‌های پیشنهادی که بدون بودجه باقی مانده‌اند. این هزینه‌ها همچنین در تنظیم الویت‌ها و سرمایه‌هایی نقش دارند که بدون آن این الویت‌ها را دریافت نمی‌کنند. این تصمیم‌ها اغلب در گیر ارزش‌های ریسک‌پذیر، مانند هزینه‌های

¹⁹ opportunity costs

شکست در برابر موفقیت، و همان‌طور که گیلوویچ اشاره می‌کند، عقیده «پرتاب موفق» هستند، یعنی درک این نکته که «موفقیت، موفقیت در پی دارد و شکست، شکست»

۲-۱۵-۴- ارزیابی تطبیقی

ارزیابی تطبیقی، یا مقایسه‌ها، مورد ارزیابی را با موارد دیگر یا رقبای اصلی خود مقایسه می‌کند. یعنی، مورد ارزیابی با روش‌های دیگری که همان فواید یا مشابه آن را از همان منابع به دست آورده‌اند، تطبیق داده می‌شود. معمولاً این نوع ارزیابی نیاز به دست کم تطبیق با یک روش اقتصادی دارد که به یک اندازه کارا باشد، همچنین یک روش پرهزینه‌تر که اگر چه پرهزینه‌تر است، سود بیشتری را تولید می‌کند. این مقایسه‌ها باید درون محدودیت‌های منابع قابل دسترس مورد ارزیابی قرار گیرند. ارزیابی تطبیقی به عنوان بخشی از فرآیند تعیین الویتهای پژوهشی ملی یا سرمایه‌گذاری‌های مربوط به آن و نیز در بازنگری طرح‌های پیشنهادی پژوهشی برای تأمین اعتبار به صورت ضمنی یا صریح انجام می‌شود.

۲-۱۵-۵- ارزیابی تعمیم‌پذیری

ارزیابی تعمیم‌پذیری کم و بیش معادل مفهوم اعتباریابی خارجی است، یعنی گستره‌ای مورد ارزیابی، یا برخی عناصر یا ابعاد آن می‌تواند به دیگر شرایط با نتایج مشابه تعمیم داده شود. این بعد شامل تعمیم به شرایط دیگر (شرایط اجتماعی، سیاسی، و فیزیکی)، دیگر کارکنان و نیروی انسانی در مقیاس کمتر یا بیشتر (مقیاس‌بندی)، دیگر دریافت‌کنندگان (یعنی جامعه واقعی)، همچنین به قابلیت‌های صدور، انتقال، حمل، پایداری، طول عمر، ماندگاری، و انعطاف مسأله مورد ارزیابی است. اسکریون ترکیب این موارد را به عنوان «ابعاد تعمیم‌سازی» اطلاق می‌کند.

اساس چنین استنباط‌هایی دانش کامل درباره‌ی مورد ارزیابی در گستره‌ای وسیع است، تا این تعمیم‌ها عاری از هر گونه تردید باشند، نه اینکه از طریق تجربه چند مورد عملی حاصل شده باشند. در اصل، برای ارزیابی تعمیم‌پذیر لازم است که پیش‌بینی‌هایی درباره نتایج سناریوهای دیگر انجام شود، و هر چند

پرمخاطره است، اما این امر گاهی بیشترین نقش ارزیابی را برای بهبود جهان خواهد داشت. ارزیابی تعمیم‌پذیر، چنانکه شرح داده شد، نیاز به استنتاج‌های کمی متفاوت برای ارزیابی علم و فناوری و تحقیق و توسعه نسبت به آنچه که برای یک پژوهشگر یا یک نمونه یا بخشی از پژوهش در بیشتر موارد دارد، یعنی تعمیم‌پذیری طرح‌های تحقیق و توسعه یا علم و فناوری به شرایط گوناگون، قابلیت تغییر یا کارایی یک پژوهشگر بعد از انتقال به مؤسسه، واحد، یا مرکز دیگر، یا تعمیم‌پذیری نتایج پژوهشی به موارد مشابه که در همه این موارد منطق اصلی (یا گستره‌ای که ارزیابی یا ابعاد آن می‌توانند به نمونه‌های گوناگون با نتایج یکسان تعمیم داده شوند) یکی است.

۲-۱۶- حوزه‌های ارزیابی

در حال حاضر، چندین حوزه یا شاخه فرعی برای ارزیابی شناسایی شده است، که از آن جمله، «هفت بزرگ» (Scriven, 2005d, p. 237 in Coryn 2007) است که حوزه‌های ارزیابی را محصولات، عملکردها، افراد، برنامه‌ها، خط‌مشی‌ها، طرح‌های پیشنهاد، و دارایی معرفی می‌کند. «هفت بزرگ» سابقه عملی زیاد و در عین حال سابقه کمتری برای بحث‌های روش‌شناسی دارد. در بیشتر موارد، ارزیابی پژوهش یا پژوهشگر نمونه‌ای از ارزیابی محصول و عملکرد است، عملکرد یک کشور، سازمان، گروه پژوهشی، یا پژوهشگر و محصولات پژوهشی آنها نیز اغلب مستلزم بکارگیری عناصر ارزیابی افراد و طرح‌های پژوهشی است. در هر صورت، منطق کلی کار ارزیابی در هر هفت حوزه مشابه است.

هر یک از این هفت حوزه، با ارزیابی پژوهش پیوند زیادی دارند و سطوح این پیوندها در حال افزایش است. برای نمونه، ارزیابی افراد پژوهشگر و طرح‌های پژوهشی نمونه‌هایی از ارزیابی در سطح بسیار کوچک است، اما ارزیابی برنامه‌های پژوهشی، سازمان‌ها، یا حوزه‌ها منطبق با ارزیابی در سطح بزرگ است. در صورتی که عملکرد یا دارایی پژوهشی می‌تواند در سطح کوچک ارزیابی شوند. (بسته

به این که دارایی یا عملکرد متعلق به یک فرد پژوهشگر باشد یا یک سازمان یا مؤسسه پژوهشی). همچنین، ارزیابی در سطح بسیار کوچک محصولات یا افراد پژوهشی، اغلب اساس سطوح کوچک یا نسبتاً کوچک ارزیابی برنامه‌ها یا خط مشی‌ها هستند.

۲-۱۶-۱- ارزیابی برنامه

از میان حوزه‌های هفت بزرگ، ارزیابی برنامه‌ها بیشترین توجه‌ها را به سوی خود جلب کرده است و بهترین اصول، رویه‌ها، و امور را دارد. با اینکه بیشتر مورخان عرصه ارزیابی ظهور ارزیابی برنامه‌های نظام‌مند را به دهه ۱۹۶۰ نسبت می‌هند، اما ادعا شده است که ریشه آن به قرن نوزدهم در بریتانیا برمی‌گردد. که در آنجا کمیسیون‌های دولتی، از طریق ارزیابی برای اصلاح آموزش، قوانین، بیمارستان‌ها، پرورشگاه‌ها، و بهداشت همگانی تلاش می‌کردند.

ارزیابی برنامه‌ها چه در بریتانیا در نیمه قرن ۱۸۰۰ ظهور یافته باشد یا در ایالات متحده در نیمه قرن ۱۹۰۰، شکوفایی آن با توسعه روش‌ها، مدل‌ها، نظریه‌ها، رویکردهای خاص ارزیابی صورت گرفته است. اما ارزیابی هنوز هم باید از لحاظ نظری توسعه یابد، هرچند که اسکریون بنیادهای آن را مدل چندبعدی خود تشریح کرده است. ارزیابی برنامه‌های پژوهشی، مراکز تحقیق و توسعه، مؤسسه‌ها، و ... سابقه نسبتاً طولانی دارند. اما در درک و کاربرد منطق آنها ضعف‌های جدی وجود دارد.

۲-۱۶-۲- ارزیابی افراد

در حوزه ارزیابی افراد، اولین تلاش‌های مستند مربوط به سلسله پادشاهان چینی و مصری در بیش از ۲۰۰۰ سال پیش است. این ارزیابی‌ها برای مشخص کردن بهترین استعدادها برای دولت انجام می‌شدند و در زمان امپراطوری تانگ (۶۱۸-۹۰۷ بعد از میلاد) در چین تکمیل شد. در نیمه دوم قرن بیستم، ارزیابی مدرن افراد با مدیریت علم تیلور (۱۹۱۱-۲۰۰۳) که بر کارایی تأکید داشت، شکل گرفت.

۲-۱۶-۳- ارزیابی عملکرد

به دلیل اینکه ارزیابی عملکرد به سنجش و آزمون مربوط است، کمی بیش از ارزیابی افراد توسعه پیدا کرد. در سال 1972، وقتی که ویلیام فاریش نمره گذاری کمی برای پاسخ های صحیح و نادرست را که امکان معدل گیری نمره های آزمون ها را فراهم می کرد، جایگزین سنجش های کیفی سنتی عملکرد دانشجویان نمود، روش های مدرن هوش سنجی شروع به توسعه کرد.

در حالی که ارزیابی عملکرد در ابتدا برای سنجش عملکرد دانشجویان در آزمون ها برای دانشجو، کلاس، مدرسه، یا یک منطقه استفاده می شد، اما سابقه طولانی در استفاده در داوری عملکرد و قهرمانی های ورزشی در المپیک و سایر رقابت های ورزشی دارد. ارزیابی عملکرد به عنوان اساسی برای ارزیابی انسجام عملکرد یک پژوهشگر یا یک مرکز تحقیق و توسعه در رابطه با کیفیت آنها دارد. اغلب، عملکردهای گذشته اساسی برای تخصیص منابع، مانند سرمایه گذاری پژوهشی هستند و نقش مهمی در ارزیابی طرح های پیشنهادی دارند.

۲-۱۶-۴- ارزیابی محصول

ارزیابی محصول، پیشینه طولانی دارد و از هزاران سال پیش هنرمندان، صنعتگران، پیشه وران، و جوامع حرفه ای محصولات را ارزیابی می کردند. اما در 50 سال گذشته، این ارزیابی به طور چشم گیری گسترده تر، فراگیرتر و پیشرفته تر شده است. بر خلاف سابقه طولانی کاربرد ارزیابی محصول، بخصوص ارزیابی محصولات مشتریان، به دلیل وابستگی به کیفیت زندگی، کمبودهایی در این ارزیابی به چشم می خورد که درک ناکافی منطق این ارزیابی از جمله آنها است.

همچنین عدم هدف گرایی، منطق نادرست فنون کاربرد، زیبایی، و مهارت های سوگیرانه و نامرتبط از جمله کاستیهای آن است. این مسائل در ارزیابی محصولات پژوهشی مخصوصاً در کاربرد معیارهای

صریح کیفی برای تعیین ماهیت یک محصول کیفی وجود دارند، مثلاً آن پژوهش اگر از دیدگاه فلسفی یا معرفت‌شناختی انجام شود، خوب است.

۲-۱۶-۵- ارزیابی خط مشی

ارزیابی خط مشی در ایالات متحده در دهه‌های ۱۹۶۰ و ۱۹۷۰ به طور همزمان با ارزیابی برنامه‌ها و از زمان پژوهش‌های کاربردی، اقتصاد خرد، نظریه سازمانی، مدیریت عمومی، روانشناسی اجتماعی، و افزایش علاقه به نقش قانون در سیاست‌گذاری‌های عمومی ظهور یافت. هرچند این حوزه نقاط مشترکی با ارزیابی برنامه‌ها (مانند منطق کاری و روش‌شناسی) دارد، ارزیابی خط مشی معمولاً هم گذشته‌نگر است و خط‌مشی‌های اجرا شده را ارزیابی می‌کند و هم ارزیابی آینده‌نگر برای خط‌مشی‌های ممکن یا مقایسه با خط‌مشی‌های دیگر است.

برخلاف حوزه‌های دیگر ارزیابی، ارزیابی خط‌مشی بخصوص نوع آینده‌نگارانه آن، اغلب مسائل در حال حاضر را مورد بررسی قرار می‌دهد، که نیاز به روش‌های رسمی برای مدل‌سازی برای تعیین مزیت‌ها، و ارزش‌های هر آنچه که دولت‌ها برای انجام یا عدم انجام انتخاب می‌کنند. یا ارزیابی خط‌مشی‌ها یا تصمیم‌های دیگر دولتی که برای رسیدن به بهترین خط‌مشی‌ها انجام می‌شوند. این حوزه در ارزیابی پژوهش دو نقش دارد: اول اینکه خط‌مشی‌های مسائلی چون الویت‌ها، سرمایه‌گذاری و تخصیص و توزیع سرمایه‌ها را ارزیابی می‌کند، ثانیاً، این حوزه خط‌مشی‌هایی را ارزیابی می‌کند که پژوهش‌های با سرمایه دولتی مرتبط هستند.

۲-۱۶-۶- ارزیابی دارایی

ارزیابی دارایی حوزه جدیدی از ارزیابی است که بیشتر بخش‌های آن در دست توسعه است. در حالیکه در دنیای تجارت، دارایی اشاره به میزان سرمایه‌گذاری‌ها دارد، در هنر، معماری، موسیقی، آموزش، فناوری، و بیشتر شکل‌های پژوهش معمولاً به کل یا منتخبی از روش‌های حرفه‌ای اطلاق

می‌شود. در حدود دهه 1980، استفاده از ارزیابی دارایی برای ارزیابی دانشجویان در حوزه‌هایی چون نگارش، ریاضی، انگلیسی، آغاز شد و در برخی موارد جایگزین یا مکمل ارزیابی عملکرد، آزمون و سنجش به عنوان شاخص‌های موفقیت و یادگیری دانشجویان شد. در اوایل دهه 1990، این ارزیابی برای ارزیابی آموزش، پژوهش، و خدمات اعضای هیئت علمی به کار گرفته شد.

اخيراً دارایی‌های وبی برای ارزیابی دوره‌ها، برنامه درسی و مؤسسات آموزشی استفاده می‌شود. اما، ارزیابی دارایی هنوز باید برخی از مشکلات جدی خود را حل کند، نظیر چگونگی برخورد با معیارها یا ارزش‌های متناقض مانند تنوع در برابر تخصص، ریسک در برابر عدم ریسک، و کمیت در برابر کیفیت. این حوزه اغلب اساسی برای ارزیابی عملکرد پژوهشگر، ارزیابی افراد پژوهشگر، یا برای سرمایه‌گذاری‌های ملی در علم و تحقیق و توسعه است.

۲-۱۶-۷- ارزیابی طرح پیشنهادی

ارزیابی طرح‌های پیشنهادی و بخصوص طرح‌های پیشنهادی پژوهشی که برای تأمین اعتبار ارسال می‌شوند، پیشینه نسبتاً طولانی دارند و تا حدی با امور مرتب سازی و پیشگویانه مشخص می‌شوند. این نوع از ارزیابی‌ها معمولاً ارزیابی توانایی پیشنهاد دهنده برای اجرای موفق پژوهش در آینده و ارزش نتایج پژوهشی پیش‌بینی شده است.

معمولاً این ارزیابی‌ها از طریق پانل‌های هم‌ترازخوانی برای قضاوت یا برای پیش‌بینی عملکرد آینده بر اساس عملکرد گذشته انجام می‌شوند. این نوع از ارزیابی معمولاً برای تعیین مزیت‌های عقلایی و اهمیت پژوهش پیشنهاد شده به کار می‌رود. اما، همان‌طور که به تازگی شاهد هستیم، این نوع از ارزیابی کاستی‌های جدی در مطالعه و اجرای آن دارد. برای نمونه، کمیته‌های بازنگری طرح‌ها و داوران اغلب با کمبود رویه‌های استاندارد و قاعده مند برای سنجش مواجهند، این ارزیابی‌ها اغلب از طریق علائق خاص دست‌کاری می‌شوند، اغلب به لحاظ سیاسی کنترل می‌شوند، باعث تضارب علائق می‌شوند، و به طور

صحیح از بین نمی روند، گاهی فقط هرج و مرج های امور بازنگری طرح های پیشنهادی را ساده می کنند، و غیره.

۲-۱۷- مدل های ارزیابی

۲-۱۷-۱- مدل *Transdisciplinary* ترارشته ای

ترارشته ای یعنی آنچه که بنیانی متفاوت از رشته های اصلی دارد، برای نمونه رشته های سنتی آکادمیک و دسته ای از رشته ها که برخی ابزارها، روش ها یا رهیافت ها را برای استفاده رشته های اصلی فراهم می کنند. رشته ای در بین دیگر رشته ها شامل آمار، سنجش، منطق، و ارزش یابی است. منطق، با حوزه های کاربردی منطق علوم اجتماعی و غیره رشته بسیار عمومی محسوب می شود. اما ارزشیابی احتمالاً عمومی تر از همه است. اما هر دوی این ها عمومی تر از آمار یا سنجش هستند.

در دیدگاه یا مدل ترارشته ای نیاز به درک چگونگی و چرایی توسعه ارزشیابی از یک اقدام به اقدامی بسیار تخصصی و حرفه ای، به رشته ای تخصصی، و سرانجام به رشته و ترارشته ای مستقل مانند اخلاق، آمار، و سنجش است. این درک، تا حدی از سه ویژگی هرم مدل ترارشته ای به دست می آید که آن را بین رشته ای می سازد. این ویژگی ها عبارتند از معرفت شناسی، علوم سیاسی و علوم انتظامی.

ویژگی معرفت شناسانه مدل ترارشته ای ناشی از دید بی طرفانه در ارزشیابی است. این ویژگی مفهوم متضادی است و بر خلاف معانی یا تعریف های موجود از مفاهیم ارزیابی در رشته ها، اندیشه ها، یا از سوی اشخاص مختلف، ناشی است که از یک واقعیت مجرد ارائه می شود. عینیت به عنوان قابلیت سازگاری گزاره های عینی محسوب می شود و از اقدام ها یا نگرش های گزاره های انتزاعی مستقل و متمایز است. اصل یک گزاره درست یا نادرست بودن آن است و انواع شکل های آن شامل قضیه ها و فرمول های علمی و ریاضی و همچنین قوانین و فرآیندهای منطقی است. بنابراین، دید عینی گرایی

ارزشیابی اظهار می کند شایستگی، ارزش، و اهمیت مورد ادعای ارزیابی در اصل و عمل امکان پذیر است و اصل آن بر منطق و دلیل و در صورت درک درست، عینیت قرار دارد.

بنابراین، ارزشیابی عینی گرا بر اساس این نظریه است که محصول اخلاقی نیز عینی است و مستقل از احساس انسانی یا فردی است

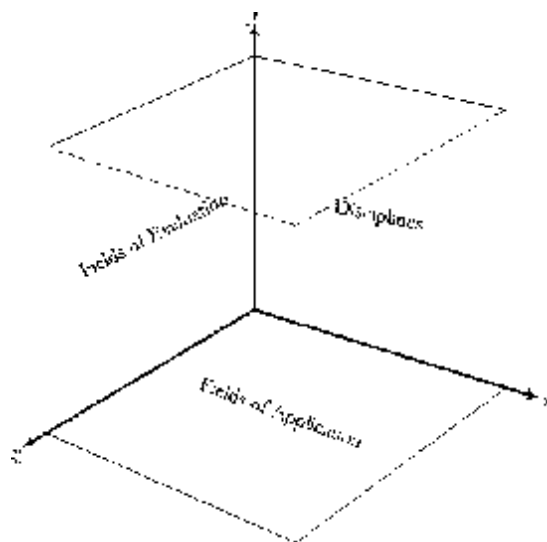
ویژگی سیاسی مدل ترارشته ای این است که از دید مشتری گرایی تنظیم شده است. یعنی، منطق یا توجیه وجود یک برنامه، خط مشی، یا محصول باید نیازهای کاربران را تأمین کند در این دیدگاه، ارزیابی برای کاربران الویت قائل است و بنابراین کارکرد اصلی آن تعیین شایستگی، ارزش، یا اهمیت یک برنامه، خط مشی، یا محصول به لحاظ میزان کارایی و اثرگذاری بر کاربران است، مخصوصاً آنهایی که خدمات یا محصول را دریافت می کنند، یا باید دریافت کنند، و آنهایی که در ازای این دریافت هزینه پرداخت می کنند. هر چند دیدگاه مشتری گرایی به لحاظ منطقی و عقیده عمومی پذیرفته شده است، با این وجود استافلیم در رده بندی و تحلیل خود از مدل های ارزیابی و رهیافت های آن، ارزشیابی مشتری گرا را به عنوان مسأله بسیار بزرگی عنوان می کند که مستلزم مهارت و تخصص بسیار است.

ویژگی رشته ای مدل ترارشته ای، شبیه آمار، اخلاق، و منطق است که ارزیابی رشته ای محسوب می شود و می تواند با مطالعه آن و بهبود برخی ابزارها برای کاربرد میان و درون رشته های دیگر مورد استفاده قرار گیرد. ویژگی رشته بودن دیدگاه ترارشته ای ارزیابی می تواند به سه بخش تفکیک شود: رشته ها (مانند هنر، علوم انسانی و اجتماعی، فناوری، علوم طبیعی)، حوزه های ارزشیابی (یعنی محصولات، نحوه اجرا، افراد، برنامه ها، خط مشی ها، طرح های پیشنهادی، و دارائی ها)، و حوزه های کاربردی (مانند آموزش، بهداشت، و خدمات انسانی).

این نگاه خاص ارزشیابی، با وجود معناهای متفاوت، اغلب اشاره به ارزشیابی مشتری گرا، مشتری محور، یا مبتنی بر نیاز دارد، علاوه بر این، به دلیل اهمیت ارزیابی به عنوان یک خدمت به کاربر که

معمولاً تلخیصی است، تحت تأثیر برخی نظریه پردازان است. این نظریه پردازان طوری از آن یاد می کنند که گویی وظیفه اصلی، ارزیابی است. کاربران به تأمین اهداف طراحان محصولات یا برنامه ها خیلی علاقه مند نیستند بلکه تنها تا حدی علاقه مند به پیشرفت برنامه هستند (یعنی ارزیابی تکوینی) و علاقه عمده آنها تأمین نیازهای خودشان است.

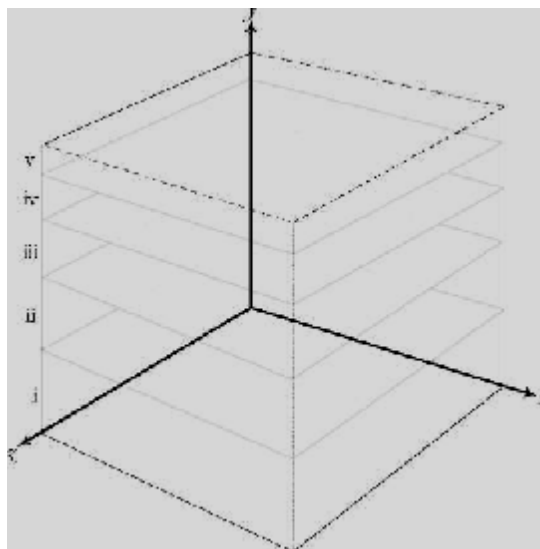
از نظر گرافیکی، عناصر رشته ای مدل ترارشته ای می تواند به صورت صفحه های فضایی در سه بعد آورده شود که هر صفحه نمایانگر یکی از سه مؤلفه یک رشته است. همان طور که در شکل 2-4 نشان داده شده است، صفحه بین محور x و y رشته ها را نشان می دهند، صفحه عمودی بین محورهای y و z حوزه های ارزیابی، و صفحه افقی بین محورهای x و z حوزه های کاربردی را نشان می دهند. هر ارزیابی خاصی می تواند سپس به عنوان یک نقطه یا حجم (مثل مکعب) در این فضای سه بعدی قرار گیرد.



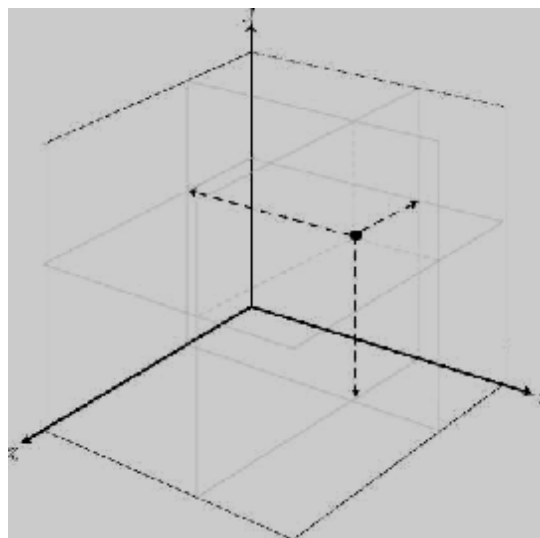
شکل ۲-۳ مدل ترارشته ای

از نظر مفهومی، مدل ترارشته ای، ابزاری مفید برای ارزیابی پژوهشگران و پژوهش آنها و همچنین برای انواع دیگر موارد ارزیابی است. برای نمونه در ارزیابی پژوهشی مربوط به زیست شناس مولکولی، هر یک از مؤلفه های مدل ترارشته ای می توانند به طور فضایی ترسیم شوند. همان طور که در شکل 2-5 نشان داده شده است، در جایی به عنوان رشته علوم طبیعی طبقه بندی شده است (x,y) ، به عنوان محصول

ارزیابی می‌شود (y,z) و در حوزه سلامت کاربرد دارد (x,z) ، این امر به معنای رویه دقیق و پیشرفته‌ای مثل نقشه‌برداری با استفاده از سیستم‌های سه بعدی نیست، بلکه به عنوان ابزاری عملی برای نمایش مصور از ماهیت مسأله مورد ارزیابی یا سایر متعلقات آن است.



شکل ۲-۴ نقشه فضایی از یک رشته و حوزه‌های ارزیابی و کاربرد در مدل تراشته‌ای



شکل ۲-۵ مدل تراشته‌ای و خانه ارزیابی

۲-۱۷-۲ مدل RAE

پیش از دهه 1980، بودجه پژوهشی دانشگاه‌ها در انگلستان، توسط کمیته‌هایی با محوریت موضوع پژوهش‌ها، اختصاص داده می‌شد و معیارهای غیرعینی را اساس ارزیابی‌های خود قرار می‌دادند. عدم شفافیت این فرآیند همراه با مسائلی درباره سهم پژوهش دانشگاهی در بودجه ملی، منجر به طراحی سیستمی شفاف‌تر و عینی‌تر برای ارزیابی پژوهش شد. طرح ارزیابی پژوهش (آ.ای.ای.)²⁰ اولین بار در سال 1986 در قالب پروژه آزمایشی کوچکی و در سال 1989 سیستم کامل‌تر و شفاف‌تر آن اجرا شد و همین سیستم نیز در سال 1992 تکرار شد. در سال 1996 محدوده بسیار بزرگتری شامل 50000 از کارکنان دانشگاهی اجرا شد که از سوی دانشگاه‌های خود به عنوان فعالان پژوهشی کاندید شده بودند و توسط 69 کمیته با حوزه‌های متمایز مورد ارزیابی قرار گرفتند. در این سال، این طرح از فرآیند اطمینان از کیفیت به سوی رقابت برای کسب بودجه گذر کرد.

یکی از اصول مهم این طرح، شفافیت آن است و داده‌هایی که پس از هر ارزیابی در دسترس عموم قرار می‌گیرند. در انگلستان تکیه صرف بر سنج‌های کمی مورد پذیرش نبود و دلیل شکل گرفتن این طرح نیز همین مسأله بود. در نتیجه این طرح، پشتیبانی از توسعه شاخص‌هایی بود که سه حوزه گسترده هزینه، زیرساختها و کیفیت پژوهش را در بر می‌گرفتند و از طرف دیگر فرآیندی بود که طی آن نتایج حاصل از شاخص‌ها توسط گروه کوچکی از کارشناسان تفسیر می‌شد. (Butler and McAllister 2007)

۲-۱۷-۳ مدل ATP

در ایالات متحده، بسیاری از آژانس‌های سرمایه‌گذاری و سازمان‌های پژوهشی با اجرای برنامه‌های ارزیابی در چالش بودند. تلاش‌هایی هم که گاهی برای ارزیابی انجام می‌شد، صرفاً به دلیل اجبار نیروهای مثل قانون دولت اجرا می‌شدند. این مسائل سبب ایجاد یکی از برجسته‌ترین طرح‌های عمومی

²⁰ Research Assessment Exercise (RAE)

خصوصی با نام طرح فناوری پیشرفته (آ.ت.پ.) شد. طراحی و ایجاد این طرح، آزمایشگاهی را برای یادگیری بیشتر در مورد ارزیابی بخصوص در حوزه علم و فناوری شد.

مأموریت این طرح:

- افزودن به پایه دانش علمی و فنی کشور
 - تقویت توسعه فناوری پیشرفته و تجاری نمودن آن توسط شرکت‌های آمریکایی
 - ارتقاء فعالیتهای همکارانه تحقیق و توسعه
 - اصلاح فرآیندهای تولیدی
 - اطمینان از همکاری مناسب بین تجارت‌های کوچک
 - افزایش رقابت بین شرکتهای آمریکایی
 - تولید گسترده بر اساس سودآوری
- جنبه‌ها و سازوکارهای اجرایی:
- توافقات همکاری بین صنایع برای انجام پژوهشهای با هزینه مشترک
 - تمرکز بر پژوهشهای با ریسک بالا برای توسعه فناوریهای قدرتمند
 - انتخاب از بین پروژههای رقیب با استفاده از همترازخوانی و سایر معیارهای منتشر شده
 - فراهم کردن امکانات برای همه پروژههای سرمایه‌گذاری شده
- لازم است که همه پروژهها اهدافی با تعاریف مناسب و مسیرهای مشخص برای نحوه تأثیرگذاری بر

فناوری و اقتصاد اتخاذ کنند. (Ruegg, Feller et al. 2003)

۲-۱۷-۴- بررسی انواع مدل‌های جهانی و مقایسه آنها با یکدیگر

کورین (2007) مدل‌های ارزیابی در 16 کشور مختلف را بررسی کرد، در این مطالعه به منابع مورد

استفاده، سنجه‌ها، قابلیت پایایی و روایی این سنجه‌ها، زمان‌بندی و توالی ابعاد اجرایی آنها و رهیافت‌های

تحلیلی مورد کاربرد در تعیین شایستگی های مدل های ارزیابی پژوهش در سطح ملی پرداخته شده است. طرح این مطالعه، بر اساس روش شناسی های گوناگون در حوزه مقایسه پژوهشها و ارزیابی است. ارزیابی تطبیقی (مقایسه ای)، اشاره به پژوهشی دارد که در آن یک ارزیابی و یافته های آن در چارچوبی تطبیقی تنظیم می شوند. استفاده از روش تطبیقی می تواند در چند روش مشخص شود. برای مثال، به عنوان ابزاری برای تحلیل شباهت ها و تفاوت های سیستم ها به منظور درک، توضیح، یا تفسیر سیستم ها یا پدیده های متفاوت به کار گرفته شوند. اما، اصول روش شناسی که بر اساس آن فرآیندهای ارزیابی تطبیقی باید پایه ریزی شوند، همیشه واضح نیستند. در هر صورت و با وجود واضح نبودن روش شناسی ها، چهار اصل مشترک برای روش ارزیابی تطبیقی وجود دارند: 1) انتخاب موضوع مورد ارزیابی 2) سطح مقایسه، 3) شفاف سازی ویژگی های مورد کاربرد در مقایسه، و 4) چارچوب تحلیلی منطقی.

در مطالعه کورین، موضوعات مورد ارزیابی (یعنی مدل های ارزیابی پژوهشی در سطح ملی)، با روش نمونه گیری مشخصی از بین 272 کشور، مناطق وابسته، و دیگر موجودیتهای جهانی انتخاب شدند که معیارهای این انتخاب عبارت بودند از: 1) شایستگی اطلاعات با توجه به میزان سرمایه گذاری و ارزیابی پژوهش و 2) قابلیت دستیابی اطلاعات به زبان انگلیسی. این نمونه با بیش از دو سوم قدرت های برتر جهانی، همچنین تعداد زیادی از ابرقدرتهای پژوهشی به لحاظ میزان تولیدات علمی و بودجه دولتی اختصاص داده شده به پژوهش و تحقیق و توسعه همخوانی داشت. همچنین این نمونه که شامل 16 کشور بود، تقریباً 80 درصد کل تولیدات علمی جهان را در فاصله سالهای 1998-2002 در اختیار داشت.

مطالعه حاضر، در رده های توصیفی و سنجشی قرار می گیرد و از نظر زمانی نیز گذشته نگر محسوب می شود. در این مطالعه دو گروه هر یک شامل شش داور مسئولیت رتبه بندی آزادانه و جامع مدل های 16 کشور را متعهد شدند همچنین پانل هایی نیز ایجاد شد.

این بررسی در مرکز ارزیابی²¹ ایالات متحده و در بین چند ایالت این کشور و همین طور در نیوریلند برای به حداقل رساندن سوگیری های فرهنگی انجام گرفت. مرکز ارزیابی یک مرکز تحقیق و توسعه در سطح بین المللی است که مأموریت آن گسترش تئوری، اجرا، و کاربرد ارزیابی است.

۲-۱۷-۴-۱- ابزار

ابزار به کار گرفته شده در این بررسی به این ترتیب بودند:

1. توصیفات جمع آوری شده در مراحل پیش

2. پرسشنامه برای داوران؛

3. برگه های امتیازبندی

4. اسناد و دستورالعمل ها

5. رضایت نامه ها

۲-۱۷-۴-۲- سنجها

در این پژوهش بر اساس پنج ویژگی صحت، اعتبار، کاربرد، هزینه- سودمندی، و اخلاق 25 شاخص تهیه شدند که هر یک توسط داوران و با نمره های 0-10 رتبه بندی و سپس اعتباریابی شدند.

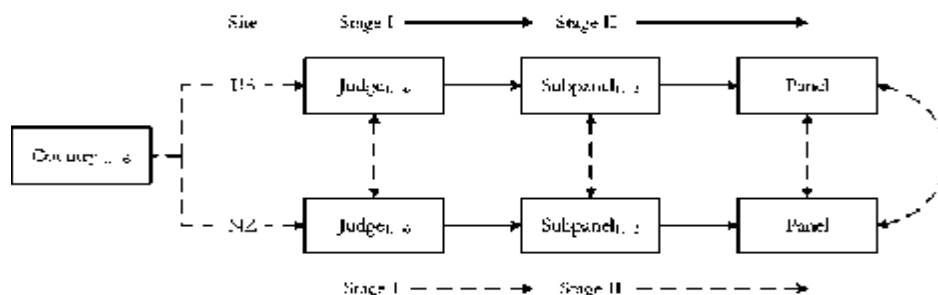
۲-۱۷-۴-۳- روش تحلیل

روش این پژوهش شامل دو مرحله است که هدف از مرحله اول، درجه بندی داوران به روش رتبه بندی آزاد و هدف از مرحله دوم ایجاد یک گروه موافق (پانل) می باشد که بر اساس روش بررسی رضایت مندی آنها ایجاد شده است.

شکل 2-6 روش های به کار گرفته شده بین گروه ها را نشان می دهد. در این شکل خط های تیره

روش کار درونی داوران، پانل های فرعی، و پانل ها، و خط چین ها روش های کار بین آنها را نشان می دهد.

²¹ Evaluation Centre (EC)



شکل ۲-۶ روش تحلیل

برای رتبه بندی پانل ها و داوران از چهار روش تحلیل استفاده شد: نمره گذاری، تهیه پروفایل، ترکیب، و قضاوت انسانی. از نمره گذاری برای ترکیب رتبه های شاخص های پانل و داوران بر اساس پنج شاخص (صحت، اعتبار، کاربرد، هزینه _ سودمندی، و اخلاق)، از ترکیب برای تبدیل عملکردهای ابعاد چندگانه در هر کشور مورد مطالعه به صورت یک نتیجه سنجشی درباره مدل ارزیابی پژوهشی هر کشور، از تهیه پروفایل برای ایجاد پروفایل ابعاد چندگانه در هر کشور به لحاظ پنج شاخص گفته شده، و بالاخره از تحلیل قضاوت انسانی برای بررسی نحوه وزن گذاری مشخصه های هر مدل استفاده شد. (Coryn 2007)

۲-۱۸- چالشهای ارزیابی پژوهش

تحلیل های استنادی و انتشاراتی ابزارهای مشهوری برای ارزشیابی پژوهش شده اند، اما برخی مشکلات روش شناختی هنوز بدون پاسخ مانده اند. گاهی ابهام هایی که بعد از آن پیش می آید، نیز سبب تردیدهایی مبنی بر کاربرد این ابزارها به عنوان استانداردهایی در فرآیندهای سیاست گذاری و تصمیم گیری مدیریتی می شوند. علاوه بر مشکلات مربوط به سازمان ها و واحدهای سازمانی، چالش دیگر تفاوت ارتباطات بین سطوح سازمانی در بین رشته های گوناگون است. این فرض که شخص

می تواند همانندها^{۲۲} را از نظر پارامترهای سازمانی مقایسه کند از دیدگاه مطالعات ارتباطاتی مشکل را است. با توجه به ظهور پیشرفت های علمی جدید (مثل هوش مصنوعی) در ترتیبات بسیار متفاوت سازمانی، لازم است برای انجام مقایسه صحیح، ابتدا واحدهای مورد تحلیل، تعریف شوند. با این همه باز هم، سازمان های علمی به راحتی قابل بررسی و سنجش نیستند.

با توجه به این که سنجش ارتباطات علمی در رابطه با طرح های دانشگاهی، علمی، و صنعتی رسمی است، اما پیشرفت نظام مند این ارتباطات در طی نیمه دوم قرن بیستم، سیستم ارجاع برای ارزیابی پژوهش ها را نیز تغییر داده است. در اروپا، به طور چشم گیری سطح بین المللی به سوی توسعه سیاست های نوآورانه سوق داده شده است تا در جهت بین المللی کردن صنعت و جهانی سازی نمودن نوآوری ها تلاش کند. البته علم از همان آغاز خود، به سمت جهانی شدن گرایش داشته است، ولی نمود فرآیند پژوهش در این پیشرفتهای جهانی، در ارزیابی پژوهش و بررسی های علم سنجی منعکس شده است.

به عبارت دیگر، سیستم های مورد مطالعه پیچیده تر شده اند. برای مثال، این پرسش مطرح می شود که آیا همکاری بین المللی در علم و همکاری های علمی ملی، در طی دهه 1990 به عنوان نیروی فرعی جدید با ویژگی هایی غیر از همکاری های علمی در سطح کشوری ایجاد شده است؟ آیا می توان نیروی فرعی متفاوتی (مثلاً جهانی) فرض نمود و سپس آن را مورد سنجش قرار داد؟ آیا می توان این بعد جدید از بروندادهای علمی را برای طرح هایی که برای مدیریت سازمانی در درون سیستم های ملی توسعه می یابند، در نظر گرفت؟ این پرسش ها، ابهام هایی را در مواجهه با علم، و اقتصاد و بافتهای سیاسی به وجود می آورد. سیستم ها و زیرسیستم های فرعی ایجاد شده در ابعاد متفاوتی با هم ارتباط دارند و ارزیابی بخشی از کد گذاری این ارتباطات شده است (Leydesdorff 2005).

²² like with like

ارزیابی‌هایی که از پژوهش‌ها در هلند و در دانشگاه‌ها و پژوهشگاه‌های تابع یک پروتکل صورت گرفته، نیز وجود مشکلاتی را نشان می‌دهد. در سال 2003، انجمن دانشگاه‌های هلند، سازمان‌های پژوهش‌های علمی، و دانشگاه سلطنتی علوم و هنر هلند برای ارزیابی فعالیت‌های پژوهشی، پروتکل جدیدی را با عنوان «سپ»²³ منتشر کردند. اما سپ نیز روش دقیقی برای ارزیابی پژوهش ارائه نمی‌کند. این امر نشان می‌دهد که ارزیابی فعالیت‌های پژوهشی بسیار پیچیده‌تر از شمارش صرف مقالات در مجلات معتبر است.

موضوع اصلی این است که ارزیابی‌های پژوهشی شکلی از همترازخوانی هستند، یعنی کمیته‌ای از صاحب‌نظران که شایستگی‌های یک طرح پژوهشی را ارزیابی می‌کنند. اما چنین قضاوت‌هایی را سیستم‌هایی از شاخص‌ها پشتیبانی می‌کنند که تا حدی پروتکل‌های ارزیابی را با زبانی ساده توضیح می‌دهند. وقتی که سیستمی برای ارزشیابی پژوهش ایجاد می‌شود، مبنای آن یا شاخص‌های کتابسنجی باشد، یا همترازخوانی، یا ترکیبی از هر دو و یا روش‌های دیگر، دیر یا زود تحت تأثیر ویژگی‌های آن سیستم قرار خواهد گرفت. این امر به خصوص در زمانی صدق می‌کند که عملکرد آنها با توجه به این ویژگی‌ها محور تخصیص بودجه پژوهشی در آینده باشد. طبق نتایج احتمالی جوایز داده شده به تعدادی از پایان‌نامه‌ها و انتشارات علمی، این دسته‌ها بخصوص در مدت 15 سال رشد زیادی داشته‌اند. شمار پایان‌نامه‌ها به 120 درصد و «انتشارات تخصصی» تنها حدود 33 درصد رشد داشته‌اند. سیاست‌گذاری پژوهشی تنها منجر به بهره‌وری بیشتر نمی‌شود، بلکه بر انواع انتشارات بیشتر تأکید دارد.

در کل، اثربخشی روش‌های ارزشیابی پژوهشی، به تأثیر شاخص‌های مبتنی بر محرک‌هایی بستگی دارد که امکان جذب بودجه را دارند. ضروری است که پیش از معرفی این شاخص‌ها در فرآیند ارزیابی،

²³ sep (Standard Evaluation Protocol 2003-2009 For Public Research Organizations)

از تأثیرات مستقیم و غیرمستقیم محرک‌ها آگاه بود. بخصوص هنگامی که این محرک‌ها، برای تخصیص بودجه پیامدهایی در آینده داشته باشند.

طبق آنچه که گفته شد، می‌توان نتیجه گرفت که ارزیابی فعالیتهای پژوهشی در بیشتر زمینه‌های علوم انسانی و اجتماعی نمی‌تواند تنها بر اساس شاخص‌های ساده و یک‌دست کتابسنجی و تنها روی مجلات انگلیسی زبان انجام شوند. روش‌های قابل کاربرد در یک رشته لزوماً برای رشته‌های دیگر معتبر نیستند. (KNAW 2005)

در 20 سال گذشته، تحقیقات و نظریه‌ها درباره پژوهش، همچنین تغییرات نوآوری و فناوری تا حد زیادی تحت تأثیر ویژگی پویا و نظام‌مند این فعالیت‌ها قرار گرفته‌اند. این امر چالش زیادی برای ارزیابی در پی دارد، زیرا به این معنی است که روابط علت و معلولی در این حوزه‌ها بسیار پیچیده‌تر از آن است که ما پیش از این فکر می‌کردیم. مقایسه ساده الف و ب کافی نیست، زیرا تغییر یک امر داخلی است. این امر منطق‌های جدیدی را برای نفوذ ایجاد می‌کند که در آن نه تنها نقص در بازار بلکه نقص در قابلیت‌ها، رفتار، مؤسسه‌ها، و شرایط قانونی به عملکرد سیستم ضربه می‌زنند و نفوذ آنها را توجیه می‌کنند. این نفوذها نیز خود به ارزیابی نیاز دارند. همچنین فرصتی مهم برای پیوند ارزیابی‌ها به طور مستقیم‌تر با سیاست‌گذاری فراهم می‌شود، زیرا سیاست‌گذاری در پایان به مدیریت عملکرد تمام سیستم‌ها در کشورها، مناطق، قاره‌ها، و صنایع منتهی می‌شود. این روند تا سطح نفوذهای فردی برای ارزیابی نیز می‌رسد، زیرا این نفوذها ابزار سیاست‌گذاری هستند، اما ارزیابی باید برای گستره وسیع‌تری که در آن این ابزار با محیط خود، تعامل می‌یابد، در نظر گرفته شود. (Arnold 2004)

3

ارزیابی پژوهش

در حوزه ها و کشورهای جهان

۴-۱- ارزیابی پژوهش در حوزه‌های علمی

۴-۱-۱- ارزیابی در حوزه‌های علوم انسانی و اجتماعی و هنر (HASS)

بر خلاف تصور رایج از دانشگاه به عنوان یک سازمان منزوی، در بسیاری از کشورهای در حال توسعه، امور ارتباطی پژوهشگران دانشگاهی در علوم طبیعی کاربردی به طور موفقیت‌آمیز به کاربران و واسطه‌های اطلاعاتی هدایت شده‌اند تا از این طریق ارتباط مستقیمی با بخش تولید ایجاد شود. چنین ارتباطی در مطالعات کتابسنجی مرسوم مشاهده نمی‌شود و معمولاً برای سیاست‌گذاران علمی ناشناخته‌اند.

ارزیابی پژوهش در علوم انسانی و اجتماعی مشکلات خاصی دارد. پژوهش‌های بزرگ در زبان‌های محلی و در نشریات، کتاب‌ها، یا گزارش‌های محلی که در نمایه‌های بین‌المللی قرار نمی‌گیرند، اغلب مسائل مهم محلی یا ملی را که در سطح کلان علمی چندان اهمیت ندارند، مورد بررسی قرار می‌دهند.

کاربرد وسیع شاخص‌های استنادی به عنوان معیارهای کلیدی در حوزه‌های رسمی ارزیابی، به عنوان سوء تعبیر جدی پارامترهای کیفی در علوم انسانی و اجتماعی شناخته می‌شوند. همچنین، در بسیاری از

کشورهای در حال توسعه، علوم اجتماعی خود را به جای علوم «مفید»، علوم «حیاتی» معرفی می کنند که در این وضعیت، خطر سیاسی برای تقابل نیروهای قدرتمند و فرسوده نامرتبط، یا غیرعملی بیشتر است.

نهادینه کردن علوم اجتماعی در بین کشورها یا منطقه ها برابر نیست: در برخی مکانها، دانشگاه ها سرمدار هستند، در برخی دیگر پژوهش های اجتماعی در بخش غیرانتفاعی انجام شده اند. برای مثال از 106 مرکز پژوهشی علوم اجتماعی که اعضای شورای علوم اجتماعی آمریکا را تشکیل می دهند، 61 مرکز به عنوان سازمان های غیر انتفاعی مستقل بعد از 1971 ایجاد شده اند. این حرکت منجر به افزایش بیشتر در جلب حمایت بین المللی برای پژوهش های علوم اجتماعی شد که بیشتر آن برای اهداف توسعه کاربردی و همچنین در برخی موارد برای رشد نظامی، دیکتاتوری در برخی کشورها بوده است. در این کشورها، دانشمندان علوم اجتماعی از دانشگاه ها به سوی پژوهشگاه های مورد حمایت آژانس های اروپایی و امریکای شمالی به عنوان محل امن پژوهشی فرار کرده اند.

نهادینه سازی نسبی که در پژوهش های علوم اجتماعی در بخش خصوصی امریکای لاتین به وجود آمده، توانایی علوم اجتماعی را در تولید دانش محلی کاهش داده است، زیرا مراکز غیرانتفاعی بسیار کمی آموزش تحصیلی را ارائه می دهند، و همین طور مشکلات مالی و بودجه پژوهش ها برای ایجاد هماهنگی علائق کوتاه مدت مؤسسه های مالی یا مشتریان خارجی وجود دارد.

اصلاحات اجتماعی و اقتصادی در امریکای لاتین از نیمه دهه 1980، باعث ایجاد تقاضا برای پژوهش های کاربردی در علوم اجتماعی در مسائل سیاست اجتماعی مانند روند سیاست گذاری، حکومتی، تمرکززدایی، مقرون به صرفه بودن برنامه ها، و ارزیابی میزان اثرگذاری برنامه ها شده است. اما پژوهشگران حوزه سیاست اجتماعی و حامیان آنها این دغدغه را دارند که مشتری و کاربران آنها چه کسانی هستند: سایر پژوهشگران، سیاست گذاران، روشنفکران، مدیران برنامه ها، رهبران جوامع، گروه های فرعی یا عموم مردم؟

یکی از سخت ترین وضعیت های پژوهشی این است که حامیان پژوهشی به ندرت عموم مردم هستند. یک مدل ساده مبتنی بر تقاضا از یک طرح پژوهشی عملی نخواهد بود زیرا تصمیم برای خرید، سرمایه گذاری، و استفاده از نتایج پژوهشی با دولتمردان، بانک ها یا سرمایه داران، و جامعه مورد نظر است. برای نیل به «تعامل پایدار» به قوانین پیچیده سازمانی نیاز است که با ایجاد و مدیریت در جوامعی که تمایز قطعی بین سیاست و سیاست گذاری وجود ندارد، در چالش باشد. (Davis and Carden 1998)

۴-۱-۲- ارزیابی در حوزه های انسانی و هنرهای خلاق

علیرغم آنچه که توسط نمایه استنادی علوم انسانی و هنر تحت پوشش قرار گرفته شده است، حوزه هنرهای خلاقانه به میزان زیادی در متون کتابسنجی مورد غفلت واقع شده است. با این حال این حوزه با به چالش کشاندن دیدگاه های مرسوم در مورد آنچه که در بعد کمی نه در بعد کتابسنجی، مدرک محسوب می شوند، متحول کننده ترین حوزه است. استرن، مدرک را در هنرهای خلاق در استرالیا تعریف می کند و معتقد است که مدرک می تواند اشکال مختلفی داشته باشد که کلمات نگارشی تنها یکی از آنها محسوب می شود. او اظهار می دارد که انتشار یک چیز به معنای آن است که آن را به اطلاع عموم می رسانیم تا نظرات، دانش، داستانها یا اطلاعات یک نویسنده توسط عموم مشاهده، نقد، تحلیل و ارزیابی شود و در این مسیر، نظرات نویسنده که بر اساس دانش قبلی او است، برای اظهار نظر و آزمون همگان و استفاده آنها قابل دسترس گردد. (Strand 1998, 53-5) این فرآیند می تواند قالبهای متفاوتی داشته باشد، مانند تولید و توزیع کتابها، گزارش ها، مقالات و آنچه که شامل کلمات، طرحها یا تصاویر باشد. اما به طور گسترده تر، انتشار یک اثر می تواند مستلزم شنیدن، دیدن، خواندن، یا تجربه شدن آن در سایر روشها باشد و این امر برای ارائه آثار در روشهای ناپایدار و کمتر ملموس نظیر اجرا روی صحنه یا دیگر محملهای عمومی صدق می کند (1998, 54) از این دیدگاه می توان گفت که عملکرد خلاق نیز از جمله مدارک محسوب می شود.

تنوع گسترده‌ای را که در اشکال بالقوه مدارک وجود دارد، در چند روش قابل بررسی است. برای مثال، در مورد موسیقی، مدرک متنی است که قابل خواندن باشد، حال چه دیداری، چه شنیداری یا به شکل نوشته باشد. در اپرا مدرک چیزی است که در روی صحنه اجرا، روی دیسک فشرده ثبت، و در همان لحظه نیز فیلم آن تهیه می‌شود. استراند ادعا می‌کند که مدرک در معماری، یعنی ساختمان واقعی، نه عکسها یا توصیفات نوشته شده از آن در مقالات مجلات. استرند تأکید می‌کند که با دیدی خاص به ارزشیابی مدارک در هنرهای خلاق، قاعده‌ای ایجاد می‌شود که طبق آن تنها از نظر نوشتاری، مقاله یک نقاش درباره نقاشی خودش یک مدرک محسوب شود. اما خود نقاشی، مدرک نیست. یک مقاله انتقادی در مورد یک تصنیف موسیقی یک مدرک محسوب می‌شود در حالیکه اجرای آن تصنیف مدرک نیست و حتی خود تصنیف هم یک مدرک محسوب نمی‌شود (Donovan 2005).

۴-۱-۳- ارزیابی در حوزه اقتصاد

داناوان^۱ و باتلر^۲ (2005)، طی پیش‌آزمونی قصد داشتند با استفاده از یک روش تطبیقی آگاهانه^۳، قابلیت‌هایی را برای انواع سنجه‌های مورد استفاده در ارزیابی، شناسایی کنند و شواهدی را در این رابطه، بخصوص در حوزه‌هایی که به طور کامل با شاخص‌های استاندارد نشان داده نشده اند، شواهدی را کسب و ارزیابی نمایند.

آنها این هدف را در حوزه اقتصاد دنبال و پژوهش خود را در بین اقتصاددانان اجرا کردند. نتیجه آن بود که اقتصاددانان، گسترش تحلیل کتابشناسی برای منابع غیردست اول را رد کردند. زیرا این تحلیل‌ها و شاخص‌های آنها هیچ اطلاعات مفیدی بخصوص در مورد یکی از مهمترین مؤلفه‌های ارتباطات علمی یعنی مقالات کاربردی ارائه نمی‌کند.

¹Donovan

²Buttler

³peer-informed comparative process

سنجه‌های دیگری که مربوط به جوایز و افتخارات اخذ شده بود، باز هم سنجه‌های معتبری شناخته نشده‌اند اما به عنوان نمایشگر وجهه و شهرت یک واحد پژوهشی پذیرفته شده‌اند.

همچنین، مسأله حل نشده درباره شاخص‌های ضریب تأثیر است که در واقع به جای تعیین میزان جنبه کیفی پژوهش، به تأثیر آن می‌پردازد.

در کل، نتیجه گرفته شد که شرکت کنندگان در این بررسی تمایل داشتند سلسله‌مراتبی را ایجاد کنند و در محدوده شاخص‌ها وزن‌دهی نمایند، تا مشخص شود؛ تا چه حدی بر شاخص‌های مبتنی بر مدارک، نشانگرهای اعتباری یا دیگر قضاوت‌های کیفیت محور تکیه می‌شود. (Donovan 2005)

۴-۱-۴- ارزیابی در حوزه علوم سیاسی

باتلر (2007)، در بررسی خود از نتایج گزارش آر.ای.ای در سال 2001 (یکی از پیشرفته‌ترین سازمان‌های ارزیابی در انگلستان) در علوم سیاسی استفاده کرد تا کاربرد استنادها به عنوان یک سنجه و ارتباط آن با دیگر شاخص‌ها را ارزیابی کند. داده‌های 4400 مورد ارسالی به پانل علوم سیاسی آر.ای.ای. بررسی شدند. 28128 استناد مورد تحلیل، نه تنها مربوط به مقالات مجلات بودند؛ بلکه به همه مدارک شامل کتاب، و فصلی از کتاب مربوط می‌شدند. نتایج نشان داد که استنادها مهمترین ابزار پیش‌گویی بروندهای آر.ای.ای. هستند. همچنین یافته‌ها حاکی از ضرورت توسعه شاخص‌های کمی معتبر بودند که برای ارزیابی کیفیت پژوهش نیاز به ارزیابی‌های داوری کمیته‌های بزرگ را رفع می‌کرد و کتاب‌سنجی‌ها باید بخش مهمی از چنین بدنه‌ای از شاخص‌های کمی را تشکیل دهد. (Butler and

(McAllister 2007)

۴-۱-۵- ارزیابی در حوزه موزه و باستان شناسی

در سال 2004-2005 «موزه باستان شناسی استاوانجر»¹ در نروژ، وضعیت پژوهش خود را طی سال های 1994-2004 مورد ارزیابی قرار داد. هدف اصلی این پژوهش، گسترش پژوهش های آتی بر پایه تجارب ده ساله این موزه بود. در این بررسی، چندان به ارزیابی کیفی پژوهش های افراد و سازمان به لحاظ معیارهای آکادمیک، پرداخته نشده است. برای ارزیابی انتشارات علمی و اسناد و فعالیت های پژوهشی این موزه، خودارزیابی و بازدید از فعالیت ها، گروه ارزیابی ای شکل گرفت کار این گروه بر اساس این واقعیت بود که این موزه، فعالیت پژوهشی دارد و پژوهش های آن باید نسبت به سایر پژوهشگاه ها ویژگی های خاصی داشته باشند. در این ارزیابی، به این موزه به عنوان یک مؤسسه چند رشته ای نگریسته شد.

نتایج این بررسی نشان داد که:

- افراد 20 درصد از وقت کاری خود را صرف پژوهش می کنند
- افراد تمایل دارند در حوزه های مورد علاقه خود کار کنند، نه حوزه های که مؤسسه به آنها تحمیل می کند.
- هدف از مؤسسه باید این باشد که پژوهش باید بخش مهمی از فعالیت های این موزه باشد و برای خارج از مؤسسه نیز نمود تصویر واضحی را ارائه کند (Hedeager 2005)

۴-۲- روشهای ارزیابی پایان نامه ها و رساله ها در دانشگاه ها

دانشگاه جایی نیست که در آن دانشجویان استخرهایی (حوضچه هایی) فرض شوند که باید سطل، سطل اطلاعات در آنها ریخت بلکه آنان چشمه های نهفته ای هستند که باید بجوشند و خلاقانه در بستر

¹ Museum of Archaeology, Stavanger (AmS)

علم جاری شوند. برای این کار دانشجویان باید «روش‌های اندیشیدن» را بیاموزند، نه اینکه فقط «اندیشه»‌های این و آن متفکر را از بر کرده و برگه‌های امتحانی را با آن سیاه کنند. یکی از راه‌های اندیشیدن و آموختن، تأمل عبرت برانگیز در کار و آثار دیگران است.

خاکی (1384) در کتاب خود ارزیابی پایان‌نامه‌ها و رساله‌ها را دارای دو بعد فرهنگی و علمی می‌داند که هر دانشگاه با توجه به ماهیت خود در بعضی بعدها تأکید بیشتر و در بعضی بعدها تأکید کمتری دارد.

۴-۲-۱- بعد فرهنگی

در این بعد ارزیاب باید معیارهای زیر را در ارزیابی خود لحاظ کند.

1. رعایت انصاف و عدالت در بازشناسی نقاط قوت و ضعف اثر مورد بررسی
2. بکارگیری معیارهای علمی در ارزیابی مسائل
3. پرهیز از ورود در حریم شخصی افراد
4. مستندسازی تمامی موارد ارزیابی شده باز ذکر آدرس دقیق در متن مورد بررسی
5. پرهیز از حرف‌های کلی و شفاهی

۴-۲-۲- بعد علمی

در این بعد ارزیاب؛ ساختار، محتوا، شیوه نگارش و روش‌های علمی بکار گرفته شده را مورد توجه قرار می‌دهد و دیدگاه‌های خود را در مورد هریک از این مسائل به‌طور نظام‌مند مطرح می‌سازد.

اجزای این بعد را می‌توان در رابطه با یک پایان‌نامه و رساله تحصیلی به شرحی که در پی می‌آید دسته‌بندی کرد:

۴-۲-۱- ارزیابی ساختاری

در این بخش سازماندهی و چارچوب بکار گرفته شده در تدوین پایان نامه و رساله مورد بررسی و ارزیابی قرار می گیرد و ارزیاب در پی پاسخگویی به سؤالاتی مانند سؤالات زیر بر می آید:

1. آیا اجزای «صفحه عنوان» به طور دقیق تنظیم شده است؟
2. آیا فصل بندی مطالب در کلیت پایان نامه به طور دقیق انجام گرفته است؟
3. آیا عناصر و اجزای هر فصل طبق استانداردهای موجود در هریک از فصل ها گنجانده شده است؟
4. آیا فهرست های مطالب، جداول و اشکال به طور دقیق سازماندهی شده اند؟
5. ...

۴-۲-۲- ارزیابی روشی

در این قسمت؛ ارزیاب پایان نامه را از زاویه روش تحقیق در مفهوم عام آن مورد بررسی قرار می دهد و سؤالاتی مانند سؤالات زیر را پی می گیرد:

1. آیا در موضوع تحقیق و بیان مسأله «اصل تحدید» رعایت شده و با مسأله اصلی رابطه آن حفظ شده است؟
2. آیا در بیان «مسأله اصلی تحقیق» نکات لازم رعایت شده اند؟
3. آیا بیان فرضیه ها از نظر «ساختاری» و «محتوایی» واجد شرایط لازم هستند؟
4. آیا روش نمونه گیری و فرمول انتخاب شده برای تعیین حجم نمونه، مناسب با موضوع و روش تحقیق دارد و به خوبی تشریح شده اند؟
5. ...

۴-۲-۳-۳- ارزیابی محتوایی

در این بخش ارزیاب به این نکته توجه می کند که آیا محتوای پایان نامه مناسبت لازم را با عنوان، موضوع و فرضیه های تحقیق دارد و درجه این مناسبت را در قالب سؤالات زیر دنبال می کند:

1. آیا از آخرین منابع اصلی و داخلی و همچنین دستاوردهای علمی محققان دیگر بهره گیری

شده است؟ (پیشینه تحقیق)

2. آیا رساله های دیگر که به نحوی با موضوع مرتبطند بررسی شده اند؟

3. آیا برای تهیه محتوا، با رعایت اختصار از منابع متنوع استفاده شده است؟

4. آیا در تحلیل داده ها برای قضاوت در مورد فرضیه ها، استنتاجات منطقی و مستدل صورت

پذیرفته است؟

5. آیا چارچوب نظری تحقیق به خوبی تبیین شده است؟

6. ...

۴-۲-۳-۴- ارزیابی نگارشی

در این قسمت ارزیاب، از منظر اصول و قواعد نگارش و ویراستاری به متن مورد نظر توجه می کند و آن را مورد کند و کاو قرار می دهد، سؤالاتی چون:

1. آیا قواعد، و نکات عمومی مربوط به نگارش رعایت شده است؟

2. آیا دسته بندی های لازم که زمینه ساز انتقال سریع پیام پایان نامه و یا رساله باشد صورت گرفته

است؟

3. آیا اصول دستور زبان فارسی مورد توجه قرار داده شده است؟

4. آیا حروف (فونت ها) مناسب برای تایپ انتخاب شده است؟

به طور کلی می توان گفت، سؤالات اساسی ارزیابی را می توان در قالب جدول زیر خلاصه کرد.

عالی ۷۵-۱۰۰	خوب ۵۰-۷۴	متوسط ۲۵-۴۹	ضعیف ۰-۲۴	خصوصیات مورد ارزیابی
				- عنوان روشن و دقیق است؟ - مسأله به خوبی مشخص و بیان شده است؟ - محدودیت‌های مطالعه بیان شده است؟ - راه‌کارها خوب توضیح داده شده و مناسب برای حل مسأله هستند؟ - فرضیات قابل دفاع هستند؟ - طرح تحقیق حقوق انسانی را به هم نمی‌زند؟ - اصطلاحات مهم تعریف شده‌اند؟ - سؤالات مهم به وضوح توضیح داده شده‌اند؟ - فرضیات، عناصر با سؤالات تحقیقی به خوبی بیان شده‌اند؟ - فرضیات، عناصر یا سؤالات تحقیق حاصل مطالعه منابع موجود در موضوع است؟ - ارتباط تحقیق یا مطالعات قبلی کاملاً روشن است؟ - بررسی منابع و مآخذ موضوع و ارجاع به آنها به خوبی انجام شده است؟ - فرآیند کار به تفصیل تشریح شده است؟ - فرآیند کار، مناسب حل مسأله است؟ - جمعیت و نمونه به وضوح مشخص شده‌اند؟ - روش نمونه‌گیری مناسب است؟ - مطالب به روشنی نوشته شده است؟ - متغیرهای ناخواسته کنترل شده‌اند؟ - روش جمع‌آوری داده‌ها توضیح داده شده‌اند؟ - روش جمع‌آوری اطلاعات مناسب حل مسأله است؟ - اعتبار و قابل اعتماد بودن داده‌ها توضیح داده شده است؟ - روش‌های مناسب برای تحلیل داده‌ها انتخاب شده است؟ - جمله‌بندی و نقطه‌گذاری‌ها صحیح است؟ - غلط‌های تایپی و چایی کم است؟ - هجا و دستور زبان صحیح است؟ - مطالب به روشنی نوشته شده است؟ - لحن نوشتار بی‌طرفانه است؟ - درجه‌بندی کلی از خلاقیت و اهمیت مسأله چگونه است؟ - جداول و نمودارها به خوبی مورد استفاده قرار گرفته‌اند؟ - یافته‌های اصلی مورد بحث قرار گرفته و به تحقیقات قبلی مربوط‌اند؟ - اهمیت یافته‌ها تشریح شده است؟ - ارتباط بین تحقیق و یافته‌ها با استدلال قوی و منطقی بیان شده است؟ - نتایج به خوبی بیان شده‌اند؟ - نتایج بر یافته‌ها متکی است؟ - تعمیم به خوبی شکل گرفته است؟

عالی ۷۵-۱۰۰	خوب ۵۰-۷۴	متوسط ۲۵-۴۹	ضعیف ۰-۲۴	خصوصیات مورد ارزیابی
				- محدودیت‌ها و نقاط ضعف تحلیل مورد بحث قرار گرفته است؟ - پیشنهادهایی برای مطالعات بعدی بیان شده است؟

جدول ۳-۱ راهنمای ارزیابی پایان‌نامه (خاکی ۱۳۸۴)

همچنین نادری (۱۳۸۵) پنجاه معیار یا ویژگی را برای ارزیابی پایان‌نامه‌ها و رساله‌ها ذکر نموده است

که به‌طور خلاصه در جدول زیر ارائه می‌شود:

کاملاً نامناسب	ضعیف	متوسط	خوب	عالی	کاربرد ندارد	معیار یا ویژگی
						۱. حوزه مورد مطالعه به دقت و به روشنی شناسایی شده است؟ ۲. عنوان پایان‌نامه با اندازه کافی خلاصه و رسا بیان شده است؟ ۳. میان جنبه‌های «عام» و «خاص» موضوع پایان‌نامه، تناسب معقولی موجود است؟ ۴. ترتیب مؤثری از کلمات در ساخت عنوان بکار رفته است؟ ۵. بیان و شرح مشکل یا مشکلات اساسی و به‌طور کلی اهمیت موضوع تحقیق رسا است؟ ۶. تجزیه و تحلیل واقعیات مسلم، فرض‌ها و نیز بیان ارتباط آنها با موضوع تحقیق روشن است؟ ۷. بنیاد و اساس‌گزینش متغیرها یا عوامل مورد مطالعه منطقی و معقول است؟ ۸. در ارائه روابط و همبستگی‌ها واقعیات و مفاهیم مرتبط با موضوع، نظم و ترتیب منطقی اعمال شده است؟ ۹. اصول واژه‌بندی و نشانه‌گذاری در بیان و شرح موضوع تحقیق رعایت شده است؟ ۱۰. رابطه مستقیمی میان موضوع تحقیق و فرض یا فرض‌های مطرح شده موجود است؟ ۱۱. فرض یا فرض‌ها به روشنی در مفاهیم عملی بیان شده است؟ ۱۲. ویژگی‌های مفروض بر جامعه تحقیق به روشنی و در چارچوب عملی مناسبی بیان و تعریف شده است؟ ۱۳. توضیح روایی و پایایی اندازه‌های مورد مطالعه مناسب است؟ ۱۴. مرزبندی‌های محدودیت‌های دوگانه تحقیق مشخص است؟ ۱۵. محدودیت‌های خارج از کنترل محقق با موضوع تحقیق هماهنگی دارد؟ ۱۶. محدودیت‌های در اختیار محقق مناسب مطرح شده است؟ ۱۷. واژه‌ها و اصطلاحات اساسی و مهم بسادگی و روشنی تعریف شده است؟ ۱۸. مطالب ارائه شده در سابقه و موضوع تحقیق مرتبط با موضوع تحقیق است؟ ۱۹. مطالب ارائه شده در سابقه و موضوع تحقیق دارای سازمان و نظم منطقی است؟ ۲۰. مراجع و مآخذ مطالب آمده در سابقه موضوع تحقیق به روشنی مشخص است؟ ۲۱. مطالب مندرج در سابقه موضوع تحقیق، اساسی و معقول و منطقی برای استنتاج و تعمیم‌هایی درباره فرض‌ها یا سؤال‌های تحقیق بوجود آورده است؟ ۲۲. نقل قول‌های مستقیم و غیر مستقیم آمده در متن گزارش مشخص و روشن است؟

کاربرد ندارد	عالی	خوب	متوسط	ضعیف	کاملاً نامناسب	معیار یا ویژگی
						۲۳. روش تحقیق اختیار شده به خوبی توضیح داده شده است؟ ۲۴. روش تحقیق بکار گرفته شده برای پاسخ یابی با این مسأله و موضوع متناسب است؟ ۲۵. نقاط قوت و ویژگی های برجسته روش تحقیق اختیار شده روشن است؟ ۲۶. توجیه چرایی انتخاب روش تحقیق، مناسب است؟ ۲۷. نارسائی ها و نقاط ضعف روش تحقیق مورد استفاده مشخص نیست؟ ۲۸. خصوصیات ویژه جامعه مادر و نمونه تحقیق به خوبی توضیح داده شده است؟ ۲۹. آزمودنی ها (به لحاظ ارتباط با ماهیت موضوع تحقیق و نیز تعداد) مناسب انتخاب شده اند؟ ۳۰. روش نمونه گیری متناسب با موضوع تحقیق انتخاب شده است؟ ۳۱. روش نمونه گیری به روشنی توضیح داده شده است؟ ۳۲. روش جمع آوری اطلاعات اختیار شده، مناسب است؟ ۳۳. روش جمع آوری اطلاعات به خوبی توضیح داده شده است؟ ۳۴. روش جمع آوری اطلاعات به درستی مورد بهره برداری قرار گرفته است؟ ۳۵. مراحل تهیه و تنظیم وسیله یا وسایل جمع آوری اطلاعات روشن است؟ ۳۶. روایی و پایایی وسیله یا وسایل جمع آوری اطلاعات مشخص شده است؟ ۳۷. روش آماری و تحلیل، مناسب انتخاب شده است؟ ۳۸. روش آماری و تحلیلی بدرستی به کار گرفته شده است؟ ۳۹. فرمول ها سطوح معنی دار بودن آماری به روشنی توضیح داده شده است؟ ۴۰. نتایج به وضوح ارائه شده است؟ ۴۱. جمع بندی نتایج واضح و مشخص است؟ ۴۲. جمع بندی نتایج با دلایل مستند و اساسی پشت بندی شده است؟ ۴۳. توضیح و توجیه تعمیم نتایج به لحاظ مفاهیم آماری، مناسب است؟ توضیح و توجیه تعمیم نتایج به جامعه مادر مناسب است ۴۴. به لحاظ نتایج تحقیق، محقق به آگاهی یا آگاهی های جدیدی دست یافته است؟ ۴۵. افراد دیگری نیز از نتایج این تحقیق سود می برند؟ ۴۶. گزارش از نظر واژه بندی و جمله بندی روان و واضح است؟ ۴۷. گزارش دارای روایی ظاهری و صوری مناسبی است؟ ۴۸. گزارش اصولاً دارای سازمان منطقی و عقلانی است؟ ۴۹. به طور کلی گزارش تحقیق تصویری بی غرضانه و بی طرفانه از نگرشی علمی و عقلانی از موضوع مورد بررسی است؟

جدول ۳-۲ راهنمای ارزیابی پایان نامه نادری (۱۳۸۵)

۴-۳- جمع بندی

یکی از انواع منابعی که از گذشته تاکنون در کتابخانه‌ها، به‌ویژه در کتابخانه‌های دانشگاهی استفاده می‌شده‌اند، پایان‌نامه‌های کارشناسی ارشد و رساله‌های دکتری هستند. این منابع که جزء منابع ردیف اول در کتابخانه‌ها و از مهم‌ترین و اصلی‌ترین منابع پژوهشی محسوب می‌شوند، علاوه بر آنکه اغلب اولین تجربه جدی پژوهشی دانشجویان تحصیلات تکمیلی محسوب می‌شوند، نشان‌دهنده کیفیت و وضعیت آموزش و پژوهش در دانشگاه‌ها نیز به‌شمار می‌روند. این مدارک علاوه بر آنکه منعکس‌کننده سطح پژوهش و حوزه‌های علمی مورد توجه در دانشگاه هستند (Yu Zhenglu & Yuntao 2006)، می‌توانند راهگشا و راهنمای پژوهشگران و دانشجویان برای تکمیل یا آغاز پژوهشهای آینده نیز باشند.

همانگونه که در بالا اشاره هر دانشگاه در ارزیابی پایان‌نامه‌ها و رساله‌ها شد روش‌های مدونی دارد؛ ولی سیستمی به‌طور یکپارچه جهت ارزیابی آنها وجود ندارد. همچنین ارزیابی کمی پایان‌نامه‌ها سبب خواهد شد که معیار مناسبی جهت ارزیابی عملکرد دانشگاه‌ها حاصل شود.

منابع و مأخذ

خاکی، غلامرضا. ۱۳۸۴. روش تحقیق با رویکردی به پایان نامه نویسی. تهران: مرکز تحقیقات علمی کشور
نادری، عزت‌الله، و مریم سیف نراقی. ۱۳۸۵. روش‌های تحقیق و چگونگی ارزشیابی آن در علوم انسانی. تهران: دفتر
تحقیقات و انتشارات بدر.

- Arnold, E. (2004). "Evaluating research and innovation policy: a systems world needs systems evaluations." RESEARCH EVALUATION 13(1): 3-18.
- Avital, M. and F. Collopy (2001). "Assessing Research Performance: Implications for Selection and Motivation." Sprouts: Working Papers on Information Environments, Systems and Organizations 1(1): 40-61.
- Barré, R. (1990). "Strategic Processes and S&T Indicators: Towards a Key Role in R&D Management Systems." The Research System in Transition. Dordrecht: Kluwer Academic Publishers: 227-239.
- BBSRC (2005). Evaluation Strategy for Research Programme and Schemes, Biotechnology and Biological Sciences Research Council.
- Bozeman, B., J. S. Dietz, et al. (2001). "Scientific and technical human capital: an alternative model for research evaluation." International Journal of Technology Management 22(7): 716-740.
- Butler, L. and I. McAllister (2007). "Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science".
- Campbell, D. F. J. (2005). "Evaluation of University Research in Europe: Conceptual Framework and Comparative Typology".
- CIVR (2006). Guidelines for Research Evaluation, Committee for the Evaluation of Research.
- Coryn, C. L. S. (2007). EVALUATION OF RESEARCHERS AND THEIR RESEARCH: TOWARD MAKING THE IMPLICIT EXPLICIT. Michigan, Western Michigan. Doctor of Philosophy: 468.
- Davis, C. H. and F. Carden (1998). "Research effectiveness and R&D evaluation in developing countries".
- Donovan, C. (2005). Quantitative indicators for research assessment – a literature review. Research Evaluation and Policy Project, The Australian National University.
- Gläser, J., T. H. Spurling, et al. (2004). "Intraorganisational evaluation: are there least evaluable units?" RESEARCH EVALUATION 13(1): 19-32.
- Hansson, F. (2005). Research evaluation and organizational learning—a possible co-existence in the universities?
- Hedeager, L. (2005). Research evaluation A report for the years 1994-2004. Härnösand, Museum of Archaeology, Stavanger.
- KNAW (2005). Judging Research on its Merits. An advisory Report by the Council for the Humanities and the Social Sciences Council, Royal Netherlands Academy of Arts and Sciences.
- Leydesdorff, L. (2005). "The evaluation of research and the evolution of science indicators." Current Science 89(9): 1510-1517.
- Leydesdorff, L. (2008). "Caveats for the Use of Citation Indicators in Research and Journal Evaluations." Journal of the American Society for Information Science and Technology 59(2): 278-287.

-
- Nelson, B. (2005). "Research Quality Framework: Assessing the quality and impact of research in Australia." Final Advice on the Preferred Model, DEST.
- Papponetti, V. and M. Bucchi (2007). "Research Evaluation as a Policy Design Tool: Mapping Approaches Across a Set of Case Studies".
- Roessner, D. (2002). "Outcome Measurement in the United States: State of the Art." American Association for the Advancement of Science, Boston, MA.
- Rudd, E. (1988). "The evaluation of the quality of research." Studies in Higher Education 13(1): 45-57.
- Ruegg, R., I. Feller, et al. (2003). A Toolkit for Evaluating Public R&d Investment: Models, Methods, and Findings from ATPs First Decade, US Department of Commerce, Technology Administration, National Institute of Standards and Technology.
- van Raan, A. F. J. (2005). "Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods." Scientometrics 62(1): 133-143.
- van Raan, A. F. J. and T. N. van Leeuwen (2002). "Assessment of the scientific basis of interdisciplinary, applied research Application of bibliometric methods in Nutrition and Food Research." Research Policy 31(4): 611-632.
- Warner, J. (2000). "A critical review of the application of citation studies to the Research Assessment Exercises." Journal of Information Science 26(6): 453.
- Weingart, P. (2003). "Evaluation of research performance: the danger of numbers." Bibliometric analysis in science and research: Applications, Benefits and Limitations: 7-19.
- Wooding, S. and J. Grant (2003). Assessing Research: The Researchers' View, RAND Corporation.