

ارزشیابی چندسطحی تلفیقی: تدوین مدل (مطالعه موردی طرح همیار مشاور مدارس)

♦ دکتر سعید اکبری زردخانه^۱ ♦ پگاه گودرزی^۲ ♦ دکتر محمد مصطفوی^۳

چکیده:

هدف پژوهش حاضر تدوین مدل ارزشیابی جامع با قابلیت ارزشیابی برنامه‌های آموزشی با جامعه هدف گسترده و لایه‌های مدیریتی و اجرایی متعدد از جمله برنامه همیار مشاور مدارس است. روش پژوهش از نظر هدف کاربردی، از نظر شیوه گردآوری داده‌ها از نوع توصیفی و از نظر ماهیت داده‌ها کیفی است. نمونه به روش غیر تصادفی هدفمند از اسناد علمی اجرایی برنامه‌های اجتماعی و نیز پیشینه علمی-پژوهشی مرتبط با موضوع پژوهش انتخاب شده است. یافته‌ها نشان دادند که مدل ارزشیابی چندسطحی باید از تلفیق الگوی منطقی به‌منزله الگوی عمومی طرح‌ریزی و ارزشیابی، سیپ از رویکرد مدیریتی و کرک پاتریک از رویکرد مبتنی بر هدف شکل گرفته شود. در مدل مذکور ضمن توجه به بخشهای طراحی تا اجرای برنامه، سطوح مختلف ذی‌نفعان در ارزشیابی مشارکت داده می‌شوند. برنامه همیار مشاور مدارس را با توجه به گستردگی، تنوع اهداف و سطوح افراد چندلایه درگیر در آن می‌توان با این مدل چندسطحی مورد ارزشیابی قرار داد، سپس راهکارهای اصلاحی اولویت‌بندی شده را در اختیار سیاستگذاران برنامه قرار داد تا بتوان میزان تحقق‌پذیری اهداف و اثربخشی برنامه را افزایش داد. به‌کارگیری این مدل برای ارزشیابی برنامه‌هایی وسیع که در آن بررسی سطوح چندگانه مدیریتی و اجرایی برنامه و ایجاد تعامل میان نتایج ارزشیابی آنها دارای اهمیت است، به‌منزله راهکاری مفید پیشنهاد می‌شود.

کلید واژگان: ارزشیابی، چندسطحی تلفیقی، برنامه همیار مشاور، تدوین مدل، طراحی مدل

© تاریخ دریافت: ۹۸/۱/۱۵

© تاریخ پذیرش: ۹۹/۱/۵

۱. نویسنده مسئول: استادیار سنجش و اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه شهید بهشتی. s_akbari@sbu.ac.ir
۲. دانش‌آموخته دوره دکتری روانشناسی سلامت، دانشگاه آزاد اسلامی، واحد کرج. p.goodarzy66@gmail.com
۳. دکتری مشاوره، سازمان آموزش و پرورش استثنایی کشور. mostafavi90@gmail.com

بیان مسأله

طی دو دهه گذشته همزمان با رشد معنادار تعداد برنامه‌های آموزشی پیشگیرانه و اجتماعی سازمانها، در دنیا جنبش عمومی تولید مستندات به‌منزله پشتوانه و شواهد علمی برای توجیه کاربرد این برنامه‌ها شکل گرفته است. از نشانه‌های آمادگی سازمانها برای اتخاذ رویکرد مبتنی بر شواهد، تعهد به ارزشیابی^۱ و سنجش نتایج برنامه‌هاست (چیلنسکی، اولسون، شولته، پرکینز و اسپات^۲، ۲۰۱۵). در این راستا در عصر حاضر برنامه‌ها باید با دید انتقادی بررسی و ارزشیابی شده و آنها که ارزش و کارایی بیشتری دارند ادامه یابند (فارسی و شریف، ۲۰۱۴). ارزشیابی به فرایند نظام‌مند گردآوری و تفسیر اطلاعات برنامه به‌منظور دستیابی و تحقق بخشی به اهداف آن، اطلاق می‌شود (نیازی، عرب، فرزبان‌پور و محمودی‌مجدآبادی، ۲۰۱۶). امروزه ارزشیابی تقریباً مرکز ثقل همه سیاستها و راهبردهای کیفیت آموزشی محسوب می‌شود و توجه به کیفیت آموزشی یکی از عوامل بسیار مهم و تأثیرگذار بر سرنوشت جامعه است و هر ارگان یا سازمانی برای ارتقای سطح کیفی و کمی خدمات و نظامهای آموزشی خود به ارزشیابی مستمر خدمات با روش‌شناسیهای گوناگون نیاز دارد (زانگ، زلر، گریفیث، متکاف و ویلیامز^۳، ۲۰۱۱). با این حال ارزشیابی یکی از چالشهای بزرگ سازمانها و بیشترین موردی است که در نظامهای آموزشی مورد غفلت واقع شده است (آگوئینیز و کریگر^۴، ۲۰۰۹) و همین امر موجب شده است که سازمانها با وجود صرف هزینه‌های بسیار در امر آموزش، نتیجه موردنظر را دریافت نکنند و برنامه‌ها اغلب پاسخگوی نیازهای فردی و اجتماعی افراد نباشند (بخشی حاجی‌خواجه‌لو و ستادی، ۱۳۹۶). وجود این مشکلات، تصمیم‌گیرندگان نظامهای آموزشی را به سمت سؤالاتی سوق داده است مانند اینکه چه راههایی برای اثربخشی بیشتر و بهبود دائمی برنامه‌های آموزشی وجود دارد (زیتزمان و واینهارت^۵، ۲۰۱۸). با انتخاب و به‌کارگیری الگوهای مناسب ارزشیابی، به این سؤالات، پاسخ داده می‌شود و زمینه ارائه اطلاعات برای اصلاح و بهبود برنامه در اختیار افراد قرار می‌گیرد (مدرسی و مدرسی، ۱۳۹۴). با توجه به تنوع برنامه‌های آموزشی از نظر محتوا، مخاطبان و نتایج، استفاده از یک الگوی خاص ارزشیابی برای تمام این برنامه‌ها کارایی یکسانی نداشته و بهتر است برای سطوح مختلف برنامه‌های آموزشی از الگوهای متناسب با آن استفاده نمود (همه‌مرادی، خراسانی و فتحی‌واجارگاه، ۱۳۹۳).

الگوهای گوناگون ارزشیابی را صاحب‌نظران معرفی کرده‌اند که با توجه به انتظارات ارزشیابان و تأکید ویژه بر جنبه‌های خاصی از برنامه (مدرسی و مدرسی، ۱۳۹۴) و نیز بر اساس اهداف و موضوعات ارزشیابی با یکدیگر تفاوت دارند (سیف، ۱۳۹۷). برخی از این الگوها بر ارزشیابی رفتار یادگیرنده، مواد آموزشی یا

1. Evaluation
2. Chilenski, Olson, Schulte, Perkins & Spoth
3. Zhang, Zeller, Griffith, Metcalf & Williams
4. Aguinis & Kraiger
5. Sitzmann & Weinhardt

شیوه‌های مربوط به آن، برخی بر اهمیت نقش ارزشیابی در برنامه‌ریزیهای آموزشی و برخی دیگر بر اهمیت گردآوری اطلاعات برای تصمیم‌گیری تمرکز دارند (همه‌مرادی و همکاران، ۱۳۹۳)؛ بنابراین هر الگوی ارزشیابی بر جنبه خاصی از برنامه توجه و از سایر جنبه‌های مهم غفلت کرده است و به دلیل وجود همین فواید و کاستیها در هر یک از الگوها، اغلب ترکیبی از آنها در ارزشیابی برنامه‌ها توصیه می‌شود (ولف^۱، ۱۳۹۷). امروزه در دنیا نیاز به توسعه مدل‌های ارزشیابی وجود دارد (سینگ و تندن^۲، ۲۰۱۶) و تدوین چارچوبی برای ارزشیابی آگاهانه معمولاً از انطباق دیدگاه‌های متفاوت در ارزشیابی آموزشی حاصل می‌شود (ولف، ۱۳۹۷). از مهم‌ترین دلایل این امر می‌توان به چندسطحی بودن و پیچیدگی برنامه‌های آموزشی (فتح‌آبادی، محمدی‌آریا، کاکائی و نودری، ۱۳۹۰)، تعاملات پویا و همزمان مؤلفه‌ها و متغیرهای مختلف در آموزش، اهداف آموزشی متنوع سازمانها (حسین‌پور و قدیری‌علمداری، ۱۳۹۵) و طیف گسترده عوامل سازمانی و فردی در طراحی آموزشی (خراسانی و حسینی‌ضرابی، ۱۳۹۳) اشاره کرد. در نظر گرفتن همه این موارد موجب تنوع و تعدد اهداف ارزشیابی می‌شود و برای دستیابی به این اهداف چندگانه، ارزشیابان به سمت توجه به عناصر گوناگون تأثیرگذار در برنامه و بهره‌گیری از رویکردهای متناسب با هر یک از عناصر سوق داده می‌شوند. نادیده گرفتن اطلاعات لازم درباره جنبه‌های مربوط به یک فعالیت آموزشی نشان‌دهنده شکست کار ارزشیابی است (ولف، ۱۳۹۷). از آنجا که برنامه‌های آموزشی گسترده اغلب اهداف و تأثیرات چندلایه و متنوع دارند (فرای و همسر^۳، ۲۰۱۲)، بهتر است ارزشیابی اثربخشی آنها با استفاده از منابع اطلاعاتی متعدد صورت گیرد (دورسی، شیفمن، ردکر، هایتکمپر، مک‌کلاسکی^۴، ۲۰۱۴). کمال مطلوب آن است که آموزش باید در چندین سطح و با استفاده از روشهای چندگانه ارزشیابی شود و به این ترتیب احتمال ارزشیابی مؤثر بیشتر خواهد بود. مقبولیت و پذیرش طرح‌های ارزشیابی ترکیبی به جای استفاده صرف از یکی از روشها، از جمله گامهای بزرگ در پیشرفت و توسعه ارزشیابی سازمانهاست (خراسانی و حسینی‌ضرابی، ۱۳۹۳). در این راستا گندمکار (۲۰۱۸) نیز بر استفاده از برنامه ارزشیابی جامع تأکید کرده و در نظر گرفتن همه‌جانبه مدلها و نظریه‌های ارزشیابی موجود را موجب غنی‌تر ساختن فعالیت ارزشیابان دانسته است. علی‌رغم تأکید پژوهشها بر لزوم ارزشیابی همه‌جانبه برنامه‌ها، همچنان فقدان الگویی جامع با تمرکز بر بهبود کیفیت و دستیابی به آموزشهای اثربخش احساس می‌شود و برنامه‌های اجتماعی بسیاری در سازمانها اجرا می‌شوند که مدیران و سیاستگذاران برای تصمیم‌گیری درباره ادامه یا توقف آنها اطلاعات مستندی در دست ندارند (زیتزمان و واینهارت، ۲۰۱۸).

یکی از این برنامه‌های اجتماعی گسترده، طرح همیار مشاور است که از سال ۱۳۸۰ در اداره امور تربیتی و مشاوره وزارت آموزش و پرورش ایران با اعتقاد به توانمندی دانش‌آموزان برای مشارکت

1. Wolf
2. Singh & Tandon
3. Frye & Hemmer
4. Dorsey, Schiffman, Redeker, Heitkemper, McCloskey

در گسترش خدمات مشاوره‌ای رشدی (تأکید بر توانمندی، نیازها، نقاط مثبت و نگاه به آینده) و به‌منظور ارتقای سلامت روان در مدرسه و جامعه و نیز استفاده از ظرفیت همسالان در انتقال آموزشهای تحصیلی، شغلی و آموزه‌های بهداشت روان طراحی و اجرا شده است (شیوه‌نامه اجرایی همیار مشاور، ۱۳۸۰). با توجه به اینکه این طرح دارای اهداف چندگانه و متنوع است و سابقه اجرایی نزدیک به دو دهه را در مدارس سطح کشور دارد و نیروی انسانی، هزینه، انرژی و زمان بسیار برای طرح‌ریزی و اجرای آن اختصاص داده شده است، نگاه جزئی‌نگرانه به تمام سطوح برنامه از زمان طراحی تا نتایج آن و ارائه مستندات و شواهد علمی در زمینه میزان تحقق‌پذیری اهداف اختصاصی و اهداف سند تحول بنیادین آموزش و پرورش و تعیین اثر برنامه بر گروه هدف، آموزش و پرورش و نیز در سطح جامعه، به‌منزله پشوتوانه علمی-اجرایی برای ادامه آن ضروری به‌نظر می‌رسد. با بررسی پیشینه پژوهشهای صورت گرفته در زمینه طرح همیار مشاور مشخص شده که تاکنون اقدام نظام‌مندی به‌منظور ارزشیابی و حصول اطمینان از اجرای درست برنامه در کشور صورت نگرفته است.

از این‌رو پژوهش حاضر با هدف پاسخگویی به این سؤالات شکل گرفته که اولاً مدل جامع ارزشیابی برنامه‌های اجتماعی به‌نحوی که ضمن توجه به گامهای تدوین و اجرای برنامه، سطوح افراد مختلف درگیر در برنامه را در نظر بگیرد، کدام است؟ و ثانیاً برنامه همیار مشاور آموزش و پرورش را به‌منزله یک برنامه اجتماعی گسترده چگونه می‌توان در قالب مدل پیشنهادی پژوهش گنجاند؟

روش پژوهش

مطالعه حاضر از نظر هدف کاربردی^۱، از نظر شیوه گردآوری داده‌ها از نوع توصیفی^۲ و از نظر ماهیت داده‌ها کیفی^۳ است. جامعه مشتمل بر اسناد علمی و اجرایی مرتبط با برنامه‌های اجتماعی گسترده، اعم از آیین‌نامه‌ها و دستورالعملهای اجرایی، محتوای آموزشی و روش تدریس، نتایج نظرسنجی عوامل اجرایی، آمارهای مهم آموزشی، تولیدات علمی-پژوهشی، میزان و شاخصهای تخصیص بودجه برنامه و نیز مشتمل بر کتابها، مقالات و پایان‌نامه‌های مرتبط با موضوع پژوهش بوده است. در این مطالعه به‌منظور نظام‌مند کردن جست‌وجوها، سعی بر آن بوده است که در بررسی شواهد علمی و اجرایی موردنیاز تدوین مدل ارزشیابی تا حد ممکن از نمودار جریان پریزما^۴ بهره گرفته شود. نمونه مشتمل بر اسناد علمی و اجرایی مرتبط با برنامه در سازمانها و نیز کتابها و پایان‌نامه‌های داخلی و خارجی با موضوع ارزشیابی آموزشی در پایگاههای علمی معتبر بوده است که به روش غیراحتمالی هدمند مورد بررسی قرار گرفته‌اند.

مقالات به‌طور مشخص در مجلات حیطة مدیریت آموزشی و برنامه‌ریزی درسی با کلیدواژگان

1. Applied research
2. Descriptive
3. Qualitative
4. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)

ارزشیابی آموزشی، الگوهای ارزشیابی، ارزشیابی برنامه‌های اجتماعی، تدوین مدل ارزشیابی، الگوی جامع ارزشیابی، مدل تلفیقی ارزشیابی، مدل چندسطحی ارزشیابی، الگوی ارزشیابی مدیریت‌گرا، الگوی ارزشیابی مبتنی بر هدف که در سالهای ۱۳۹۰ تا ۱۳۹۸ به زبان فارسی در پایگاه‌های اطلاعاتی Irandoc, Magiran, SID و در سالهای ۲۰۱۰ تا ۲۰۱۹ به زبان انگلیسی در پایگاه‌های اطلاعاتی ScienceDirect, Google Scholar و SpringerLink نمایه شده بود، مورد جست‌وجو قرار گرفت. پس از حذف مقالات تکراری، موارد واجد شرایط و مناسب از نظر پژوهشگر به‌عنوان نمونه قابل بررسی در نظر گرفته شد و مابقی حذف شدند.

● روش اجرا

۱. برگزاری جلسات و نشستهایی با مدیران و سیاستگذاران برخی برنامه‌های اجتماعی کشوری به‌منظور شناسایی اهداف، انتظارات و نیازهای اطلاعاتی آنان برای پیشبرد برنامه‌ها و مشکلاتی که در این زمینه با آن مواجه‌اند.
۲. بررسی مستندات سازمانی مرتبط با برنامه‌های اجتماعی، اعم از آیین‌نامه، دستورالعمل، محتوا و بودجه به‌منزلهٔ مرتبط‌ترین اسناد مربوط به موضوع پژوهش.
۳. بررسی پیشینه نظری مرتبط با الگوهای عمومی ارزشیابی آموزشی به‌منظور آشنایی با اهداف، انواع، اجزا و دسته‌بندی آنها.
۴. بحث و تبادل نظر متخصصان مدیریت آموزشی، برنامه‌ریزی درسی و روان‌سنجی به‌منزلهٔ خبرگان این عرصه در زمینهٔ الگوهای ارزشیابی موردنظر پژوهشگر و سایر الگوها.
۵. انتخاب الگوهای ارزشیابی متناسب با موضوع و هدف پژوهش و تلفیق آنها به نحوی که ضمن جامعیت داشتن هر الگو، محدودیتها و ضعفهای الگوی دیگر را به شکلی مناسب پوشش دهد.
۶. روایی‌یابی محتوایی مدل: از آنجایی که برای تدوین مدل حاصل مدل‌های نظری و پژوهش‌های موجود در ادبیات مطالعه شده است، از این‌رو انتظار می‌رود مدل از روایی نظری و مفهومی^۱ مناسبی برخوردار باشد. برای بررسی روایی محتوایی نیز مدل نهایی شده در اختیار ده نفر از خبرگان مدیریت آموزشی و برنامه‌ریزی درسی قرار گرفت تا از نظر مرتبط بودن^۲ مفاهیم نظری، منطقی بودن^۳ ارتباطات و انسجام^۴ مؤلفه‌های مدل، نظر خود را در طیف چهار درجه‌ای لیکرت بیان کنند. هفت تن از این متخصصان از میان اعضای هیأت علمی دانشگاه‌های تهران (حداقل با ۱۰ سال سابقه تدریس) و سه نفر نیز از فارغ‌التحصیلان دوره دکتری این دانشگاه‌ها انتخاب شدند که

1. Conceptual
2. Relatedness
3. Rationality
4. Coherency

دارای فعالیتهای پژوهشی منظم در حوزه ارزشیابی و نظامهای مرتبط با آن بودند. شاخص روایی محتوا^{۱۰} ۰/۹۰ حاصل شد. این شاخص براساس نظر یاری، پویاکیان، جعفری، علیپور و ورمزیار (۱۳۹۶) نشان از روایی مناسب است.

۷. روایی یابی بیرونی و بوم‌شناختی^{۱۱}: این دو روش روایی‌یابی برای بررسی میزان تعمیم‌پذیری اقدامات و نتایج حاصل از یک بررسی به دنیای واقعی به کار می‌روند (کمبل^{۱۲}، ۱۹۵۷؛ برکوویتز و دانرستین^{۱۳}، ۱۹۸۲؛ هوک^{۱۴}، ۲۰۰۱؛ اشکرافت و رادوانسکی^{۱۵}، ۲۰۰۹؛ مک‌دونالد و تاتلر^{۱۶}، ۲۰۱۳ و هولمن، هوگ، کمز و هسلز^{۱۷}، ۲۰۲۰). برای بررسی این نوع روایی و امکان‌سنجی میزان عملی بودن مدل و همچنین تطابق یافته‌ها، در سه سطح با مدیران ارشد، مدیران میانی و پایه و همچنین کارشناسان برخی برنامه‌های اجتماعی در سازمانها، جلسات بحث گروهی برگزار شد. در این فرایند برخی از اصلاحات در مدل انجام شد و در نهایت بالای ۹۰ درصد افراد بر وجود این نوع روایی در مدل نهایی صحه گذاشتند.

۸. نهایی‌سازی مدل ارزشیابی چندسطحی تلفیقی و مصورسازی آن.

۹. بررسی و تطبیق طرح همیار مشاور با مدل ارزشیابی تلفیقی.

یافته‌ها

مدل ارزشیابی چندسطحی تدوین شده در این پژوهش حاصل تلفیق سه الگوی سیپ^{۱۸}، کرک پاتریک^{۱۹} و مدل منطقی است که چرایی انتخاب هر یک و منطق تلفیق آنها در این بخش ارائه می‌شود. یکی از الگوهایی که تمامی سطوح، عناصر و مؤلفه‌های برنامه را به‌طور جامع و براساس شاخصهای کمی و کیفی مناسب مورد ارزشیابی قرار می‌دهد، الگوی سیپ است (هاکان و سول^{۱۱}، ۲۰۱۱ و نیک‌خواه، شریف، نصر و طالبی، ۱۳۹۰). این الگو کل نگر است و برنامه را در تمامی مراحل آغازین و پیاده‌سازی (ارزشیابی تکوینی^{۱۲}) و نیز مراحل پایانی (ارزشیابی تراکمی^{۱۳}) بررسی می‌کند (استافل‌بیم^{۱۴} و ژانگ، ۲۰۱۷).

1. Content Validity Index (CVI)
2. External and ecological
3. Campbell
4. Berkowitz & Donnerstein
5. Hook
6. Ashcraft & Radvansky
7. Macdonald & Tatler
8. Holleman, Hooge, Kemner & Hessels
9. CIPP
10. Kirkpatrick
11. Hakan & Seval
12. Formative
13. Summative
14. Stufflebeam

همسو بودن الگوی سیپ با نظریه سیستمی و نظریه پیچیدگی^۱ این قابلیت را ایجاد کرده است که یک برنامه آموزشی به‌عنوان سیستم باز در نظر گرفته شود و ضمن بررسی تک‌تک عناصر، روابط پویا میان عناصر و محیط برنامه آموزشی نیز مورد بررسی قرار گیرد و تصویری وسیع و گسترده در قالب چارچوبی انعطاف‌پذیر در اختیار ارزشیابان، مدیران و تصمیم‌گیرندگان و سایر ذینفعان برنامه قرار دهد (موریو،^۲ ۲۰۱۷).

الگوی سیپ زیرمجموعه رویکرد مبتنی بر مدیریت یا الگوی تصمیم‌گرا^۳ است (استافل‌بیم و شینکفیلد^۴، ۲۰۰۷) و هدف اصلی آن ارائه اطلاعات مفید برای تصمیم‌گیرندگان برنامه آموزشی است (کلاگن^۵، استافل‌بیم و وینگیت^۶، ۲۰۰۳). الگوی ارزشیابی سیپ فرایند-محور است که فرایند طراحی، دستیابی و فراهم ساختن اطلاعات ضروری برای تصمیم‌گیرندگان از طریق ارزشیابی زمینه^۷، درونداد^۸، فرایند^۹ و محصول^{۱۰} را شامل می‌شود (لی و چا^{۱۱}، ۲۰۱۵). ارزشیابی زمینه به سنجش نیازها و مشکلات، فرصتها و امکانات بالقوه برنامه می‌پردازد. یافته‌های این بخش، خط پایه‌ای مفید برای ارزشیابی پیامدهای آتی (محصول و فرآورده‌ها) فراهم می‌کنند (مک‌لمور^{۱۲}، ۲۰۰۹). ارزشیابی درونداد به بررسی عناصر ورودی و عوامل مؤثر بر توسعه یا اصلاح برنامه می‌پردازد. ارزشیابی فرایند به چگونگی اجرای برنامه اشاره کرده و اغلب منبع اطلاعات برای بهسازی برنامه است. ارزشیابی محصول دستاوردهای پیش‌بینی‌شده و پیش‌بینی‌نشده فعالیت‌های آموزشی را بررسی می‌کند (بکر، نیمرمن، دلویل و سئو^{۱۳}، ۲۰۱۷).

الگوی سیپ در عین جامعیت، نگاه کل‌نگرانه به اهداف برنامه دارد (خراسانی و حسینی‌ضرابی، ۱۳۹۳) و از آنجایی که بروندهای یک برنامه به لحاظ اندازه یا اهمیت متفاوت‌اند و ممکن است گاهی اوقات تمایز آنها از پیامدها دشوار باشد، استفاده از مدل منطقی با قابلیت ارائه توصیفی اطلاعات به تفکیک خروجیها و پیامدهای برنامه می‌تواند در ارزشیابی کمک‌کننده باشد (فرای و همر،^{۲۰۱۲}). مدل منطقی نیز از دیرباز به‌مثابه پایه ارزشیابی، ضمن دسته‌بندی چشم‌انداز و نتایج چندسطحی عموم برنامه‌ها و به‌خصوص برنامه‌های اجتماعی و مشارکتی (دورسی و همکاران، ۲۰۱۴) چگونگی دستیابی به انتظارات برنامه در کوتاه‌مدت، میان‌مدت و بلندمدت را تشریح می‌کند (کلافام، مانینگ، ویلیامز، ابرایان و

1. Complexity theory
2. Moreau
3. Decision-oriented
4. Shinkfield
5. Kellaghan
6. Wingate
7. Context
8. Input
9. Process
10. Product
11. Lee & Cha
12. McLemore
13. Becker, Timmerman, Delville & Seo

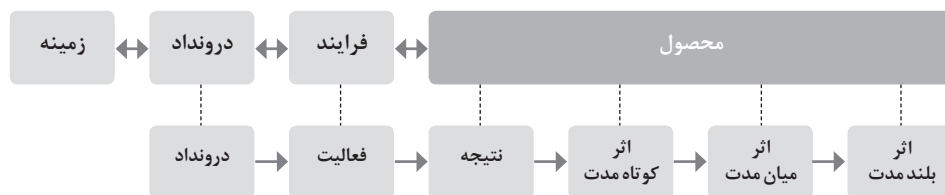
سادلند^۱، ۲۰۱۷). این مدل بر پایه تئوری برنامه^۲ بنا شده است که طبق این تئوری مداخلاتی مانند یک پروژه، برنامه، راهبرد، ابتکار عمل یا یک سیاست، به زنجیره‌ای از نتایج میانی و در نهایت به نتایج منتخب یا قابل انتظار می‌انجامد (فانل و راجرز^۳، ۲۰۱۱).

این مدل به نوعی هم نتیجه - محور^۴ و هم فرایند - محور^۵ است (لی و چا، ۲۰۱۵) و به طور کلی ارزشیابی یک برنامه را در چهار مرحله ارائه می‌کند: درونداد، فعالیت^۶، نتیجه^۷ و تأثیر^۸ (فیتزپاتریک، سندرز و وردن^۹، ۲۰۰۴). مرحله درونداد شامل منابع انسانی، مالی یا دیگر منابع ضروری برای پشتیبانی و حمایت از برنامه آموزشی است. مرحله فعالیت به اتفاقات و عناصر منجر به تولید درونداد اشاره دارد. مرحله نتیجه شامل محصولات یا خدمات در دسترس برای مخاطب برنامه است. مرحله تأثیر به سود و تغییرات حاصل از مرحله فعالیت و نتیجه، اشاره دارد که به سه دسته کوتاه‌مدت، میان‌مدت و بلندمدت تقسیم می‌شود. تأثیر کوتاه‌مدت، زمانی است که مرحله تأثیر با نتیجه ارتباطی نزدیک داشته باشد و در واقع هر دستاوردی است که مستقیماً با برنامه ایجاد شده باشد. تأثیر میان‌مدت به کاربردهای پی‌آیند تأثیر کوتاه‌مدت همچون تغییر در ارزشها، نگرشها و شرایط اطلاق می‌شود. تأثیرات بلندمدت به اثر کلی برنامه در سطح اجتماعی و کلان آن اشاره دارد (کالو و نورمن^{۱۰}، ۲۰۱۸).

اگرچه عناصر درونداد، فعالیت و برون‌داد مدل منطقی با عناصر درونداد، فرایند و محصول رویکرد سیپ اشتراکاتی دارند، اما اشاره به چند نقطه ضعف مدل منطقی می‌تواند دلیل تلفیق با الگوی سیپ را پیش بکشد (فرای و همر، ۲۰۱۲). بزرگ‌ترین ضعف مدل منطقی عدم توجه دقیق به بافت آموزشی برنامه است، به این معنا که همه چیز بعد از طراحی برنامه بررسی می‌شود. این ضعف در بخش زمینه الگوی سیپ به منزله یک بخش مستقل در ارزشیابی پوشش داده شده است (استافل‌بیم و ژانگ، ۲۰۱۷). ضمناً مدل منطقی مبتنی بر رویکرد تقلیل‌گرایی^{۱۱} است که مطابق با آن، میان عناصر برنامه روابط علت - معلولی و فرضیه خطی یا تقلیلی وجود دارد، به این معنا که تغییر در هر یک از عناصر، تأثیری قابل انتظار بر پیامد برنامه دارد و نیز انتظار می‌رود تغییر کوچک تأثیری کوچک داشته و تأثیر بزرگ در پی تغییر بزرگ اتفاق بیفتد (گروسمن و سالاس^{۱۲}، ۲۰۱۱)، اما الگوی سیپ نه تنها بر تأثیر عناصر بر یکدیگر بلکه بر ارتباطات پیچیده میان آنها تمرکز می‌کند و طبق این الگو تغییرات کوچک در عناصر

1. Clapham, Manning, Williams, O'Brien & Sutherland
2. Program theory
3. Funnell & Rogers
4. Result-oriented
5. Process-oriented
6. Activity
7. Output
8. Outcome
9. Fitzpatrick, Sanders & Worthen
10. Kalu & Norman
11. Reductionist approach
12. Grossman & Salas

برنامه ممکن است به تغییرات بزرگ در پیامدها و به عکس آن بینجامد (موریو، ۲۰۱۷). همچنین به خشک و غیرمنعطف بودن مدل منطقی به عنوان دیگر ایراد وارد بر این الگو می توان اشاره کرد، زیرا این مدل صرفاً به بررسی اهداف از پیش تعیین شده و نتایج مورد انتظار اختصاص دارد و پاسخ دهی به سایر اطلاعات برنامه مانند نتایج ناخواسته و پیش بینی نشده را محدود می کند (کوکسی، گیل و کلی^۱، ۲۰۰۱). در حالی که هدف الگوی سیپ ارائه اطلاعاتی در مورد تأثیرات خواسته و ناخواسته برنامه است (بکر و همکاران، ۲۰۱۷). در شکل ۱، دو الگوی سیپ و مدل منطقی با یکدیگر تلفیق شده اند.



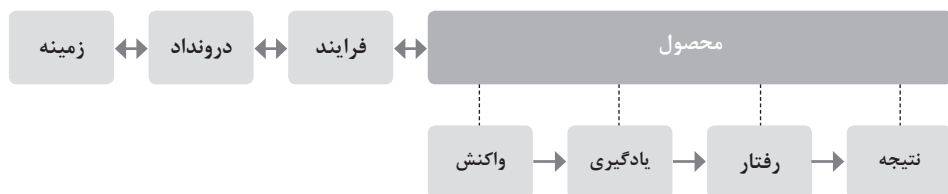
شکل ۱. تلفیق الگوی سیپ و مدل منطقی

با توجه به رویکرد بسیار مدیریتی و بیش از حد بالا به پایین الگوی سیپ و ضعف این الگو در مشارکت دادن فراگیران و ذی نفعان سطوح پایین تر برنامه در ارزشیابی (خراسانی و حسینی ضرابی، ۱۳۹۳)، استفاده از الگوهای مشارکتی برای پوشش دهی این ضعف، بررسی دقیق انواع تأثیرات برنامه بر فراگیران و نادیده گرفته نشدن اطلاعات این گروه (بینش، بختیاری فاینبری و قراری، ۱۳۹۵) ضروری به نظر می رسد. به همین دلیل الگوی کرک پاتریک به عنوان الگویی هدف-محور و متمرکز بر فراگیران (ریو، روکو، اسمیت و چانگ^۲، ۲۰۱۷) به الگوی پیشنهادی اضافه می شود. الگوی کرک پاتریک به عنوان رایج ترین الگوی ارزشیابی در سازمانها، نتایج شناختی، مهارتی، نگرشی و سازمانی یک برنامه آموزشی را مدنظر قرار می دهد (کرک پاتریک، ۲۰۰۹). این مدل دارای چهار سطح واکنش^۳، یادگیری^۴، انتقال^۵ یا رفتار و نتایج^۶، است (پارک، لی، کیم، جون و کیم^۷، ۲۰۱۷).

سطح واکنش به رضایت شرکت کنندگان از آموزش اشاره دارد. سطح یادگیری به تغییر در سطح دانش، مهارت و نگرش افراد در اثر آموزش اشاره دارد. سطح رفتار به کاربرد عملی محتوای یاد گرفته شده در محیط اطلاق می شود. سطح نتیجه به تأثیرات مثبت حاصل از سطوح واکنش، یادگیری و رفتار بر عملکرد سازمان اشاره می کند (پارک و همکاران، ۲۰۱۷). چهار سطح کرک پاتریک به صورت متوالی

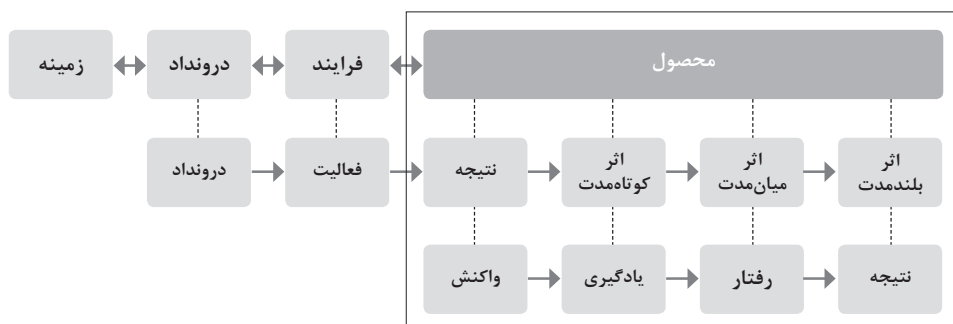
1. Cooksy, Gill & Kelly
2. Reio, Rocco, Smith & Chang
3. Reaction
4. Learning
5. Transfer
6. Results
7. Park, Lee, Kim, Jun & Kim

و سلسله‌مراتبی هستند. به این معنا که حرکت از یک سطح به سطح دیگر ارزشیابی را سخت‌تر و وقت‌گیرتر کرده است، اما اطلاعاتی باارزش‌تر را فراهم می‌آورد (پارک و همکاران، ۲۰۱۷).
 گرچه الگوی کرک‌پاتریک طبقه‌بندی مفیدی از پیامدهای برنامه ارائه می‌دهد، اما باین حال آموزشگران را به‌سوی یک ارزشیابی کامل از برنامه آموزشی هدایت نمی‌کند (بی‌تس، ۲۰۰۴). از جمله ایرادهای واردشده بر این الگو به‌عنوان یک رویکرد مبتنی بر هدف، تمرکز صرف بر اهداف نهایی و غفلت از ارزشیابی فرایند (هیز، اسکات، آبراجینسکاس، اسکچا و استاوت، ۲۰۱۶) و نیز تکیه انحصاری بر نتایج از پیش تعیین‌شده به‌عنوان ملاک قضاوت و داوری برنامه است (یاردلی و درنان، ۲۰۱۲). این ضعف الگوی کرک‌پاتریک با الگوی سیپ که زمینه، فرایند و عناصر گوناگون مؤثر بر نتایج برنامه را در نظر می‌گیرد، پوشش داده‌شده است (ریو و همکاران، ۲۰۱۷). همچنین کرک‌پاتریک همچون بسیاری از برنامه‌های کلاسیک و سنتی ارزشیابی به اثبات برنامه می‌پردازد، بنابراین در انتهای برنامه اجرا می‌شود. در مقابل، هدف الگوی سیپ بهبود و ارتقای برنامه و ارائه اطلاعات برای تصمیم‌گیری درباره برنامه است و نه اثبات آن (گندمکار، ۲۰۱۸). شکل ۲، نحوه تلفیق الگوی سیپ و الگوی کرک‌پاتریک را نشان می‌دهد.



شکل ۲. تلفیق الگوی سیپ و کرک‌پاتریک

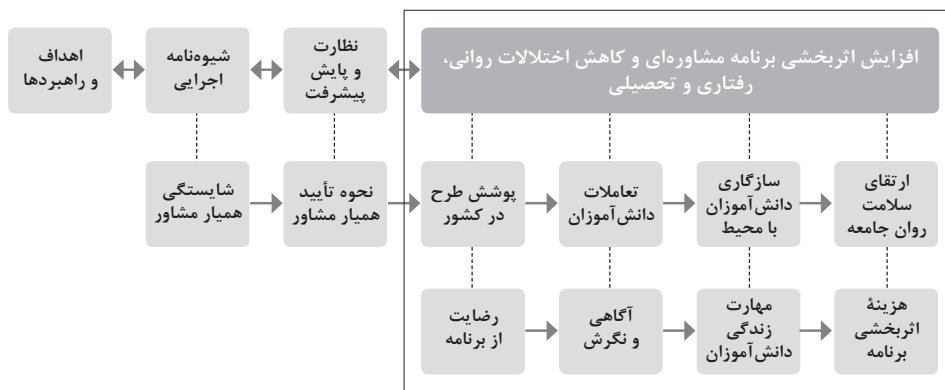
با توجه به ذکر نقاط قوت و ضعف هر یک از الگوهای ارزشیابی منتخب و دلایل مطروحه برای استفاده و تلفیق آنها با یکدیگر، مدل مفهومی تلفیقی نهایی در شکل ۳ ارائه شده است.



شکل ۳. مدل ارزشیابی چندسطحی تلفیقی

1. Bates
2. Hayes, Scott, Abraczinskas, Scaccia, Stout
3. Yardley & Dornan

شکل شماره ۳ نمای کلی از مدل ارزشیابی پیشنهادی پژوهش است که حاصل تلفیق سه الگوی سیپ (سطر اول)، مدل منطقی (سطر دوم) و الگوی کرک پاتریک (سطر سوم) است. هر مستطیل نشان‌دهنده یک بخش از الگوی ارزشیابی است. فلش دوطرفه میان بخشها ارتباط دوطرفه و سیستمی میان عناصر را نشان می‌دهد و فلش یک‌طرفه ارتباط خطی میان عناصر را نشان می‌دهد. ضمناً علامت نقطه‌چین عمودی و مشابهت الگوی تصویری داخل مستطیلهایی که با نقطه‌چین به هم وصل شده اند، نشان‌دهنده متناظر بودن نسبی آن مراحل با یکدیگر است. بزرگ‌تر بودن اندازه مستطیل محصول نشان از آن دارد که بخش محصول علاوه بر بررسیهای اطلاعاتی متعلق به الگوی سیپ، از طرفی به کمک بخشهای نتیجه و تأثیرات مدل منطقی و از طرف دیگر به کمک الگوی کرک پاتریک تکمیل می‌شود. در شکل ۴، برنامه همیار مشاور در قالب مدل ارزشیابی چندسطحی تلفیقی گنجانده شده و در هر بخش یک شاخص ارزشیابی برنامه به‌عنوان نمونه ارائه شده است.



شکل ۴. برنامه همیار مشاور در قالب مدل ارزشیابی چندسطحی تلفیقی

بحث و نتیجه‌گیری

در مدل ارزشیابی چندسطحی تلفیقی این پژوهش، سعی بر آن شده است که نگاهی جامع به سطوح متفاوت و لایه‌های مهم در برنامه آموزشی گسترده صورت گیرد؛ از اقدامات زمان طراحی و نیازسنجی برنامه (بخش زمینه) گرفته تا زیرساختها و امکانات لازم برای اجرای برنامه (بخش درونداد)، پیاده‌سازی و نظارت بر حسن اجرای برنامه (بخش فرایند)، تحقق‌پذیری تمامی اهداف پیش‌بینی شده و پیش‌بینی نشده در طول زمان (بخش محصول)، اثربخشی برنامه در فراگیران پیش از ارائه خدمات و بلافاصله پس از آن (بخش واکنش و یادگیری)، مدتی پس از آن (بخش رفتار) و نتایج مالی و استراتژیک برنامه بر سازمان (بخش نتیجه).

با طراحی و به‌کارگیری این مدل، رویکرد اداری ارزشیابی برنامه‌های آموزشی سازمانها می‌تواند

جای خود را به رویکرد علمی بدهد، زیرا این مدل بر اساس نتایج پژوهش‌های داخلی و خارجی و نیز بر اساس مبانی نظری موجود ارزشیابی و انطباق دیدگاه‌های متفاوت در این حوزه حاصل شده و مطلوبیت آن از نظر متخصصان نیز مورد تأیید قرار گرفته است. در واقع با بهره‌گیری از رویکردهای متفاوت در ارزشیابی سعی در توسعه مدل‌های ارزشیابی با نگاهی جامع به سطوح متفاوت و لایه‌های مهم در برنامه‌های آموزشی گسترده بوده است (سینگ و تندون، ۲۰۱۶).

در ضمن برای نخستین بار در سطح کشور، مدلی تلفیقی از رویکرد سیپ، کرک پاتریک و مدل منطقی طراحی شده و با این تلفیق، ضعف الگوهای ارزشیابی محدود پوشش داده شده است، زیرا به اعتقاد زیتزمان و واینهارت (۲۰۱۸) با وجود تأکید پژوهش‌ها بر لزوم ارزشیابی همه‌جانبه برنامه‌ها، همچنان فقدان الگویی جامع با تمرکز بر بهبود کیفیت و دستیابی به آموزش‌های اثربخش احساس می‌شود. همچنین فرایند - محور و هدف - محور بودن از دیگر قابلیت‌های مدل مذکور است؛ به این معنا که اقدامات پیش، ضمن و پس از اجرای برنامه مورد ارزشیابی قرار گرفته و به جای تمرکز صرف بر نتایج، ویژگی‌های فردی، زمینه‌ای و محیطی آموزش نیز مورد توجه قرار گرفته است. از این رو برای برنامه‌های نوپا و نیز برای برنامه‌های با سابقه اجرای طولانی می‌تواند کاربرد داشته باشد (استافل‌بیم و ژنگ، ۲۰۱۷). تلفیق الگوهای هدف - محور و فرایند - محور در پژوهش‌های بسیاری مورد تأکید قرار گرفته است (چو^۱، پارک، جو^۲، جئونگ^۳ و لیم^۴، ۲۰۰۹؛ فرای و هم، ۲۰۱۲؛ هیز و همکاران، ۲۰۱۶؛ بکر و همکاران، ۲۰۱۷؛ همه‌مرادی و همکاران، ۱۳۹۳).

همچنین در این الگو به چندسطحی بودن و پیچیدگی برنامه‌های آموزشی، تعاملات پویا و همزمان مؤلفه‌ها و متغیرهای مختلف در آموزش، اهداف آموزشی متنوع سازمانها و طیف گسترده عوامل سازمانی و فردی در طراحی آموزشی توجه شده است، از این رو می‌توان آن را در ارزشیابی هر برنامه اجتماعی و آموزشی با هر سطح از گستردگی مورد استفاده قرار داد. در واقع قابلیت اجرایی آن به گونه‌ای است که تنوع محتوا، مخاطب، اهداف و سازمان برگزارکننده یک برنامه در کاربرد و به کارگیری این الگو محدودیتی ایجاد نمی‌کند. در پژوهش پیتون و شیکیتانو^۵ (۲۰۱۷) نیز بر این جنبه از الگوهای جامع ارزشیابی تأکید شده است. توجه ویژه به اهداف و تحقق‌پذیری آنها را می‌توان به‌منزله یکی دیگر از نقاط قوت این الگوی ارزشیابی در نظر گرفت، به طوری که در طراحی الگو ضمن تفکیک زمانی کوتاه‌مدت، میان‌مدت و بلندمدت نتایج، مثبت و منفی بودن، پیش‌بینی‌شده و پیش‌بینی‌نشده بودن و نیز ابعاد فردی، سازمانی و اجتماعی نتایج برنامه نیز لحاظ شده است.

1. Cho
2. Jo
3. Jeung
4. Lim
5. Peyton & Scicchitano

تازگی مدل یادشده را می‌توان از دید مشارکت‌دادن ذی‌نفعان متعدد برنامه و توجه به اصل هماهنگی و ارتباط ذی‌نفعان در طراحی، تولید و ارزشیابی آموزشی نگریست، زیرا از منابع اطلاعاتی متعددی همچون گروه‌های سطح‌بندی‌شده ذی‌نفعان و افراد درگیر در برنامه در سطوح سیاست‌گذاری، مدیریتی، اجرایی، علمی و مخاطبان آموزش برای سنجش مطلوبیت برنامه بهره گرفته می‌شود. نتیجه این مشارکت در ارزشیابی، دستیابی به اطلاعات مهم و چندجانبه برای سیاست‌گذاری و تصمیم‌گیری درباره توقف، اصلاح یا تداوم برنامه است (فریمن، ۲۰۰۴).

با توجه به قابلیت‌های این مدل ارزشیابی که پیشتر ذکر شد، هر برنامه اجتماعی و آموزشی که در سطح گسترده اجرا شود می‌تواند با این مدل مورد ارزشیابی قرار بگیرد که در این پژوهش چگونگی تطبیق برنامه همیار مشاور به‌مثابه یک برنامه اجتماعی گسترده با مدل ارزشیابی تدوین شده مورد بررسی قرار گرفته است.

از آنجا که طرح همیار مشاور دارای اهداف چندلایه و متعدد است و برای پیاده‌سازی و پیشبرد مطلوب آن، سطوحی از افراد با شرح وظایف و جایگاه‌های مختلف در سطح کشور در آن فعالیت می‌کنند، مدل ارزشیابی تلفیقی این پژوهش می‌تواند برای ارزشیابی آن مناسب باشد. برای این امر، بایستی در پی شناسایی ظرفیتها، نیازها و ویژگیهای برنامه و سازمان مجری آن، شاخصهای ارزشیابی و گویه‌های هر شاخص کاملاً متناسب با برنامه تدوین شود؛ به‌عنوان مثال در بخش فرایند، شاخص کیفیت اقدامات شورای راهبری شهرستانی برنامه همیار مشاور با گویه‌های آگاهی از اهداف و راهبردهای برنامه، پایبندی به شیوه‌نامه اجرایی ابلاغی، تعامل با استان، نظارت بر حسن اجرای طرح در مدارس و دستیابی به اهداف غایی برنامه مورد ارزشیابی در یک طیف لیکرتی قرار گیرد. در نهایت برای بررسی وضعیت برنامه در میزان تحقق‌پذیری اهداف، نمرات هر گروه به گویه‌های مقیاس در حکم وضعیت موجود در نظر گرفته شده و با نمره مطلوبی که حاصل نمرات ارزشیابی مدیران و سیاست‌گذاران برنامه است، مورد مقایسه قرار می‌گیرد. بیشترین اختلاف به‌منزله اولین اولویت اصلاح و به‌ترتیب اولویت‌های دیگر براساس همین شاخص معین شده و راهکارهای بازبینی علمی و کاربردی در اختیار مدیران و سیاست‌گذاران برنامه قرار خواهد گرفت. راهکارهایی که با پیاده‌سازی آنها به احتمال زیاد میزان تحقق‌پذیری اهداف و اثربخشی برنامه بیشتر خواهد شد.

پیشنهاد‌های پژوهشی

برنامه‌های اجتماعی گسترده در سازمانها با الگوی ارزشیابی چندسطحی تلفیقی مورد ارزشیابی قرار گیرند و نیز پیشنهاد ارزشیابی برنامه همیار مشاور در آموزش و پرورش عملیاتی و اجرایی شود.

1. Freeman

- بخشی حاجی خواجه‌لو، شهرام و ستادی، صدرالدین. (۱۳۹۶). ارزشیابی اثربخشی کارگاه آموزشی احیای قلبی ریوی مغزی در محیط بالینی بر اساس الگوی کرک پاتریک. *نشریه پرستاری قلب و عروق*، ۶(۳)، ۱۶-۲۲.
- بینش، مرتضی؛ بختیاری‌فایندری، منصوره و قراری، محمد. (۱۳۹۵). اهداف و ضرورت ارزشیابی برنامه‌درسی و الگوهای آن. *مطالعات روانشناسی و علوم تربیتی*، ۸(۲)، ۵۳-۶۷.
- حسین‌پور، رضا و قدیری‌علمداری، قدیر. (۱۳۹۵). مطالعه تطبیقی الگوهای ارزشیابی آموزشی مبتنی بر هدف و سبب به‌منظور استخراج نقاط قوت و ضعف آنها. *پنجمین کنفرانس ملی توسعه پایدار در علوم تربیتی و روانشناسی، مطالعات اجتماعی فرهنگی، موسسه آموزش عالی مهر اروند، تهران.*
- خراسانی، اباصلت و حسینی‌ضرابی، حامد. (۱۳۹۳). *ارزشیابی آموزشی (از تئوری تا عمل)*. تهران: مرکز آموزش و تحقیقات صنعتی ایران.
- سیف، علی‌اکبر. (۱۳۹۷). *روش‌های اندازه‌گیری و ارزشیابی آموزشی*. تهران: دوران.
- ششویه‌نامه اجرایی همیار مشاور. (۱۳۸۰). دبیرخانه کشوری همیار مشاور، اداره مشاوره و امور تربیتی آموزش پرورش استان مازندران.
- فتح‌آبادی، جلیل؛ محمدی‌آریا، علیرضا؛ کاکائی، یزدان و نوذری، محمد. (۱۳۹۰). ارزشیابی جامع دانشجویی دانشگاه آزاد اسلامی واحد گرگان بر اساس الگوی CIIP از دیدگاه کارکنان آموزشی در سال تحصیلی ۸۶-۸۵. *مجله توسعه پژوهش در پرستاری و مامایی*، ۸(۲)، ۵۸-۶۸.
- مدرسی سریزدی، آسیه‌السادات و مدرسی سریزدی، سیده‌جمیله. (۱۳۹۴). الگوی ارزشیابی سبب و کاربرد آن در آموزش‌های رسمی فنی و حرفه‌ای. *کارآفرین*، ۲(۲)، ۳۳-۴۱.
- نیک‌خواه، محمد؛ شریف، مصطفی؛ نصر، احمدرضا و طالبی، هوشسنگ. (۱۳۹۰). امکان‌سنجی کاربرد شاخص‌های ارزیابی آموزش عالی در حوزه ارزشیابی برنامه‌درسی دوره تحصیلات تکمیلی، بر مبنای الگوی سیپ. *مدیریت و برنامه‌ریزی در نظام‌های آموزشی*، ۴(۷)، ۱۰۰-۱۳۲.
- ولف، ریچارد. (۱۳۹۷). *ارزشیابی آموزشی (مبانی سنجش توانایی و بررسی برنامه)*، ترجمه علیرضا کیامنش. تهران: مرکز نشر دانشگاهی.
- همه‌مرادی، مجید؛ خراسانی، اباصلت و فتحی‌واجارگاه، کوروش. (۱۳۹۳). ارزشیابی اثربخشی دوره‌های آموزش ضمن خدمت شرکت ملی گاز ایران بر اساس سه الگوی پاتریک، فیلیپس و سیپ. *مدیریت منابع انسانی در صنعت نفت*، ۶(۱۹)، ۱۰۳-۱۳۰.
- یاری، سعید؛ پویاکیان، مصطفی؛ جعفری؛ محمدجواد؛ علیپور، عباس و ورمزیار، سکینه. (۱۳۹۶). تهیه و اعتبارسنجی پرسشنامه ارزیابی وضعیت ایمنی پمپ بنزین‌های شهری. *مجله ارتقای ایمنی و پیشگیری از مصدومیت‌ها*، ۵(۳)، ۱۶۹-۱۸۰.

- Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology*, 60, 451-474.
- Ashcraft, M. H., & Radvansky, G. A. (2009). *Cognition* (5th ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, 27(3), 341-347.
- Becker, H., Timmerman, G. M., Delville, C., & Seo, E. (2017). A multifaceted model to evaluate interprofessional education in clinical nurse specialist programs. *Clinical Nurse Specialist*, 31(5), 243-251.

- Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist*, 37(3), 245-257.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297-312.
- Chilenski, S. M., Olson, J. R., Schulte, J. A., Perkins, D. F., & Spoth, R. (2015). A multi-level examination of how the organizational context relates to readiness to implement prevention and evidence-based programming in community settings. *Evaluation and Program Planning*, 48, 63-74.
- Cho, Y., Park, S., Jo, S. J., Jeung, C.-W., & Lim, D. H. (2009). Developing an integrated evaluation framework for e-learning. In C. X. Wang (Ed.), *Handbook of research on e-learning applications for career and technical education: Technologies for vocational training* (pp. 707-722). Hershey, PA: IGI Global.
- Clapham, K., Manning, C., Williams, K., O'Brien, G., & Sutherland, M. (2017). Using a logic model to evaluate the Kids Together early education inclusion program for children with disabilities and additional needs. *Evaluation and Program Planning*, 61, 96-105.
- Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning*, 24(2), 119-128.
- Dorsey, S. G., Schiffman, R., Redeker, N. S., Heitkemper, M., McCloskey, D. J., Weglicki, L. S., & Grady, P. A. (2014). National Institute of Nursing Research Centers of Excellence: A logic model for sustainability, leveraging resources, and collaboration to accelerate cross-disciplinary science. *Nursing Outlook*, 62(6), 384-393.
- Farsi, M., & Sharif, M. (2014). Stufflebeam's CIPP model & program theory: A systematic review. *International Journal of Language Learning and Applied Linguistic World*, 6(3), 400-406.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). White Plains, NY: Longman Publishers.
- Freeman, R. E. (2004). The stakeholder approach revisited. *Zeitschrift für Wirtschafts-und Unternehmensethik*, 5(3), 228-241.
- Frye, A. W., & Hemmer, P. A. (2012). Program evaluation models and related theories: AMEE Guide no. 67. *Medical Teacher*, 34(5), e288-299.
- Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. Jossey-Bass.
- Gandomkar, R. (2018). Comparing Kirkpatrick's original and new model with CIPP evaluation model. *Journal of Advances in Medical Education & Professionalism*, 6(2), 94-95.
- Grossman, R., & Salas, E. (2011). The transfer of training: What really matters? *International Journal of Training and Development*, 15(2), 103-120.
- Hakan, K., & Seval, F. (2011). CIPP evaluation model scale: Development, reliability and validity. *Procedia-Social and Behavioral Sciences*, 15, 592-599.
- Hayes, H., Scott, V., Abraczinskas, M., Scaccia, J., Stout, S., & Wandersman, A. (2016). A formative multi-method approach to evaluating training. *Evaluation and Program Planning*, 58, 199-207.
- Hoc, J. M. (2001). Towards ecological validity of research in cognitive ergonomics. *Theoretical Issues in Ergonomics Science*, 2(3), 278-288.
- Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach' and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, 11, 721.
- Kalu, M. E., & Norman, K. E. (2018). Step by step process from logic model to case study method as an approach to educational programme evaluation. *Global Journal of Educational Research*, 17(1), 73-85.

- Kellaghan, T., Stufflebeam, D. L., & Wingate, L. A. (2003). *The international handbook of educational evaluation*. Dordrecht: Kluwer.
- Kirkpatrick, D. L. (2009). *Implementing the four levels: A practical guide for effective evaluation of training programs*. Easyread Large Edition. ReadHowYouWant.com.
- Lee, H. J., & Cha, Y. J. (2015). A study on the effectiveness of privacy education using the CIPP model: Focusing on the perceptions of local government officials. *The Korean Journal of Local Government Studies*, 19(1), 95-119.
- Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, 13(4), Article 6.
- McLemore, A. (2009). The CIPP model. *American Chronicle*, 23(3), 35-41.
- Moreau, K. A. (2017). Has the new Kirkpatrick generation built a better hammer for our evaluation toolbox?. *Medical Teacher*, 39(9), 999-1001.
- Neyazi, N., Arab, M., Farzianpour, F., & Mahmoudi Majdabadi, M. (2016). Evaluation of selected faculties at Tehran University of Medical Sciences using CIPP model in students and graduates point of view. *Evaluation and Program Planning*, 59, 88-93.
- Park, S.-K., Lee, S.-H., Kim, T.-Y., Jun, H.-J., & Kim, T.-S. (2017). A performance evaluation of information security training in public sector. *Journal of Computer Virology and Hacking Techniques*, 13(4), 289-296.
- Peyton, D. J., & Scicchitano, M. (2017). Devil is in the details: Using logic models to investigate program process. *Evaluation and Program Planning*, 65, 156-162.
- Reio, T. G., Rocco, T. S., Smith, D. H., & Chang, E. (2017). A critique of Kirkpatrick's evaluation model. *New Horizons in Adult Education and Human Resource Development*, 29(2), 35-53.
- Singh, R., & Tandon, P. (2016). User values based evaluation model to assess product universality. *International Journal of Industrial Ergonomics*, 55, 46-59.
- Sitzmann, T., & Weinhardt, J. M. (2018). Training engagement theory: A multilevel perspective on the effectiveness of work-related training. *Journal of Management*, 44(2), 732-756.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco: Jossey-Bass.
- Stufflebeam, D. L., & Zhang, G. (2017). *The CIPP evaluation model: How to evaluate for improvement and accountability*. Guilford Publications.
- Yardley, S., & Dornan, T. (2012). Kirkpatrick's levels and education 'evidence'. *Medical Education*, 46(1), 97-106.
- Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. (2011). Using the context, input, process, and product evaluation model (CIPP) as a comprehensive framework to guide the planning, implementation, and assessment of service-learning programs. *Journal of Higher Education Outreach and Engagement*, 15(4), 57-84.