

## اشاعه استانداردهای ارزشیابی، در گزارش پژوهش‌های ارزشیابی آموزشی: کاوشی در پایان‌نامه‌های کارشناسی ارشد

معصومه قائمی پور \*

مرتضی کرمی \*\*

مقصود امین خندقی \*\*\*

### چکیده

ارزشیابی برای سلامت تعلیم و تربیت و برنامه‌های آن، امری مهم و حیاتی است. هدف این پژوهش بررسی میزان اشاعه استانداردهای ارزشیابی در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد بود. پژوهش حاضر از نظر روش، پژوهش ارزشیابی است و با توجه به اینکه گزارش‌های پایانی مدنظر قرار گرفت فراارزشیابی از نوع پایانی بود. جامعه آماری آن شامل همه پایان‌نامه‌های کارشناسی ارشد با موضوع ارزشیابی طی سال‌های ۱۳۸۲-۱۳۹۲ در چهار دانشگاه علامه طباطبائی، شهید بهشتی، تهران و فردوسی مشهد است که به شیوه سرشماری مورد بررسی قرار گرفت. صرفاً پژوهش‌هایی که با روش کمی انجام شده بودند، برای بررسی انتخاب شد. در این پژوهش از فهرست یادآور (چکلیست) فرا-ارزشیابی استافل بیم (۲۰۱۲) استفاده شد. برای مشخص کردن پایایی کدگذار از ضریب همبستگی کاپا استفاده شد که مقدار آن ۰/۹۳ به دست آمد. یافته‌ها نشان داد که همه پایان‌نامه‌ها در اشاعه معیارهای سودمندی، قابلیت کاربرد و درستی در سطح ضعیف و در اشاعه معیارهای دقت و مسئولیت‌پذیری ارزشیابی در سطح متوسط ارزشیابی شدند.

واژگان کلیدی: ارزشیابی، فراارزشیابی، استانداردهای ارزشیابی برنامه

\* کارشناسی ارشد تحقیقات آموزشی دانشگاه فردوسی مشهد

\*\* دانشیار دانشکده علوم تربیتی و روانشناسی دانشگاه فردوسی مشهد (مسئول مکاتبات: m. karami@um. ac. ir)

\*\*\* دانشیار دانشکده علوم تربیتی و روانشناسی دانشگاه فردوسی مشهد

## مقدمه

عده بسیار کمی از مریبان ممکن است اهمیت ارزشیابی را مورد تردید قرار دهند، اما در مقابل، عده زیادی، ارزشیابی را برای سلامت تعلیم و تربیت و برنامه‌های آن، امری مهم و حیاتی می‌دانند (ارنشتاین، ۱۳۸۴). با ارزشیابی برنامه‌های آموزشی، قوت و ضعف برنامه‌ها مشخص می‌شود و به برنامه‌ریزان و مجریان کمک می‌شود تا اطلاعات و شناخت لازم را برای اصلاح برنامه‌های موجود و طراحی برنامه‌های با کیفیت‌تر کسب کنند و بهترین روش‌ها و تمهیدات را برای پیشبرد اهداف برنامه به کار بندند (عابدی سروستانی، ۱۳۹۰، امین خندقی و عجم، ۱۳۹۰). هدف از نتایج ارزشیابی، سنجش اثربخشی برنامه است و به مزایای به‌دست‌آمده توجه می‌کند، پیشنهادهایی برای بهبود می‌دهد و احتمالاً جهتی برای فعالیت‌های آینده فراهم می‌کند. اهمیت نتایج ارزشیابی از آن روست که با بحث و مذاکره می‌تواند بر تصمیم‌گیری و شکل‌گیری سیاست‌های مختلف هر سازمانی مؤثر باشد (سونپال والیاس<sup>۱</sup>، ۲۰۰۹؛ والوویرتا<sup>۲</sup>، ۲۰۰۲). البته باید توجه داشت که بدون عنایت به روش‌های علمی ارزشیابی، هر ارزشیابی منتج به هدف‌های مطلوب نخواهد شد و به‌طور قطع، نیروهای مادی و انسانی را به‌طور نسبی به هدر خواهد داد. ارزشیابی ممکن است به جهت تمرکز ناکافی، اشتباهات فنی، سوء استفاده از قدرت، اجرای نامناسب، گزارش با تأخیر، یافته‌های سوگیرانه، یافته‌های مبهم، نتیجه‌گیری‌های ناموجه، تفسیرهای ناکافی یا اشتباه برای کاربران و پیشنهادهای ناموجه، کم اعتبار شود (استافل بیم<sup>۳</sup>، ۲۰۰۱).

شرایط ارزشیابی به‌طور اجتناب‌ناپذیری پیچیده و مختلف است، بنابراین ارزیاب‌ها و ذی‌نفعان به چهارچوب تفسیری برای روشن شدن مسائل و راه‌حل‌های عملی نیاز دارند (اسچویگر<sup>۴</sup>، ۲۰۰۷). کسی که به کار ارزشیابی می‌پردازد نیازمند آن است تا درباره اینکه از کدام منابع و با چه شیوه‌ای ملاک‌های ارزش‌گذاری را انتخاب کند، چگونه اطلاعات لازم را جمع‌آوری و داده‌ها را تحلیل کند و چگونه و به چه کسانی یافته‌های حاصل از ارزشیابی خود را عرضه نماید تا در جهت اعتلای وضع، مورد استفاده قرار گیرند، تصمیم بگیرد (موسی پور، ۱۳۸۱). هرگام از فرایند ارزشیابی (از نظر جزئیات ارجاع‌دهی، جمع‌آوری داده‌ها، پیش‌نویس گزارش و نتایج) به‌طور بالقوه

---

۱. Sonpal-Valias

۲. Valovirta

۳. Stufflebeam

۴. Schweigert

می‌تواند طعمه رفتارهای نامناسب باشد (دساتلز و جاکوب<sup>۱</sup>، ۲۰۱۲). ارزیاب برای جلوگیری از نوع مشکلات ارزشیابی که می‌تواند ناشی از عدم کاربست استانداردهای ارزشیابی باشد، باید علاوه بر شناخت برنامه مورد نظر از یک الگو و چهارچوب خاصی، برگرفته از استانداردها استفاده نماید. استانداردهای ارزشیابی برنامه، در ارزشیابی از ارزشیابی‌ها، یعنی فرارزشیابی‌ها<sup>۲</sup>، قابل کاربرد هستند. فرارزشیابی ابزاری قوی برای افزایش اثربخشی و کیفیت کار ارزشیابی است و با استفاده از فرارزشیابی اطمینان حاصل می‌شود که ارزشیابی‌ها، یافته‌ها و نتیجه‌گیری‌های درست ارائه می‌دهند (کوکسی و کاراسلی، ۲۰۰۵، ۲۰۰۹؛ هنری، ۲۰۰۱؛ استافل بیم، ۲۰۰۱)، برای کنترل سوگیری مهم هستند (استافل بیم و شینکفیلد؛ ۲۰۰۷، به نقل از مادزیو هاندیلا، گریفیت، فلمینگ و نساموونی<sup>۳</sup>، ۲۰۱۰؛ استافل بیم، ۲۰۰۱) و می‌توانند برای بهبود ارزشیابی‌های بعدی هدایت‌کننده باشند (لینچ و همکاران<sup>۴</sup>، ۲۰۰۳).

با وجود روشن بودن اهمیت و ضرورت فرارزشیابی، متأسفانه مطالعات خیلی کمی از این‌گونه ارزشیابی‌ها گزارش شده است (هانسن، لارنز و دانت<sup>۵</sup>، ۲۰۰۸). فیتزپاتریک و همکاران<sup>۶</sup> (۲۰۰۴)، کمبود فرارزشیابی‌ها را متذکر شده‌اند و بر افزایش استفاده از آنها برای بهبود اجرا و عمل ارزشیابی توصیه کرده‌اند (به نقل از مادزیو هاندیلا و همکاران، ۲۰۱۰). بر اساس کاوش‌های انجام شده توسط محقق، در پیشینه پژوهشی داخل کشور، با وجود انجام پژوهش‌های ارزشیابی متعدد بر اساس رویکردهای پژوهشی متنوعی مانند بازرگان، فتح آبادی و عین الهی (۱۳۷۹)، امین خندقی و باغانی (۱۳۸۹)، شجاع (۱۳۹۰) و محبی امین (۱۳۸۷)، تاکنون پژوهشی که به بررسی میزان اشاعه استانداردهای ارزشیابی در پژوهش‌های ارزشیابی آموزشی بپردازد یافت نشد. با توجه به اینکه ارزشیابی، بیشتر از هر شاخه حرفه‌ای دیگر به مقیاسی برای قضاوت ارزش دست‌آوردهایش نیاز دارد (بیول<sup>۷</sup>، ۲۰۰۰) و همچنین با توجه به اهمیت پژوهش‌های ارزشیابی آموزشی و نتایج آنها در تصمیم‌گیری‌ها، در

---

۱. Desautels & Jacob

۲. meta-evaluation

۳. Madzivhandila, Griffith, Fleming & Nesamvuni

۴. Lynch & et. al

۵. Hanssen, Lawrenz, & Dunet

۶. Fitzpatrick & et. al

۷. Beywl

این مطالعه سعی شد با استفاده از استانداردهای ارزشیابی برنامه انجمن ارزشیابی آمریکا، نسخه سوم، سال ۲۰۱۱، به فرارزشیابی پایان‌نامه‌های کارشناسی ارشد با موضوع ارزشیابی آموزشی پرداخته شود.

پژوهشگر به دنبال رعایت استانداردهای ارزشیابی در فرایند ارزشیابی نیست و فقط به گزارش‌های پایانی توجه می‌کند؛ لذا مبنای تحلیل، گزارش‌های ارزشیابی است که ارزیاب‌ها ارائه می‌دهند و از آنجا که پژوهشگر فقط به گزارش‌های پایانی توجه می‌کند، فرارزشیابی از نوع پایانی است. با توجه به اینکه ارزیاب‌ها چه در پژوهش ارزشیابی خود استانداردها را رعایت کرده یا نکرده باشند، ممکن است در گزارش خود شواهدی دال بر اینکه از استانداردها استفاده کرده‌اند را ذکر نکرده باشند، بنابراین پژوهشگر به دنبال میزان رعایت استانداردها نیست و به دنبال میزان اشاعه استانداردها در گزارش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد است. لذا با توجه به اهمیت فرارزشیابی پایانی در نمایان ساختن نقاط قوت و ضعف ارزشیابی‌ها، این پژوهش با هدف بررسی میزان اشاعه استانداردهای ارزشیابی در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد به‌منظور پاسخ به سؤالات زیر انجام شده است:

- ۱- تا چه اندازه استانداردهای سودمندی در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد اشاعه یافته است؟
- ۲- تا چه اندازه استانداردهای قابلیت کاربرد در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد اشاعه یافته است؟
- ۳- تا چه اندازه استانداردهای درستی در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد اشاعه یافته است؟
- ۴- تا چه اندازه استانداردهای دقت در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد اشاعه یافته است؟
- ۵- تا چه اندازه استانداردهای مسئولیت‌پذیری ارزشیابی در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد اشاعه یافته است؟

مبانی نظری و پیشینه

فرارزشیابی

مایکل اسکریون<sup>۱</sup> واژه فراارزشیابی را در سال ۱۹۶۹ در گزارش محصولات آموزشی معرفی کرد. اسکریون فراارزشیابی را ارزشیابی، از یک ارزشیابی تعریف کرد. نکته اساسی اسکریون این بود که با انتشار گزارش‌های نادرست و سوگیرانه، ارزیاب‌ها می‌توانند مشتریان را برای خرید محصولات آموزشی بی‌ارزش و نامرغوب گمراه کنند و سپس استفاده از آنها به زیان کودکان و جوانان باشد. بنابراین، او تأکید کرد که ارزشیابی از چنین محصولاتی، خود، باید ارزشیابی شوند (استافل بیم، ۲۰۰۰). به باور استافل بیم (۲۰۰۰)، فراارزشیابی فرایند شرح دادن، کسب کردن و به‌کار گرفتن اطلاعات توصیفی و اطلاعات قضاوتی درباره سودمندی، قابلیت کاربرد، درستی و دقت هر ارزشیابی برای هدایت ارزشیابی و گزارش نقاط قوت و ضعف آن است. ساندرز (۱۹۹۴)، معتقد است فراارزشیابی ممکن است برای اهداف تکوینی یا پایانی برای یک مطالعه به کار گرفته شود. فراارزشیابی تکوینی به ارزیاب‌ها برای طرح، اجرا، بهبود، تفسیر و گزارش مطالعات ارزشیابی‌شان کمک می‌کند و فراارزشیابی پایانی بعد از پایان یک ارزشیابی، به مخاطبان کمک می‌کند تا نقاط قوت و ضعف برنامه را ببینند (کوکسی و کاراسلی، ۲۰۰۵، ۲۰۰۹؛ استافل بیم، ۲۰۰۱). همچنین هانسن و همکاران (۲۰۰۸)، فراارزشیابی تکوینی را فراارزشیابی خط دهنده<sup>۲</sup> می‌نامند، زیرا طراحی شده‌اند تا به ارزیاب‌ها قبل و در طول انجام یک ارزشیابی کمک کنند. همچنین آنها فراارزشیابی پایانی را فراارزشیابی برگشت‌پذیر<sup>۳</sup> می‌نامند زیرا طراحی شده‌اند تا به مخاطبان کمک کنند تا درباره ارزشیابی که به اتمام رسیده است، قضاوت کنند.

استافل بیم (۲۰۰۰)، ده مرحله اصلی برای فراارزشیابی بیان می‌کند که عبارتند از: تعیین و ترتیب تعامل با ذی‌نفعان فراارزشیابی، ایجاد یک تیم فراارزشیابی دارای شرایط، تعریف سؤالات فراارزشیابی، توافق بر استانداردها برای قضاوت سیستم ارزشیابی، چهارچوب قرارداد فراارزشیابی، جمع‌آوری و بررسی اطلاعات موجود مناسب و مربوط، جمع‌آوری اطلاعات جدید در صورت نیاز با استفاده از روش‌های مختلف، تجزیه و تحلیل یافته‌ها (اطلاعات کمی و کیفی) و قضاوت تبعیت ارزشیابی

۱. Michael Scriven

۲. proactive meta-evaluation

۳. retroactive

از استانداردهای ارزشیابی انتخاب شده، آماده کردن و ارائه گزارش‌های لازم و کمک به مشتریان و دیگر ذی‌نفعان در تفسیر و کاربرد یافته‌ها. با توجه به دیدگاه پاتون<sup>۱</sup> (۱۹۹۷)، سؤال‌هایی که باید فرارزشیابی بر آنها متمرکز باشد شامل این موارد است: آیا ارزشیابی به خوبی انجام شده است؟ آیا ارزشیابی استانداردها و اصول ارزشیابی را برآورده کرده است؟. اسکریون (۲۰۰۷)، نشان می‌دهد که استفاده از چک‌لیست یا استانداردهایی مانند استانداردهای ارزشیابی برنامه که توسط کمیته مشترک ارائه شده است، می‌تواند به فرارزشیابی کمک کند (به نقل از مادزیوهاندیلا و همکاران، ۲۰۱۰).

### استانداردهای ارزشیابی برنامه

استانداردهای ارزشیابی برنامه (۲۰۱۱) متشکل از پنج معیار سودمندی (مقبولیت ارزیاب، توجه به ذی‌نفعان، اهداف مذاکره‌ای، ارزش‌های روشن، اطلاعات مرتبط، فرایند و محصولات معنی‌دار، گزارش و ارتباط به موقع و مناسب و نگرانی برای عواقب و تأثیر)، قابلیت کاربرد (مدیریت پروژه، رویه‌های قابل اجرا، مقبولیت بافت و کاربرد منابع)، درستی (جهت دهی کلی و پاسخگو، توافقات رسمی، احترام و حقوق شرکت کنندگان انسانی، صراحت و انصاف، شفافیت و افشا، تعارض در علایق و مسئولیت مالی)، دقت (تصمیمات و نتایج موّجه، اطلاعات معتبر، اطلاعات پایا، برنامه روشن و توصیف موقعیت، مدیریت اطلاعات، تجزیه و تحلیل و طراحی‌های درست، استدلال واضح ارزشیابی و ارتباط و گزارش) و مسئولیت‌پذیری ارزشیابی (مستند سازی ارزشیابی، فرارزشیابی داخلی و فرارزشیابی خارجی) است. سودمندی، یعنی مفید بودن ارزشیابی‌ها برای مشتریان و مخاطبانی که می‌خواهند در جریان ارزشیابی قرار بگیرند. قابلیت کاربرد، یعنی تا چه حد ارزشیابی کارآمد و مقرون به صرفه است، درستی یعنی پایبندی به اصول اصلی که شامل: اخلاق حرفه‌ای، کدهای اخلاقی و موافقت‌نامه‌های قراردادی است. معیار دقت، به راست‌گویی، اعتبار<sup>۲</sup>، قابلیت اعتماد بودن یافته‌های ارزشیابی، تفسیر و قضاوت درباره کیفیت اشاره دارد و معیار مسئولیت‌پذیری ارزشیابی، به جمع‌آوری کافی اسناد و مدارک فرایند، موافقت‌نامه‌ها،

۱. Patton

۲. validity

اطلاعات و گزارش‌ها و دیدگاه‌های فراارزشیابی‌های داخلی و خارجی که بر بهبود و مسئولیت‌پذیری فرایندها و محصولات متمرکز شده‌اند، اشاره دارد (یاربروق و همکاران ۲۰۱۱، به نقل از روهه و بودریو، ۲۰۱۳).

### پژوهش‌های انجام شده در داخل و خارج کشور

یکی از پژوهش‌هایی که در زمینه استانداردهای ارزشیابی در داخل کشور انجام گرفته است، پژوهشی است که صادقی (۱۳۸۵) با هدف بررسی وضعیت موجود نظام ارزشیابی پیشرفت تحصیلی پایه پنجم ابتدایی مدارس دخترانه دولتی شهر تهران در چهارچوب استانداردهای ارزشیابی آموزشی، انجام داده است. سؤال اصلی پژوهش، این بود که آیا نظام ارزشیابی پیشرفت تحصیلی پایه پنجم ابتدایی با استانداردهای مربوط فاصله‌ای دارد. یافته‌ها نشان داد که از نظر آموزگاران و مدیران وضعیت موجود نظام ارزشیابی پیشرفت تحصیلی از نظر استانداردهای سودمندی، قابلیت کاربرد، درستی و دقت در حالت نسبتاً مطلوبی قرار دارد؛ بنابراین، وضعیت موجود نظام ارزشیابی پیشرفت تحصیلی در دوره ابتدایی در شهر تهران با وضعیت مطلوب فاصله دارد.

تورس<sup>۱</sup> (۲۰۰۵)، در پژوهشی با عنوان فراارزشیابی از یک ارزشیابی دوره زبان دوم، نقاط قوت و ضعف ارزشیابی دوره مقدماتی را از لحاظ معیارهای سودمندی، قابلیت کاربرد، درستی و دقت، شناسایی کرد و توصیه‌هایی برای بهبود ارزشیابی مطرح کرد که عبارت‌اند از: نیاز به ایجاد روابط بهتر با ذی‌نفعان و درک روشن از دغدغه‌ها و دیدگاه‌های ذی‌نفعان، بستن قراردادهای جامع‌تر و روشن‌سازی مسئولیت‌ها و حقوق مربوط به استفاده از گزارش‌ها و اطلاعات.

لین<sup>۲</sup> (۲۰۰۷)، پژوهشی با هدف فراارزشیابی از ارزشیابی مدرسه در مدارس ابتدایی شهر کائوسیونگ، انجام داد. نتایج نشان داد که استانداردهای سودمندی در سطح عالی بودند. در استانداردهای سودمندی، مقبولیت ارزیاب عالی اما زمینه ارزشیابی ضعیف بود. در استانداردهای قابلیت کاربرد، رویه‌های قابل اجرا عالی اما اثر

۱. Torres

۲. Lin

بخشی هزینه ضعیف بود. در استانداردهای درستی تعامل‌های انسانی عالی اما کاربرد نتایج ضعیف بود. در استانداردهای دقت، تشریح اهداف و رویه‌ها عالی اما سوابق نوشته شده درباره برنامه و نوشتن گزارش‌های ارزشیابی ضعیف بود.

کوکسی و کاراسلی (۲۰۰۹)، پژوهشی با عنوان فرارزشیابی در عمل، انتخاب و کاربرد معیار به‌منظور آگاه کردن ذی‌نفعان از نقاط قوت و ضعف ارزشیابی انجام دادند. نمونه پژوهش شامل نه گزارش، هشت مقاله و یک کتاب بود. پژوهشگران دریافتند که تقریباً تمام فرارزشیابی‌ها از منابع داده‌های متعدد (بررسی گزارش‌ها، بررسی اسناد و مدارک دیگر، مصاحبه و جلسات با تیم ارزشیابی، بررسی و تحلیل مجدد داده‌های اصلی، مصاحبه و جلسات با ذی‌نفعان دیگر، تکرار فعالیت‌های جمع‌آوری داده‌ها) استفاده کردند.

مادزیوهان‌دیلا و همکاران (۲۰۱۰)، پژوهشی با عنوان "فرارزشیابی در دولت و مؤسسات دولتی: مطالعه‌ای موردی از مرکز استرالیایی برای تحقیقات کشاورزی بین‌المللی (ACIAR)"، انجام دادند. فرارزشیابی از ۶۳ اثر گزارش‌های ارزشیابی انجام گرفت. یافته‌ها نشان داد که ۱۹ نمونه گزارش، از استانداردهای، شناسایی ذی‌نفعان، رویه‌های قابل اجرا، مقبولیت سیاسی، توافقات رسمی، حقوق شرکت کنندگان انسانی، تعامل‌های انسانی، مسئولیت مالی، تجزیه و تحلیل داده‌های کیفی و استفاده از فرارزشیابی، استفاده نکرده‌اند یا کم استفاده کرده‌اند. همچنین، حداقل هفتاد درصد استانداردها تا حدی مورد مخاطب قرار گرفته‌اند و در استفاده از استانداردهای ارزشیابی برنامه در قضاوت کیفیت ارزشیابی ACIAR موفق بودند. نتایج نشان داد که در یک مدل فرارزشیابی نمی‌توان بر همه استانداردهای ارزشیابی برنامه به‌طور مساوی تأکید کرد. استانداردهای ارزشیابی برنامه می‌توانند در هنگام تدوین هر مدل ارزشیابی اصلی مهم باشند اما باید به با ملاحظه بافت ارزشیابی، به کار برده شوند.

روهه و بودریو (۲۰۱۳)، در پژوهشی با عنوان "استانداردهای ارزشیابی برنامه ۲۰۱۱: چهارچوبی برای کیفیت ارزشیابی‌های برنامه‌های آموزشی پزشکی"، معضلات، تنش‌ها و خطرات ارزشیابی را نشان دادند و برای آن پیشنهادهایی همچون این موارد، ارائه کردند: ۱- مذاکره و شفاف‌سازی درباره نقش ذی‌نفعان و ارزیاب‌ها ۲- استفاده از روش‌های آمیخته برای بحث درباره نقاط قوت و ضعف برنامه، تهیه داده‌های کیفی در صورت نبود داده‌های کمی، و رعایت احتیاط در اعتباریابی ابزارها در شرایط و بافت‌ها، ۳- در ابتدا، بیان روشن منطق نظری برنامه و اهمیت اجتماعی و ابعاد مهم



ارزش و شایستگی برنامه، ۴- احتیاط در انتخاب و استفاده از مدل‌های ارزشیابی برنامه و تناسب آن با اهداف و نیازهای اطلاعاتی.

با بررسی مرور پیشینه تجربی تحقیقات انجام شده در داخل کشور، مشخص شد که فقط یک مطالعه مروری توسط سبزواری (۱۳۸۴) با عنوان "فرارزشیابی و معیارهای آن"، انجام شده است که به معرفی چهارچوب فرارزشیابی، معیارهای فرارزشیابی و در انتها به معرفی استانداردهای ارزشیابی برنامه نسخه دوم پرداخته است. در پژوهش صادقی (۱۳۸۵) از استانداردهای ارزشیابی دانش‌آموزان و از ابزار پرسشنامه محقق ساخته برای نظرسنجی استفاده شده است. در بررسی مرور پیشینه تجربی تحقیقات انجام شده در خارج از کشور، بیشتر تحقیقات فرارزشیابی، از ارزشیابی یک دوره یا مدرسه بود و تقریباً همه آنها در صدد شناسایی نقاط ضعف و قوت ارزشیابی‌ها با استفاده از استانداردهای مربوط بودند و پیشنهادهایی برای بهبود مطرح کرده‌اند. در اغلب آنها از استانداردهای ارزشیابی نسخه دوم استفاده شده است. فقط در پژوهش روهه و بودریو (۲۰۱۳) از استانداردهای ارزشیابی برنامه (۲۰۱۱) استفاده شده است که به بررسی معضلات، تنش‌ها و خطرات ارزشیابی پرداخته و پیشنهادهایی مطرح کرده‌اند. همان‌طور که ملاحظه شد، پژوهشی که مشابه پژوهش حاضر، یافت نشد.

### روش پژوهش، جامعه و نمونه

پژوهش حاضر از نظر روش، پژوهش ارزشیابی است، به این معنی که پژوهشگر به ارزشیابی از ارزشیابی‌ها می‌پردازد. با عنایت به اینکه گزارش‌های ارزشیابی محرمانه تلقی می‌شوند لذا پژوهش‌های ارزشیابی که توسط دانشجویان تحصیلات تکمیلی انجام شده است و گزارش‌های آن در کتابخانه‌های دانشگاه‌ها موجود است به عنوان جامعه آماری برگزیده شد. بر این اساس جامعه آماری، پایان‌نامه‌های کارشناسی ارشد دانشجویان دانشگاه‌های تهران، علامه طباطبائی، شهید بهشتی و فردوسی مشهد، طی ۱۱ سال اخیر بود و پژوهش‌هایی که با روش کمی انجام شده بودند، انتخاب شدند. در بین رشته‌های علوم تربیتی معمولاً دانشجویان رشته‌های برنامه‌ریزی آموزشی، تحقیقات آموزشی، مدیریت آموزشی و برنامه‌ریزی درسی و در بین رشته‌های روان‌شناسی معمولاً دانشجویان رشته‌های سنجش و اندازه‌گیری (روان‌سنجی) به این نوع پژوهش می‌پردازند. در این چهار دانشگاه به دلیل دایر بودن رشته‌های مورد اشاره و

اساتیدی که در زمینه ارزشیابی فعالیت می‌کنند، پژوهش‌های ارزشیابی بیشتر بود و همچنین دسترسی به این دانشگاه‌ها امکان‌پذیرتر بود. همچنین، دلیل انتخاب دانشگاه فردوسی علاوه بر دسترسی آسان به پایان‌نامه‌های این دانشگاه این بود که برای بررسی‌های اولیه و رفع اشکالات احتمالی در هنگام ارزشیابی پایان‌نامه‌ها در شهر تهران و دانشگاه‌های مربوط ابتدا پایان‌نامه‌های دانشگاه فردوسی بررسی شود.

حجم جامعه تحقیق شامل حدود ۹۹ پایان‌نامه بود که در نهایت تعداد ۸۰ پایان‌نامه مورد بررسی قرار گرفت. با توجه به اینکه در زمان حضور پژوهشگر در شهر تهران، دانشگاه شهید بهشتی به دلیل برگزاری المپیاد تعطیل بود، بنابراین محقق به مرکز مدارک علمی ایران مراجعه کرد و در آنجا هشت پایان‌نامه با موضوع مربوط از دانشگاه شهید بهشتی یافت. همچنین در دانشگاه تهران به دلیل محدودیت در استفاده از نسخه‌های چاپی و کامل نبودن نسخه‌های الکترونیکی، محقق توانست تعداد ۱۳ پایان‌نامه را بررسی کند. از ۸۰ پایان‌نامه موجود، همه موارد بررسی شد. به دلیل محدود بودن جامعه، از شیوه سرشماری استفاده شد. به‌طور کلی، ۶۶/۲۵ درصد از پایان‌نامه‌ها مربوط به دانشگاه علامه طباطبایی، ۱۶/۲۵ درصد از پایان‌نامه‌ها مربوط به دانشگاه تهران، ده درصد از پایان‌نامه‌ها مربوط به دانشگاه شهید بهشتی، ۷/۵ درصد از پایان‌نامه‌ها مربوط به دانشگاه فردوسی مشهد بود. بیشترین تعداد پایان‌نامه‌ها مربوط به سال ۱۳۸۸ (۱۶/۳ درصد) و کمترین تعداد پایان‌نامه مربوط به سال ۱۳۸۲ (۱/۳ درصد) بود. بیشترین تعداد پایان‌نامه‌ها (۳۳/۸ درصد) مربوط به رشته برنامه‌ریزی آموزشی، ۲۳/۸ درصد از پایان‌نامه‌ها مربوط به رشته مدیریت آموزشی و ۱۳/۸ درصد از پایان‌نامه‌ها مربوط به رشته روان‌سنجی، ۷/۵ درصد از پایان‌نامه‌ها مربوط به رشته برنامه‌ریزی درسی، ۳/۸ درصد از پایان‌نامه‌ها مربوط به رشته تحقیقات آموزشی و ۱۷/۵ درصد از پایان‌نامه‌ها مربوط به سایر رشته‌ها (تکنولوژی آموزشی، آموزش ابتدایی، آموزش بزرگسالان، حسابداری، مدیریت دولتی و تربیت بدنی) بود. همچنین ۳۵ درصد پایان‌نامه‌ها از الگوی ارزشیابی درونی، ۱۷/۵ درصد پایان‌نامه‌ها از الگوی ارزشیابی سیپ و ۲۵ درصد پایان‌نامه‌ها از الگوی ارزشیابی پاتریک استفاده کرده بودند.

در این پژوهش، از فهرست یادآور (چک‌لیست) برای جمع‌آوری داده‌ها استفاده شد. این فهرست، معیارهای سودمندی، قابلیت کاربرد، درستی، دقت و مسئولیت‌پذیری ارزشیابی را می‌سنجد. پژوهشگران فهرست یادآور (چک‌لیست) را بر اساس استانداردهای ارزشیابی برنامه (۲۰۱۱) و بر اساس فهرست فرارزشیابی استافل بیم (۲۰۱۲) و همچنین بر اساس بررسی‌های دو فهرست فرارزشیابی کوتاه و بلند استافل بیم (۱۹۹۹) تدوین کردند. با توجه به اینکه استانداردهای ارزشیابی برنامه (۲۰۱۱)، ارزشیابی کارکنان (۱۹۸۸) و ارزشیابی دانش‌آموز (۲۰۰۳)، وجود دارد (استافل بیم، ۲۰۰۴)؛ بنابر این بر حسب موضوع در این پژوهش از استانداردهای ارزشیابی برنامه (۲۰۱۱)، استفاده شد. در فهرست استافل بیم (۲۰۱۲)، برای معیار سودمندی هشت استاندارد، برای قابلیت کاربرد هفت استاندارد، برای درستی ۷ استاندارد، برای دقت هفت استاندارد و برای مسئولیت‌پذیری ارزشیابی سه استاندارد در نظر گرفته شده است و هر استاندارد ده نشانگر دارد. در این پژوهش بر اساس بررسی‌های انجام شده شش نشانگر از ده نشانگر که مصدافی در پایان‌نامه داشتند (گزارش‌های نهایی ارزشیابی)، انتخاب شدند. نشانگرهایی مد نظر قرار گرفت که بتوان برای آنها شواهد عینی پیدا کرد. همچنین استاندارد مسئولیت مالی که یکی از استانداردهای معیار درستی بود چون در پایان‌نامه مصدافی نداشت حذف گردید. برای معیار مسئولیت‌پذیری ارزشیابی هم دو استاندارد فرارزشیابی داخلی و فرارزشیابی خارجی با هم ادغام شد و به صورت یک استاندارد فرارزشیابی در فهرست یادآور در نظر گرفته شد. در مجموع پنج معیار، ۲۸ استاندارد و ۱۶۸ نشانگر مورد بررسی قرار گرفت. نمره‌گذاری بر اساس فهرست فرارزشیابی، فرم کوتاه، استافل بیم (۱۹۹۹) است. قضاوت درباره هر استاندارد به این صورت است، که اگر هر شش نشانگر برای استاندارد در اثر مورد ارزشیابی مشاهده شد، عالی، اگر از شش نشانگر، پنج نشانگر مشاهده شد، خیلی خوب، اگر چهار نشانگر مشاهده شد، خوب، اگر دو یا سه نشانگر مشاهده شد، متوسط و اگر یک نشانگر مشاهده شد یا اصلاً مشاهده نشد، ضعیف می‌باشد. سپس برای هر معیار، قضاوت به شکل درج شده در جدول (۱) بود.

جدول (۱) شیوه نمره‌گذاری معیارهای سودمندی، قابلیت کاربرد، درستی، دقت و

#### مسئولیت‌پذیری ارزشیابی

نمره گذاری ارزشیابی برای سودمندی	<input type="checkbox"/> ۳۰ (٪۹۳) تا ۳۲ عالی
----------------------------------	--

<p>تعداد رتبه عالی (۸-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 4 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خیلی خوب (۸-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 3 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خوب (۸-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 2 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه متوسط (۸-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 1 \times \underline{\hspace{1cm}}</math></p> <p>نمره کل = <math>\underline{\hspace{1cm}}</math></p> <p>۲۹ (۶۸٪) تا ۲۲: خیلی خوب <math>\square</math></p> <p>۲۱ (۵۰٪) تا ۱۶: خوب <math>\square</math></p> <p>۱۵ (۲۵٪) تا ۸: متوسط <math>\square</math></p> <p>۰ (۰٪) تا ۷: ضعیف <math>\square</math></p> <p>..... <math>\times 100 = \dots\dots\dots \div 32 =</math> (نمره کل).....</p>	<p>تعداد رتبه عالی (۸-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 4 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خیلی خوب (۸-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 3 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خوب (۸-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 2 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه متوسط (۸-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 1 \times \underline{\hspace{1cm}}</math></p> <p>نمره کل = <math>\underline{\hspace{1cm}}</math></p> <p>۱۵ (۹۳٪) تا ۱۶: عالی <math>\square</math></p> <p>۱۱ (۶۸٪) تا ۱۴: خیلی خوب <math>\square</math></p> <p>۸ (۵۰٪) تا ۱۰: خوب <math>\square</math></p> <p>۴ (۲۵٪) تا ۷: متوسط <math>\square</math></p> <p>۰ (۰٪) تا ۳: ضعیف <math>\square</math></p> <p>..... <math>\times 100 = \dots\dots\dots \div 16 =</math> (نمره کل).....</p>
<p>تعداد رتبه عالی (۶-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 4 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خیلی خوب (۶-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 3 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خوب (۶-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 2 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه متوسط (۶-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 1 \times \underline{\hspace{1cm}}</math></p> <p>نمره کل = <math>\underline{\hspace{1cm}}</math></p> <p>۲۳ (۹۳٪) تا ۲۴: عالی <math>\square</math></p> <p>۱۷ (۶۸٪) تا ۲۲: خیلی خوب <math>\square</math></p> <p>۱۳ (۵۰٪) تا ۱۶: خوب <math>\square</math></p> <p>۷ (۲۵٪) تا ۱۲: متوسط <math>\square</math></p> <p>۰ (۰٪) تا ۶: ضعیف <math>\square</math></p> <p>..... <math>\times 100 = \dots\dots\dots \div 24 =</math> (نمره کل).....</p>	<p>تعداد رتبه عالی (۶-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 4 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خیلی خوب (۶-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 3 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خوب (۶-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 2 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه متوسط (۶-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 1 \times \underline{\hspace{1cm}}</math></p> <p>نمره کل = <math>\underline{\hspace{1cm}}</math></p> <p>۳۰ (۹۳٪) تا ۳۲: عالی <math>\square</math></p> <p>۲۹ (۶۸٪) تا ۲۲: خیلی خوب <math>\square</math></p> <p>۲۱ (۵۰٪) تا ۱۶: خوب <math>\square</math></p> <p>۱۵ (۲۵٪) تا ۸: متوسط <math>\square</math></p> <p>۰ (۰٪) تا ۷: ضعیف <math>\square</math></p> <p>..... <math>\times 100 = \dots\dots\dots \div 32 =</math> (نمره کل).....</p>
<p>تعداد رتبه عالی (۲-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 4 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خیلی خوب (۲-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 3 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خوب (۲-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 2 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه متوسط (۲-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 1 \times \underline{\hspace{1cm}}</math></p> <p>نمره کل = <math>\underline{\hspace{1cm}}</math></p> <p>۷ (۹۳٪) تا ۸: عالی <math>\square</math></p> <p>۵ (۶۸٪) تا ۶: خیلی خوب <math>\square</math></p> <p>۳ (۵۰٪) تا ۴: خوب <math>\square</math></p> <p>۱ (۲۵٪) تا ۲: متوسط <math>\square</math></p> <p>۰ (۰٪): ضعیف <math>\square</math></p> <p>..... <math>\times 100 = \dots\dots\dots \div 8 =</math> (نمره کل).....</p>	<p>تعداد رتبه عالی (۲-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 4 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خیلی خوب (۲-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 3 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه خوب (۲-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 2 \times \underline{\hspace{1cm}}</math></p> <p>تعداد رتبه متوسط (۲-۰) <math>\square</math> <math>\underline{\hspace{1cm}} = 1 \times \underline{\hspace{1cm}}</math></p> <p>نمره کل = <math>\underline{\hspace{1cm}}</math></p> <p>۷ (۹۳٪) تا ۸: عالی <math>\square</math></p> <p>۵ (۶۸٪) تا ۶: خیلی خوب <math>\square</math></p> <p>۳ (۵۰٪) تا ۴: خوب <math>\square</math></p> <p>۱ (۲۵٪) تا ۲: متوسط <math>\square</math></p> <p>۰ (۰٪): ضعیف <math>\square</math></p> <p>..... <math>\times 100 = \dots\dots\dots \div 8 =</math> (نمره کل).....</p>

**پایایی:** اصطلاح پایایی شامل سه نوع متمایز پایایی (پایداری<sup>۱</sup>، بازنمایی<sup>۲</sup> و دقت<sup>۳</sup>) است. پایداری، میزان عدم تغییر یا تحول یک فرایند در طول زمان است. پایداری در

۱. stability

۲. reproducibility

۳. accuracy

شرایط آزمون مجدد آزمون آشکار می‌شود. به عبارت دیگر، کدگذار عملی را که قبلاً انجام داده است تکرار می‌کند و انتظار دارد که انحراف عمده‌ای بین نتایج آن دو نباشد و نتیجه‌گیری کند که داده‌هایش پایا است (کرپیندورف، ترجمه ناییبی، ۱۳۸۳: ۱۷۷). در این پژوهش برای تعیین پایایی کدگذاری، از نوع اول پایایی، یعنی پایداری استفاده شد، زیرا نمونه‌های پژوهش در شهر تهران بودند و امکان استفاده از فرد دیگری که در زمینه ارزشیابی آموزشی آگاهی داشته باشد و با آموزش، به عنوان کدگذار دوم فعالیت کند، نبود. بنابراین برای ایجاد تسلط در محقق، ابتدا پایان‌نامه‌های دانشگاه فردوسی مشهد بررسی شد و درستی آن مورد بازبینی قرار گرفت. برای نیل به این هدف، از همبستگی کاپایکوهن برای تعیین میزان همخوانی و توافق نظر دو کدگذاری انجام شده توسط محقق استفاده شد. بدین منظور بیست درصد پایان‌نامه‌ها (۱۶ پایان‌نامه) در فاصله زمانی ده روز توسط محقق دوباره بررسی شدند. میزان توافق بین این دو کدگذاری ۰/۹۳۷ تعیین شد.

### شیوه اجرای پژوهش

ابتدا پس از ترجمه و تأیید فهرست یادآور توسط دو متخصص در زمینه ارزشیابی، برای بررسی نهایی سؤالات فهرست و رفع اشکالات، پایان‌نامه‌های دانشگاه فردوسی بررسی شد و پس از رفع اشکالات به دانشگاه‌های واقع در شهر تهران، مراجعه و داده‌ها جمع‌آوری شد. همچنین برای پرهیز از سوگیری و خطاهای احتمالی، برای هر نشانگر، "بخش‌ها"ی احتمالی که ممکن است پژوهشگران درباره نشانگرها، توضیحاتی در پایان‌نامه خویش بیان کرده باشند مشخص شد. همچنین معیاری با عنوان "تا چه حد" برای هر نشانگر مشخص شد به منظور اینکه، تصمیم پژوهشگر مبنی بر اینکه آیا نشانگر مربوط اشاعه شده است، از ذهنی بودن خارج و به عینیت نزدیک شود. به عنوان مثال در معیار سودمندی، استاندارد توجه به ذی‌نفعان، نشانگر اول بیانگر شناسایی روشن مشتریان است که مشخص گردید فصل (۱) (بیان مسئله یا قبل از اهداف) یا در فصل (۳) (جامعه) "بخش‌ها"ی مناسب بررسی این موضوع است و "تا چه حد" قابل قبول اینکه حداقل سه مشتری ارزشیابی را نام برده باشد.

### یافته‌های توصیفی پژوهش

به‌منظور پاسخگویی به سؤال‌های پژوهش، یافته‌های حاصل از تحلیل توصیفی پایان‌نامه‌ها با استفاده از فهرست یادآور فراارزشیابی و به‌صورت شاخص‌های آمار توصیفی فراوانی و درصد فراوانی در قالب جدول یافته‌های توصیفی اشاعه استانداردهای معیار سودمندی، قابلیت کاربرد، درستی، دقت و مسئولیت‌پذیری ارزشیابی ارائه شده است.

سؤال ۱: تا چه اندازه استانداردهای سودمندی در پژوهش‌های ارزشیابی در پایان‌نامه‌های کارشناسی ارشد اشاعه یافته است؟

جدول (۲) یافته‌های توصیفی مربوط به اشاعه استانداردهای معیار سودمندی

معیار	استانداردها		شاخص آماری					
	مقبولیت	توجه به ذی‌نفعان	اهداف مذاکره‌ای	ارزش‌های روشن	اطلاعات مرتبط	فرایند و محصولات معنی‌دار	گزارش و ارتباط به موقع و مناسب	نگرانی برای عواقب و تأثیر
	فراوانی	فراوانی	فراوانی	فراوانی	فراوانی	فراوانی	فراوانی	فراوانی
مقبولیت ارزیاب	۸۰	۸۰	۰	۰	۰	۰	۰	۰
	درصد	۱۰۰	۱۰۰	۰	۰	۰	۰	۰
توجه به ذی‌نفعان	۸۰	۸۰	۰	۰	۰	۰	۰	۰
	درصد	۱۰۰	۱۰۰	۰	۰	۰	۰	۰
اهداف مذاکره‌ای	۸۰	۴۲	۳۸	۰	۰	۰	۰	۰
	درصد	۵۲/۵	۴۷/۵	۰	۰	۰	۰	۰
ارزش‌های روشن	۸۰	۳۸	۴۲	۰	۰	۰	۰	۰
	درصد	۴۷/۵	۵۲/۵	۰	۰	۰	۰	۰
اطلاعات مرتبط	۸۰	۰	۸۰	۰	۰	۰	۰	۰
	درصد	۱۰۰	۰	۰	۰	۰	۰	۰
فرایند و محصولات معنی‌دار	۸۰	۰	۲۱	۴۰	۱۹	۰	۰	۰
	درصد	۰	۲۶/۳	۵۰/۰	۲۳/۸	۰	۰	۰
گزارش و ارتباط به موقع و مناسب	۸۰	۸۰	۰	۰	۰	۰	۰	۰
	درصد	۱۰۰	۰	۰	۰	۰	۰	۰
نگرانی برای عواقب و تأثیر	۸۰	۷۷	۳	۰	۰	۰	۰	۰
	درصد	۹۶/۳	۳/۸	۰	۰	۰	۰	۰

با توجه به جدول (۲) در همهٔ پایان‌نامه‌ها اشاعه استانداردهای مقبولیت ارزیاب، توجه به ذی‌نفعان و گزارش و ارتباط به موقع و مناسب در سطح ضعیف، در ۴۷/۵ درصد اهداف مذاکره، در سطح متوسط و در ۵۲/۵ درصد در سطح ضعیف، در ۵۲/۵ درصد ارزش‌های روشن، در سطح متوسط و در ۴۷/۵ درصد در سطح ضعیف، در همهٔ پایان‌نامه‌ها اطلاعات مرتبط، در سطح متوسط، در ۲۳/۸ درصد فرایند و

محصولات معنی‌دار، در سطح بسیار خوب، در ۵۰ درصد در سطح خوب و در ۲۶/۳ درصد در سطح متوسط، در ۳/۸ درصد نگرانی برای عواقب و تأثیر، در سطح متوسط و در ۹۶/۳ درصد در سطح ضعیف ارزشیابی شدند.

سؤال ۲: تا چه اندازه استانداردهای قابلیت کاربرد در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد اشاعه یافته است؟

جدول (۳) یافته‌های توصیفی مربوط به اشاعه استانداردهای معیار قابلیت کاربرد

معیار	استانداردها	شاخص آماری	عالی	خیلی خوب	خوب	متوسط	ضعیف	کل
قابلیت کاربرد	مدیریت پروژه	فراوانی	۰	۰	۰	۰	۸۰	۸۰
		درصد	۰	۰	۰	۰	۱۰۰	۱۰۰
	رویه‌های قابل اجرا	فراوانی	۰	۰	۰	۶	۷۴	۸۰
		درصد	۰	۰	۰	۷/۵	۹۲/۵	۱۰۰
	مقبولیت بافت	فراوانی	۰	۰	۰	۷۴	۶	۸۰
		درصد	۰	۰	۰	۹۲/۵	۷/۵	۱۰۰
	کاربرد منابع	فراوانی	۰	۰	۰	۵۹	۲۱	۸۰
		درصد	۰	۰	۰	۷۳/۸	۲۶/۳	۱۰۰

با توجه به جدول (۳)، در همه پایان‌نامه‌ها اشاعه استاندارد مدیریت پروژه، در سطح ضعیف، در ۷/۵ درصد رویه‌های قابل اجرا، در سطح متوسط و ۹۲/۵ درصد در سطح ضعیف، در ۹۲/۵ درصد مقبولیت بافت، در سطح متوسط و ۷/۵ درصد در سطح ضعیف، در ۷۳/۸ درصد کاربرد منابع، در سطح متوسط و ۲۶/۳ درصد در سطح ضعیف ارزشیابی شدند.

سؤال ۳: تا چه اندازه استانداردهای درستی در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد اشاعه یافته است؟

جدول (۴) یافته‌های توصیفی مربوط به اشاعه استانداردهای معیار درستی

معیار	استانداردها	شاخص آماری	عالی	خیلی خوب	خوب	متوسط	ضعیف	کل
درستی	جهت‌دهی کلی و پاسخگو	فراوانی	۰	۰	۸۰	۰	۰	۸۰
		درصد	۰	۰	۱۰۰	۰	۰	۱۰۰
	توافقات رسمی	فراوانی	۰	۰	۰	۰	۸۰	۸۰
		درصد	۰	۰	۰	۰	۱۰۰	۱۰۰
	احترام و حقوق	فراوانی	۰	۰	۰	۰	۸۰	۸۰

۱۰۰	۱۰۰	۰	۰	۰	۰	درصد	شرکت‌کنندگان انسانی
۸۰	۰	۱۹	۵۹	۲	۰	فراوانی	صراحت و انصاف
۱۰۰	۰	۲۳/۸	۷۳/۸	۲/۵	۰	درصد	
۸۰	۵	۷۵	۰	۰	۰	فراوانی	شفافیت و افشا
۱۰۰	۶/۳	۹۳/۸	۰	۰	۰	درصد	
۸۰	۷۹	۱	۰	۰	۰	فراوانی	تعارض در علاقه‌ها
۱۰۰	۹۸/۸	۱/۳	۰	۰	۰	درصد	

با توجه به جدول (۴)، در همه پایان‌نامه‌ها اشاعه استانداردهای جهت‌دهی کلی و پاسخگو، در سطح خوب و توافقات رسمی و احترام و حقوق شرکت‌کنندگان انسانی در سطح ضعیف ارزشیابی شدند. در ۲/۵ درصد صراحت و انصاف، در سطح خیلی خوب، ۷۳/۸ درصد در سطح خوب و ۲۳/۸ درصد در سطح متوسط، در ۹۳/۸ درصد شفافیت و افشا، در سطح متوسط و ۶/۳ درصد در سطح ضعیف و در ۱/۳ درصد تعارض در علایق، در سطح متوسط و ۹۸/۸ درصد در سطح ضعیف ارزشیابی شد.

سؤال ۴: تا چه اندازه استانداردهای دقت در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد اشاعه یافته است؟

جدول (۵) یافته‌های توصیفی مربوط به اشاعه استانداردهای معیار دقت

معیار	استانداردها	شاخص آماری	عالی	خیلی خوب	خوب	متوسط	ضعیف	کل
معیار	تصمیمات و نتایج موجه	فراوانی	۲	۳۰	۳۹	۹	۰	۸۰
		درصد	۲/۵	۳۷/۵	۴۸/۸	۱۱/۳	۰	۱۰۰
	اطلاعات معتبر	فراوانی	۳	۲۰	۵۵	۲	۰	۸۰
		درصد	۳/۸	۲۵/۰	۶۸/۸	۲/۵	۰	۱۰۰
	اطلاعات پایا	فراوانی	۰	۰	۰	۷۶	۴	۸۰
		درصد	۰	۰	۰	۹۵/۰	۵/۰	۱۰۰
	برنامه روشن و توصیف موقعیت	فراوانی	۰	۰	۰	۲۹	۵۱	۸۰
		درصد	۰	۰	۰	۳۶/۳	۶۳/۸	۱۰۰
	مدیریت اطلاعات	فراوانی	۰	۰	۱۰	۷۰	۰	۸۰
		درصد	۰	۰	۱۲/۵	۸۷/۵	۰	۱۰۰
	تجزیه و تحلیل و طراحی‌های درست	فراوانی	۰	۰	۰	۷۹	۱	۸۰
		درصد	۰	۰	۰	۹۸/۸	۱/۳	۱۰۰
استدلال واضح (روشن) ارزشیابی	فراوانی	۰	۴۲	۲۸	۱۰	۰	۸۰	
	درصد	۰	۵۲/۵	۳۵/۰	۱۲/۵	۰	۱۰۰	



۸۰	۰	۸۰	۰	۰	۰	فراوانی	ارتباط و گزارش
۱۰۰	۰	۱۰۰	۰	۰	۰	درصد	

با توجه به جدول (۵)، در ۲/۵ درصد پایان‌نامه‌ها اشاعه استاندارد تصمیمات و نتایج موجه، در سطح عالی، ۳۷/۵ درصد در سطح خیلی خوب، ۴۸/۸ درصد در سطح خوب و ۱۱/۳ درصد در سطح متوسط، در ۳/۸ درصد اطلاعات معتبر، در سطح عالی، ۲۵ درصد در سطح خیلی خوب، ۶۸/۸ درصد در سطح خوب و ۲/۵ درصد در سطح متوسط، در ۹۵ درصد اطلاعات پایا، در سطح متوسط و ۵ درصد در سطح ضعیف، در ۳۶/۳ درصد پایان‌نامه‌ها اشاعه استاندارد برنامه روشن و توصیف موقعیت، در سطح متوسط و ۶۳/۸ درصد در سطح ضعیف، در ۱۲/۵ درصد مدیریت اطلاعات، در سطح خوب و ۸۷/۵ درصد در سطح متوسط، در ۹۸/۸ درصد تجزیه و تحلیل و طراحی‌های درست، در سطح متوسط و ۱/۳ درصد در سطح ضعیف، در ۵۲/۵ درصد استدلال واضح ارزشیابی، در سطح خیلی خوب، ۳۵ درصد در سطح خوب و ۱۲/۵ درصد در سطح متوسط ارزشیابی شدند و همه پایان‌نامه‌ها در اشاعه استاندارد ارتباط و گزارش، در سطح متوسط ارزشیابی شدند.

سؤال ۵: تا چه اندازه استانداردهای مسئولیت‌پذیری ارزشیابی در پژوهش‌های ارزشیابی آموزشی در پایان‌نامه‌های کارشناسی ارشد اشاعه یافته است؟

جدول (۶) یافته‌های توصیفی مربوط به اشاعه استانداردهای معیار مسئولیت‌پذیری

#### ارزشیابی

معیار	استانداردها	شاخص آماری	عالی	خیلی خوب	خوب	متوسط	ضعیف	کل
مسئولیت‌پذیری ارزشیابی	مستندسازی	فراوانی	۰	۰	۴۷	۳۳	۰	۸۰
	ارزشیابی	درصد	۰	۰	۵۸/۸	۴۱/۳	۰	۱۰۰
	فراارزشیابی	فراوانی	۰	۰	۰	۰	۸۰	۸۰
		درصد	۰	۰	۰	۰	۱۰۰	۱۰۰

با توجه به جدول (۶)، در ۵۸/۸ درصد پایان‌نامه‌ها استاندارد مستندسازی ارزشیابی، در سطح خوب و ۴۱/۳ درصد در سطح متوسط و همه پایان‌نامه‌ها در اشاعه استاندارد فراارزشیابی، در سطح ضعیف ارزشیابی شدند.

در مجموع، به‌منظور بررسی اینکه تا چه اندازه معیارهای سودمندی، قابلیت کاربرد، درستی، دقت و مسئولیت‌پذیری ارزشیابی در پژوهش‌های ارزشیابی در پایان‌نامه‌های کارشناسی ارشد اشاعه شده است، از شاخص‌های فراوانی و درصد استفاده شد. یافته‌ها در جدول (۷) ارائه شده است.

جدول (۷) یافته‌های توصیفی مربوط به اشاعه معیارهای سودمندی، قابلیت کاربرد، درستی، دقت و مسئولیت‌پذیری ارزشیابی

معیار	شاخص آماری	عالی	خیلی خوب	خوب	متوسط	ضعیف	کل
سودمندی	فراوانی	۰	۰	۰	۰	۸۰	۸۰
	درصد	۰	۰	۰	۰	۱۰۰	۱۰۰
قابلیت کاربرد	فراوانی	۰	۰	۰	۰	۸۰	۸۰
	درصد	۰	۰	۰	۰	۱۰۰	۱۰۰
درستی	فراوانی	۰	۰	۰	۰	۸۰	۸۰
	درصد	۰	۰	۰	۰	۱۰۰	۱۰۰
دقت	فراوانی	۰	۰	۰	۸۰	۰	۸۰
	درصد	۰	۰	۰	۱۰۰	۰	۱۰۰
مسئولیت‌پذیری ارزشیابی	فراوانی	۰	۰	۰	۸۰	۰	۸۰
	درصد	۰	۰	۰	۱۰۰	۰	۱۰۰

با توجه به جدول (۷)، همه پایان‌نامه‌ها در اشاعه معیارهای سودمندی، قابلیت کاربرد و درستی در سطح ضعیف و همه پایان‌نامه‌ها در اشاعه معیار دقت و مسئولیت‌پذیری ارزشیابی در سطح متوسط ارزشیابی شدند.

## بحث و نتیجه‌گیری

### بحث درباره سؤال اول پژوهش

بر اساس یافته‌های به دست آمده نتیجه گرفته شد، اشاعه استانداردهای مقبولیت ارزیاب، توجه به ذی‌نفعان، گزارش و ارتباط به موقع و مناسب و نگرانی برای عواقب و تأثیر در سطح ضعیف، اهداف مذاکره‌ای در سطح نسبتاً ضعیف، اطلاعات مرتبط، در سطح متوسط، ارزش‌های روشن، در سطح نسبتاً متوسط و فرایند و محصولات معنی‌دار در سطح نسبتاً خوب ارزشیابی شدند. به‌طور کلی همه پایان‌نامه‌ها در اشاعه

معیار سودمندی در سطح ضعیف بودند. این نتایج با یافته‌های پژوهش مادزیوهاندیلا و همکاران (۲۰۱۰)، که فرارزشیابی از ۶۳ گزارش ارزشیابی، و نشان دادند که در گزارش ۱۹ نمونه از مطالعات، از استاندارد، شناسایی ذی‌نفعان، استفاده نکرده یا کم استفاده کرده‌اند، مشابه است. ولی با پژوهش لین (۲۰۰۷) که فرارزشیابی از ارزشیابی مدرسه در مدارس ابتدایی شهر کائوسیونگ بود و نتایج پژوهش آنها نشان داد، به‌طور کلی وضعیت اجرای ارزشیابی مدرسه در شهر کائوسیونگ خوب بود و معیار سودمندی در سطح عالی و استاندارد مقبولیت ارزیاب بود، متفاوت است.

در تبیین یافته‌های پژوهش می‌توان این‌گونه بیان نمود که دانشجویان در انجام پژوهش‌های ارزشیابی بیش از اینکه دغدغه توجه به ذی‌نفعان ارزشیابی و علائق و نظرات آنها را داشته باشند در برآوردن استانداردهای پژوهش دانشگاهی اهتمام دارند. بر این اساس در تعیین اهداف و رویه‌های ارزشیابی به جای مذاکره با ذی‌نفعان و ملاحظه نظرات آنها این کار صرفاً مبتنی بر مبانی نظری صورت می‌گیرد و در نتیجه، فرجام پژوهش‌های ارزشیابی بیش از آنکه مبنای تصمیم‌گیری و عمل ذی‌نفعان ارزشیابی باشد، بیشتر، گزارش‌هایی برای دریافت گواهینامه پایان دوره است. به زعم کاپلان و همکاران (۱۹۹۹)، استانداردهای شناسایی ذی‌نفعان و مقبولیت ارزیاب باید در مرحله درگیر کردن ذی‌نفعان در ارزشیابی مورد توجه قرار بگیرد.

با توجه به نتایج، از جمله موارد بسیار مهم که به‌شدت مغفول مانده است، توجه نکردن به نیازها و دغدغه‌های ذی‌نفعان از ارزشیابی برنامه مورد نظر است. ضروری است که در ابتدای پژوهش ارزشیابی، افراد ذی‌نفع مشخص شوند؛ زیرا آنها می‌توانند در تعیین اهداف و سؤال‌های ارزشیابی، انتخاب طرح تحقیق، تفسیر نتایج و چگونگی گزارش یافته‌ها و مشخص کردن افرادی که باید به آنها گزارش شود، کمک کننده باشند. ترسی و تنو (۱۹۹۵)، اعتقاد دارند که برنامه آموزشی زمانی موفق خواهد بود که اطلاعات حاصل از ارزشیابی در مورد برنامه نشان دهد که نیازهای افراد ذی‌نفع رفع شده است (عباسیان، ۱۳۸۵). همچنین، تورس (۲۰۰۵)، نیاز به ایجاد روابط بهتر با ذی‌نفعان و درک روشن از دغدغه‌ها و دیدگاه‌های ذی‌نفعان را به عنوان پیشنهادی برای بهبود ارزشیابی دوره زبان دوم ارائه کرد.

بحث درباره سؤال دوم پژوهش

بر اساس یافته‌های به دست آمده نتیجه گرفته شد، اشاعه استانداردهای مدیریت پروژه و رویه‌های قابل اجرا در سطح ضعیف و مقبولیت بافت و کاربرد منابع در سطح متوسط ارزشیابی شدند. به‌طور کلی همه پایان‌نامه‌ها در اشاعه معیار قابلیت کاربرد در سطح ضعیف بودند. این نتایج با یافته‌های پژوهش مادزیوهان‌دیلا و همکاران (۲۰۱۰) که نشان دادند که در گزارش ۱۹ نمونه از مطالعات، از استانداردهای رویه‌های قابل اجرا، مقبولیت سیاسی استفاده نکرده یا کم استفاده کرده‌اند، مشابه است. ولی با پژوهش لین (۲۰۰۷) که فرارزشیابی از ارزشیابی مدرسه در مدارس ابتدایی شهر کائوسیونگ بود و نتایج پژوهش آنها نشان داد، استاندارد رویه‌های قابل اجرا عالی بود، متفاوت است. به زعم کاپلان و همکاران (۱۹۹۹)، استانداردهای رویه‌های قابل اجرا و مقبولیت سیاسی باید در مرحله تمرکز بر طرح ارزشیابی مورد توجه قرار بگیرند. مطابق با استاندارد رویه‌های قابل اجرا، رویه‌های ارزشیابی باید مطابق با روشی که برنامه اجرا می‌شود، عملی و مناسب باشند که در پایان‌نامه‌های مورد بررسی نادیده گرفته شده است. مطابق با استاندارد مقبولیت بافت، ارزشیابی باید علاقه‌های فرهنگی و سیاسی و نیازهای افراد و گروه‌ها را سازماندهی و کنترل کند. نتایج پژوهش مادزیوهان‌دیلا و همکاران (۲۰۱۰)، نشان داد که هنگام توسعه یک مدل ارزشیابی استانداردهای ارزشیابی برنامه اصلی مهم است اما باید به شیوه‌ای که بر اساس بافت باشد به کار برده شود. بافت و زمینه ارزشیابی از مسائل مهمی است که ارزیابان باید به آن توجه داشته باشند. همچنین مطابق با استاندارد کاربرد منابع، ارزشیابی باید از منابع به‌طور مؤثر و کارآمد استفاده کند.

با توجه به اینکه استاندارد مدیریت پروژه در نسخه سوم، به استانداردهای معیار قابلیت کاربرد اضافه شده است، بنابراین در پژوهش‌هایی که تا قبل از آن انجام شده است، این استاندارد در نظر گرفته نشده است و این سبب شد همه پایان‌نامه‌ها در اشاعه استاندارد مدیریت پروژه، در سطح ضعیف ارزشیابی شوند. مطابق با استاندارد مدیریت پروژه، ارزشیابی باید از راهبردهای مدیریت پروژه مؤثر استفاده کند. روهه و بودریو (۲۰۱۳)، پیشنهاد می‌دهند که در هر برنامه ارزشیابی، یک مدیر پروژه استخدام شود یا مدیریت پروژه، آموزش داده شود.

در تبیین این یافته‌ها می‌توان این‌گونه بیان نمود که دانشجویان در انجام پژوهش‌های ارزشیابی بیش از اینکه به مناسب بودن طرح و روش‌های جمع‌آوری اطلاعات با زمینه، اهداف و فعالیت‌های ارزشیابی توجه کنند بیشتر به این نکته توجه دارند که طرح یا روشی را انتخاب کنند که به آن علاقه‌مندند و یا بنابر مبانی نظری

اهمیت خاصی دارد. به عبارتی دانشجویان معمولاً ابتدا موضوع پایان‌نامه و مدل ارزشیابی را انتخاب می‌کنند سپس دغدغه اجرای آن در سازمان و یا محیط دیگری را دارند. لذا به جای اینکه مدل، متناسب با شرایط سازمان باشد، سازمان را متناسب با مدل انتخاب می‌کنند و این باعث می‌شود بدون در نظر گرفتن زمینه، اهداف و محدودیت‌های منابع مادی و انسانی موقعیت مربوط دست به انتخاب رویه‌ها و ابزار بزنند و روش‌های جمع‌آوری داده‌ها را انتخاب کنند.

### بحث درباره سؤال سوم پژوهش

بر اساس یافته‌های به دست آمده نتیجه گرفته شد، اشاعه استانداردهای توافقات رسمی، احترام و حقوق شرکت‌کنندگان انسانی و تعارض در علاقه‌ها در سطح ضعیف، شفافیت و افشا در سطح متوسط، جهت دهی کلی و پاسخگو در سطح خوب و صراحت و افشا در سطح نسبتاً خوب ارزشیابی شدند. به طور کلی همه پایان‌نامه‌ها در اشاعه معیار درستی در سطح ضعیف بودند. این نتایج با یافته‌های پژوهش مادزیوهاندیلا و همکاران (۲۰۱۰)، که در فرارزشیابی خود از ۶۳ گزارش ارزشیابی، نشان دادند که در ۱۹ نمونه از مطالعات، از استانداردهای توافقات رسمی، حقوق شرکت‌کنندگان انسانی، تعامل‌های انسانی، استفاده نشده یا کم استفاده شده بود، مشابه است. ولی با پژوهش لین (۲۰۰۷) که فرارزشیابی از ارزشیابی در مدارس ابتدایی شهر کائوسیونگ بود و نتایج پژوهش آنها نشان داد که استاندارد تعامل‌های انسانی عالی بود، متفاوت است.

با توجه به نتایج، از جمله موارد بسیار مهم که در پایان‌نامه‌ها نادیده گرفته شده است استاندارد توافقات رسمی است. مطابق با استاندارد توافقات رسمی، موافقت‌نامه‌های ارزشیابی باید برای ایجاد تعهدات روشن و به حساب آوردن نیازها، انتظارات و زمینه‌های فرهنگی مشتریان و ذی‌نفعان مورد مذاکره قرار گیرند. در تبیین این یافته‌ها می‌توان این‌گونه بیان نمود که چون دانشجویان اهداف و سؤال‌های پژوهش خود را با توجه به دغدغه‌ها و نیازهای ذی‌نفعان در نظر نمی‌گیرند، بنابراین رسیدن به توافقات معنی پیدا نمی‌کند. همچنین همه پایان‌نامه‌ها در اشاعه استاندارد احترام و حقوق شرکت‌کنندگان انسانی در سطحی ضعیف بودند. این نتایج تاحدودی با نتایج پژوهش‌های بهمن آبادی (۱۳۹۱) که به بررسی میزان رعایت اخلاق پژوهش در رساله‌های دکتری گروه علوم انسانی دانشگاه فردوسی مشهد و زاهدپاشا، حسن‌جانی روشن و ریاحی (۱۳۸۲) که به بررسی میزان رعایت اخلاق پزشکی در پایان‌نامه‌ها و کارآزمایی‌های بالینی در دانشگاه علوم پزشکی بابل پرداختند، مشابه

می‌باشد. یکی از علت‌های این امر این است که تاکنون الزام و اجباری برای رعایت اصول اخلاقی در پژوهش‌های ارزشیابی مطرح نبوده است. علت دیگر می‌تواند مربوط به عدم آموزش اصول اخلاقی ارزشیابی در دانشگاه‌ها باشد، چرا که کتاب‌های ارزشیابی آموزشی توجه چندانی به مباحث مربوط به اصول اخلاقی در ارزشیابی نداشته‌اند و به آنها نپرداخته‌اند، بنابراین عدم آشنایی دانشجویان با این اصول، باعث عدم اشاعه استاندارد احترام و حقوق شرکت‌کنندگان انسانی از جانب آنان شده است. همچنین یکی از موارد مهم دیگر که در ارزشیابی پایان‌نامه‌ها نادیده گرفته شده است استاندارد تعارض در علاقه است که در همه پایان‌نامه‌ها نادیده گرفته شده است. در تبیین این یافته می‌توان بیان نمود که چون دانشجویان موظف به انجام پایان‌نامه خویش هستند، درگیر کردن افراد دارای شرایط دیگر را مد نظر قرار نداده‌اند. از موارد دیگر که کمتر به آن توجه شده است استاندارد شفافیت و افشا است. مطابق با این استاندارد، ارزشیابی باید توصیف کاملی از یافته‌ها، محدودیت‌ها و نتیجه‌گیری‌ها را برای همه ذی‌نفعان ارائه دهد، مگر اینکه انجام این کار تعهدات عرفی و قانونی را نقض کند که در این پژوهش متوسط ارزشیابی شد و در تبیین این یافته می‌توان بیان نمود که از نشانگرهای این استاندارد، فاش کردن محدودیت‌های ارزشیابی و گزارش تمام یافته‌ها به صورت کتبی است که عمدتاً این دو مورد در پایان‌نامه‌ها قابل مشاهده است. نتایج نشان داد که در استاندارد جهت دهی کلی و پاسخگو وضعیت رعایت نشانگرها در حالت خوب قرار دارد. در تبیین این یافته می‌توان بیان نمود که نشانگرهای این استاندارد عمدتاً به بررسی نقاط قوت و ضعف برنامه مورد ارزشیابی می‌پردازند و از آنجا که یکی از غایات هر پروژه ارزشیابی این موضوع است، لذا در پایان‌نامه‌های مورد بررسی نیز شواهد خوبی در این زمینه مشاهده شد.

### بحث درباره سؤال چهارم پژوهش

بر اساس یافته‌های به‌دست آمده نشان داد که، اشاعه استاندارد برنامه روشن و توصیف موقعیت در سطح نسبتاً ضعیف، استانداردهای اطلاعات پایه، مدیریت اطلاعات، تجزیه و تحلیل و طراحی‌های درست و ارتباط و گزارش در سطح متوسط، استانداردهای تصمیمات و نتایج موجه، اطلاعات معتبر، در سطح نسبتاً خوب و استاندارد استدلال واضح ارزشیابی، در سطح نسبتاً خوب ارزشیابی شد. به‌طور کلی همه پایان‌نامه‌ها در اشاعه معیار دقت در سطح متوسط بود. این نتایج با نتایج پژوهش لین (۲۰۰۷) متفاوت است. پژوهش لین که فراارزشیابی، از ارزشیابی مدرسه در

مدارس ابتدایی شهر کائوسیونگ بود نشان داد، استاندارد تشریح اهداف و رویه‌ها عالی بود.

در تبیین این یافته‌ها می‌توان این‌گونه بیان نمود که دانشجویان با روش‌های جمع‌آوری داده‌ها، تجزیه و تحلیل‌ها، پایایی، اعتبار، منطق نتیجه‌گیری‌ها آشنایی بیشتری دارند و در کتاب‌های روش تحقیق متعدد درباره هر کدام بحث شده است و حتی اساتید در کلاس‌های روش تحقیق به این موارد بیشتر می‌پردازند. همچنین بر اساس استانداردهای رایج در پژوهش‌های دانشگاهی روایی و پایایی ابزارهای اندازه‌گیری از الزامات خاصی برخوردار است، لذا پژوهشگران عموماً تلاش بر آورده نمودن الزامات آن دارند. بر این اساس دانشجویان توضیحات کامل‌تری درباره این موارد در پایان‌نامه خود ذکر می‌کنند. به‌زعم کاپلان و همکاران (۱۹۹۹)، ارزیاب‌ها باید در مرحله توصیف برنامه، به استاندارد تجزیه و تحلیل موقعیت، در مرحله طرح ارزشیابی به استانداردهای اطلاعات معتبر و اطلاعات پایا، در مرحله نتیجه‌گیری‌های موجه به استانداردهای تجزیه و تحلیل داده‌های کمی و تجزیه و تحلیل داده‌های کیفی توجه داشته باشند.

### بحث درباره سؤال پنجم پژوهش

بر اساس یافته‌های به دست آمده نتیجه گرفته شد، اشاعه استاندارد فرارزشیابی در سطح ضعیف و مستندسازی ارزشیابی در سطح نسبتاً خوب ارزشیابی شد. به‌طور کلی، در پایان‌نامه‌ها اشاعه معیار مستندسازی ارزشیابی‌ها در سطح متوسط بود. همچنین این نتایج با یافته‌های پژوهش مادزیوهاندیلا و همکاران (۲۰۱۰)، که فرارزشیابی از ۶۳ گزارش ارزشیابی است، نشان دادند که در ۱۹ نمونه از مطالعات، از استاندارد فرارزشیابی استفاده نکرده‌اند یا کم استفاده کرده‌اند؛ در تبیین این یافته‌ها می‌توان این‌گونه بیان نمود، عدم آشنایی دانشجویان با گام‌های فرارزشیابی و استانداردهای مربوط باعث شده است به این کار نپردازند و همچنین زمان‌بر بودن فرارزشیابی، یکی دیگر از دلایل برای نادیده گرفتن شدن این استاندارد از جانب دانشجویان است. به‌زعم کاپلان و همکاران (۱۹۹۹)، ارزیاب‌ها باید استاندارد فرارزشیابی را در همه مراحل ارزشیابی مورد توجه قرار دهند. مطابق با استاندارد مستندسازی ارزشیابی، ارزشیابی باید به‌طور کامل اهداف مذاکره شده و طرح‌ها، فرایندها، داده‌ها و نتایج اجرا شده آنها را مستندسازی کند که از الزامات هر پژوهش است.

### نتیجه‌گیری کلی

در عرصه یافته‌های این پژوهش آنچه بیش از همه هویدا می‌گردد این است که پژوهشگران با مواردی مانند جمع‌آوری اطلاعات، نمونه‌گیری، ابزارهای اندازه‌گیری، تعیین پایایی و روایی ابزار، تحلیل داده‌های کمی و یا کیفی آشنایی بیشتری دارند و از طرفی در پژوهش‌های دانشگاهی این موارد بیش از همه مورد تأکید اصحاب دانشگاه می‌باشد. لذا سطح اشاعه استانداردهای معیار دقت نسبت به استانداردهای دیگر مطلوب‌تر بود. نکته قابل توجه این است که ارزشیابی، پژوهشی است که از قواعد و اصول خاص خود پیروی می‌کند و همان‌طور که نشان داده شد در این پژوهش این قواعد و اصول در پایان‌نامه‌ها مشاهده نشده است. همچنین نتایج نشان داد در پژوهش‌های ارزیابانه، به قواعد و استانداردهای مختلف ارزشیابی توجه کافی نمی‌شود و این ناظر بر عدم درک کافی، کامل و مناسب ارزیابان نسبت به قواعد و استانداردهای مختلف ارزشیابی است.

همچنین با بررسی جامعه مورد مطالعه، ملاحظه شد علاوه بر رشته‌های برنامه‌ریزی آموزشی، تحقیقات آموزشی، برنامه‌ریزی درسی و روان‌سنجی، دانشجویان رشته‌های دیگر به این نوع از پژوهش پرداخته بودند. با اینکه دانشجویان رشته مدیریت آموزشی، در طی تحصیل خود، واحدی با عنوان ارزشیابی آموزشی نمی‌گذرانند، سهم عمده‌ای از پایان‌نامه‌ها مربوط به آن بود. و شاید این موضوع دلیلی بر نامطلوب بودن اشاعه معیارهای سودمندی، قابلیت کاربرد و درستی ارزشیابی در پایان‌نامه‌ها بود. همچنین در رشته‌هایی که این واحد را می‌گذرانند عنوان کردن مطالبی در زمینه فرارزشیابی و استانداردهای ارزشیابی در کلاس‌های مربوطه محل تردید است.

از آنجا که ارزشیابی آموزشی تکیه‌گاه محکمی برای بهبود کیفیت برنامه و نظام‌های آموزشی می‌باشد و امروزه دغدغه اصلی سکان‌داران این برنامه‌ها بهبود کیفیت است، لذا برای اینکه ارزشیابی آموزشی یافته‌های وثیقی را برای تصمیم‌گیری و عمل فراهم نماید، نیازمند توجه بیشتر به فرارزشیابی و رعایت استانداردهای آن در ارزشیابی برنامه‌ها است. به عبارت دیگر، در ارزشیابی آموزش عالی غایت اصلی کمک به بهبود کیفیت برنامه‌های درسی و آموزشی نظام آموزشی است. چنانچه



استانداردهای ارزشیابی مدنظر قرار گیرد، یافته‌های آن بیش از پیش می‌تواند به بهبود نظام آموزشی کمک کند.

از آنجا که نتایج پژوهش حاکی از این بود که اشاعه استانداردهای ارزشیابی برنامه در پژوهش‌های ارزشیابی وضعیت مطلوبی ندارد، لذا پیشنهادهایی برای بهبود ارائه می‌گردد. از آنجا که نظام تشویق در هر جایگاهی تولیدکننده رفتار مطلوب و فرهنگ‌ساز می‌باشد لذا برای ایجاد و نهادینه‌سازی استانداردهای ارزشیابی، پیشنهاد می‌شود نمره‌ای از کل نمره پایان‌نامه را به رعایت اصول اصلی ارزشیابی و استفاده از استانداردهای مربوط اختصاص داده شود. از آنجایی که به نظر می‌رسد، یکی از علل عدم رعایت استانداردها در پژوهش‌های ارزشیابی، عدم آگاهی دانشجویان نسبت به آن است؛ بنابراین، به اساتید درس ارزشیابی آموزشی توصیه می‌شود که مباحث فرارزشیابی و استانداردهای ارزشیابی (برنامه، دانش‌آموز و کارکنان) را به عنوان یکی از بخش‌های اساسی دوره آموزش خود قرار دهند. همچنین با توجه به اینکه اهداف ارزشیابی باید همسو با دغدغه‌های ذی‌نفعان باشد، توصیه می‌شود پژوهشگران در هنگام انتخاب موضوع پایان‌نامه با موضوع ارزشیابی، ابتدا ذی‌نفعان را شناسایی کنند.

از آنجا که هر پژوهشی ادامه دهنده پژوهش‌های قبلی و مقدمه‌ای برای پژوهش‌های بعدی است، لذا پیشنهادهایی برای پژوهش‌های آینده ارائه می‌گردد: با وجود رشد روز افزون ارزشیابی‌های درونی در آموزش عالی، پیشنهاد می‌شود به فرارزشیابی آنها با توجه به استانداردهای ارزشیابی برنامه در قالب پژوهشی مستقل پرداخته شود. تحلیل کتاب‌های ارزیابی‌های آموزشی به منظور بررسی میزان رسیدگی به استانداردهای ارزشیابی (برنامه، کارکنان و دانش‌آموز) و میزان به روزرسانی اطلاعات و شناسایی مشکلات مربوط به منظور تألیف کتاب‌های جامع‌تر. همچنین بررسی نظام ارزشیابی پیشرفت تحصیلی مقاطع مختلف تحصیلی دانش‌آموزان (ابتدایی و متوسطه) با استفاده از استانداردهای ارزشیابی دانش‌آموز.

## منابع

- ارنشتاین، آلن و هانکینز، فرانسیس (۱۳۸۴). *مبانی، اصول و مسائل برنامه درسی*؛ ترجمه قدسی احقر، تهران: انتشارات دانشگاه آزاد اسلامی واحد علوم و تحقیقات.
- امین خندقی، مقصود و باغانی، مریم (۱۳۸۹). *ارزشیابی فراگیران در محیط آموزش الکترونیکی با مروری بر مدل GPAM\_WATA*. پنجمین کنفرانس ملی و دومین کنفرانس بین‌المللی یادگیری و آموزش الکترونیکی. دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، ایران.
- امین خندقی، مقصود و عجم، علی‌اکبر (۱۳۹۰). *تأملی بر پیوند نظری رویکردهای ارزشیابی آموزشی با الگوهای کیفیت آموزشی و تبیین پیامدهای آن بر حوزه عمل. نامه‌ی آموزش عالی، ۴ (۱۳): ۷۳-۴۷*.
- بازرگان، عباس؛ فتح‌آبادی، جلیل و عین‌اللهی، بهرام (۱۳۷۹). *رویکرد مناسب ارزیابی درونی برای ارتقای مستمر کیفیت گروه‌های آموزشی در دانشگاه‌های علوم پزشکی. مجله روان‌شناسی و علوم تربیتی، ۶۱: ۱-۲۶*.
- بهمن‌آبادی، سمیه (۱۳۹۱). *بررسی میزان رعایت اخلاق پژوهش در رساله‌های دکتری گروه علوم انسانی دانشگاه فردوسی مشهد در طی سال‌های ۱۳۸۶ تا ۱۳۹۰*. پایان‌نامه کارشناسی ارشد. دانشکده علوم تربیتی و روان‌شناسی دانشگاه فردوسی مشهد.
- زاهدپاشا، یداله؛ حسن‌جانی روشن، محمدرضا و ریاحی، حمیدرضا (۱۳۸۲). *میزان رعایت اخلاق پزشکی در پایان‌نامه‌ها و کارآزمایی‌های بالینی در دانشگاه علوم پزشکی بابل. مجله علمی پژوهشی دانشگاه علوم پزشکی بابل، ۵ (۳): ۴۵-۵۱*.
- سبزواری، سکینه (۱۳۸۴). *فراارزشیابی و معیارهای آن، گام‌های توسعه در آموزش پزشکی، ۲ (۲): ۱۱۶-۱۲۴*.
- شجاع، کلثومه (۱۳۹۰). *ارزشیابی اثر بخشی برنامه‌های آموزش ضمن خدمت کارکنان دانشگاه فردوسی مشهد. پایان‌نامه کارشناسی ارشد، دانشکده علوم تربیتی و روانشناسی، دانشگاه فردوسی*.
- صادقی، فرشته (۱۳۸۵). *ارزشیابی نظام ارزشیابی پیشرفت تحصیلی پایه پنجم ابتدایی مدارس دخترانه دولتی شهر تهران در چهارچوب استانداردهای ارزشیابی آموزشی. پایان‌نامه کارشناسی ارشد، دانشکده علوم تربیتی و روانشناسی، دانشگاه تهران*.

- عابدی سروستانی، احمد (۱۳۹۰). چالش‌ها و راهکارهای ارزشیابی اثربخش برنامه‌های اجتماعی. *فصلنامه پژوهش*، ۳ (۲): ۱۳۹-۱۵۵.
  - عباسیان، عبدالحسین (۱۳۸۵). اثر بخشی دوره های آموزشی. *ماهنامه تدبیر*، ۱۷۰: ۵۲-۵۵.
  - کریپندورف، کلاوس (۱۳۹۱). تحلیل محتوا مبانی روش شناسی؛ ترجمه هوشنگ نایی، تهران: نشر نی.
  - محبی امین، سکینه (۱۳۸۷). *ارزیابی آموزش کارکنان دانشگاه فردوسی در مسیر توانمند سازی آنان با الگوی ارزشیابی CIPP*. پایان‌نامه کارشناسی ارشد، دانشکده علوم تربیتی و روانشناسی، دانشگاه فردوسی.
  - موسی‌پور، نعمت الله (۱۳۸۱). مبانی نظری الگوهای ارزشیابی آموزشی. *فصلنامه پژوهش و برنامه ریزی در آموزش عالی*، ۲۳ و ۲۴: ۱۱۹-۱۴۴.
- 
- Beywl, W. (2000). Standards for evaluation: On the way to guiding principles in German evaluation. In C. Russon (Ed.), *the program evaluation standards in international settings* (pp. 60-66). Kalamazoo: Western Michigan University Evaluation Center.
  - Cooksy, L. J. & Caracelli, V. J. (2005). Quality, Context, and Use Issues in Achieving the Goals of Metaevaluation. *American Journal of Evaluation*, 26 (1): 31-42.
  - Cooksy, L. J. & Caracelli, V. J. (2009). Metaevaluation in Practice: Selection and Application of Criteria. *Journal of MultiDisciplinary Evaluation*, 6 (11): 1-15.
  - Desautels, G. & Jacob, S. (2012). The ethical sensitivity of evaluators: Aqualitativestudy using a vignette design. *Evaluation*, 18 (4): 437-450.
  - Hanssen, C. E.; Lawrenz, F. & Dunet, D. O. (2008). Concurrent Meta-Evaluation ACritique. *American Journal of Evaluation*, 29 (4): 572-582.
  - Henry, G. T. (2001). How modern democracies are shaping evaluation and the emerging challenges for evaluation. *The American Journal of Evaluation*. 22 (3): 419-429.
  - Koplan, J. P.; Milstein, R. & Wetterhall, S. (1999). Framework for program evaluation in public health. *MMWR: Recommendations and Reports*. 48: 1-40.
  - Lin, S. H. (2007). *A Study on Metaevaluation of Elementary School Evaluation in Kaohsiung City*. [http://etd.lib.nsysu.edu.tw/ETD-db/ETD-search/view\\_etd?URN=etd-0729107](http://etd.lib.nsysu.edu.tw/ETD-db/ETD-search/view_etd?URN=etd-0729107) 231017.

- Lynch, D. C.; Greer, A. G.; Larson, L. C.; Cummings, D. M.; Harriett, B. S. & Dreyfus, K. S. (2003). *Descriptive Metaevaluation Case Study of an Interdisciplinary Curriculum*. *Evaluation & the health professions*, 26 (4): 447-461.
- Madzivhandila, T.; Griffith, G. R.; Fleming, E. M. & Nesamvuni, A. (2010), *Meta-evaluations in government and government institutions: A case study example from the Australian Centre for International Agricultural Research*. Paper presented at the Conference (54th), February 10-12, Adelaide, Australia.
- Patton, M. Q (1997), *Utilization-focused evaluation* (London: Thousand Oaks).
- Ruhe. V. & Boudreau. J. D. (2013). The 2011 Program Evaluation Standards: a framework for quality in medical education programme evaluations. *Journal of evaluation in clinical practice*, 19 (5): 925-932.
- Sanders, J. R. (1994). *The program evaluation standards: how to assess evaluations of educational programs*: Sage.
- Schweigert, Francis J. (2007). The priority of justice: A framework approach to ethics inprogram evaluation, *Evaluation and Program Planning*, 30: 394-399
- Sonpal-Valias, N. (2009). *MODULE 1 Outcome Evaluation: Definition and Overview*, the Vocational and Rehabilitation Research Institute.
- Stufflebeam. Daniel L. (2000). The methodology of metaevaluation as reflected in metaevaluations by the Western Michigan University Evaluation Center. *Journal of Personnel Evaluation in Education*, 14 (1): 95-125.
- Stufflebeam, D. L. (2001). The metaevaluation imperative. *The American Journal of Evaluation*, 22 (2): 183-209.
- Torres, E. B. (2005). *A metaevaluation of an evaluation of a second language course*, thesis for the degree of Master of Science (M. A), Brigham Young University.
- Valovirta, V. (2002). Evaluation utilization as argumentation. *Evaluation*, 8 (1): 60-80.