

دانشگاه فرهنگیان
فصلنامه علمی تخصصی
پژوهش در آموزش مطالعات اجتماعی
دوره سوم، شماره چهارم، زمستان ۱۴۰۰

روش‌های سنجش در آموزش جغرافیا: یک مرور نظام‌مند

ارسال: ۱۴۰۰/۲/۲۵ علی خاکی، امجدی مهدوی^۲
پذیرش: ۱۴۰۰/۶/۲۰

چکیده

امروزه، مقالات متعددی در مورد ارزیابی در آموزش جغرافیا تدوین و نگارش شده است. با این حال، این مقالات از نظر توصیه‌ها و نتیجه‌گیری‌ها بر اساس شواهد و یافته‌های تحقیق متفاوت هستند. در سطح جهان، عمل مبتنی بر شواهد در اولویت است و درک دقیق آنچه تحقیقات تجربی پیرامون این موضوع می‌گوید، ضروری است. مرور نظام‌مند، یک روش دقیق را برای دستیابی به چنین امری ارائه می‌دهد. در پژوهش حاضر، مقالاتی که به ارزیابی در آموزش جغرافیا اختصاص یافته‌اند، مورد بررسی قرار گرفته است. روش مورد استفاده در این مطالعه، مروری و بهره‌مندی از منابع کتابخانه‌ای بود. بنابراین، در ابتدا مروری بر منابع کتابخانه‌ای و تحقیقاتی در زمینه روش‌های ارزیابی در آموزش جغرافیا، به عمل آمد. تمامی منابع مورد استفاده در این مقاله از موتور جستجوی گوگل، پایگاه‌های علمی ساینس دایرکت، گوگل اسکولار، وایلی آنلاین و الزویر جستجو شد. نتیجه پژوهش حاضر، این بود که دقت در موارد زیر لازم و ضروری است: (۱) دانش و مهارت‌های اساسی جغرافیایی که دانش‌آموزان باید رشد دهند؛ (۲) ماهیت پیشرفت یادگیری در هر یک از این زمینه‌ها و (۳) انواع روش‌های ارزیابی که معیارهای معتبر و قابل اعتمادی از این پیشرفت ارائه دهند.

واژگان کلیدی: ارزیابی آموزش جغرافیا، دانش جغرافیایی، بررسی سیستماتیک.

۱ - استادیار گروه معماری، دانشکده انقلاب اسلامی، دانشگاه فنی و حرفه‌ای، تهران، ایران

۲ - دکتری تخصصی روانشناسی و آموزش کودکان استثنایی، دانشگاه تهران، تهران، ایران

مقدمه

در بین محققان اتفاق نظر کلی وجود دارد که ارزیابی تکوینی و تراکمی برای بهبود آموزش جغرافیا ضروری است (لمبرت، ۲۰۱۱، ویدن و لمبرت، ۲۰۰۶). به‌طور گسترده‌تر، نیوتن (۲۰۰۷) معتقد است که هدف اصلی ارزیابی، ایجاد پیشرفت در یادگیری دانش‌آموزان با ارائه داده‌های معتبر، قابل‌اعتماد و به‌موقع جهت اطلاع ذی‌نفعان اصلی از جمله معلمان و سیاست‌گذاران است. برای نیل به این مقصود، در مورد بهترین روش‌ها برای ارزیابی اطلاعات یادگیری دانش‌آموزان در این حوزه، به‌وضوح نیاز است. درحالی‌که ما تصدیق می‌کنیم که مطالعات، برخی جنبه‌های ارزیابی را در آموزش جغرافیا موردبررسی قرار داده‌اند، اما یک بررسی سیستماتیک از تحقیقات تجربی ضروری است.

بیش از ۷۰۰ مقاله در مورد ارزیابی در آموزش جغرافیا وجود دارد که می‌توان آن‌ها را بر اساس داده‌های اولیه جمع‌آوری شده در زمینه جغرافیایی به مقالات تجربی و غیرتجربی تقسیم کرد. موضوعات مطرح‌شده در مقالات غیرتجربی شامل ارزیابی برای یادگیری (لمبرت، ۲۰۱۱؛ ویدن و لمبرت، ۲۰۰۶)؛ ارزیابی کنترل‌شده (آرتون، ۲۰۱۱)؛ پیشرفت‌های یادگیری در جغرافیا (بنتس، ۲۰۰۵ الف، ۲۰۰۵ ب؛ ادلسون، شاولسون و ورتهایم، ۲۰۱۳؛ سولم، هویانه و بوهم، ۲۰۱۵) و معیارهای موفقیت (جورج، کلارک، دیویس و دوربین، ۲۰۰۲؛ مارچلو، ۲۰۰۹؛ اسمورز، ۲۰۰۲) است. مقالات، سایر موارد ارزیابی کلی در آموزش جغرافیا را بررسی کرده‌اند (به بتیس، ۲۰۰۱؛ ویدن، ۲۰۱۳ مراجعه کنید). گزارش‌های عمده‌ای نیز منتشر شده است که جدیدترین آن‌ها شامل یک نقشه راه برای آموزش جغرافیای قرن ۲۱ است: ارزیابی (ادلسون و همکاران، ۲۰۱۳) که بررسی می‌کند چگونه تغییرات در روش‌های ارزیابی می‌تواند از تلاش‌ها برای بهبود آموزش جغرافیا در سطح مدرسه پشتیبانی کند. این انتشارات و گزارش‌ها سهم مهمی در درک ما از اصول کلی ارزیابی دارند. با این حال، با تمرکز جهانی بر روی عمل مبتنی بر شواهد، درک دقیق آنچه پژوهش تجربی پیرامون این موضوع می‌گوید ضروری است؛ بنابراین، در این مقاله به‌طور سیستماتیک تحقیقات تجربی ارزیابی در آموزش جغرافیایی، به‌منظور برجسته‌کردن شکاف‌ها و زمینه‌های تحقیق آینده بررسی می‌شود.

پس از توضیح روش دقیق این مطالعه از جمله استراتژی جستجو، معیارها و رویکرد تحقیق، خلاصه‌ای از مشخصات عمومی تحقیق شامل ناشر، مکان، دامنه، سطح برنامه و طراحی تحقیق را ارائه می‌دهیم. سپس مباحث اصلی در پیشینه تحقیق بررسی می‌شود تا به دو سؤال اصلی بپردازد: (۱) درباره ارزیابی مبتنی بر شواهد در آموزش جغرافیا چه می‌دانیم؟ و (۲) چه شکاف‌هایی مشخص شده است که می‌تواند جهت‌گیری‌های آینده تحقیق را نشان دهد؟

روش تحقیق

روش مورد استفاده در این مطالعه، مروری و بهره‌مندی از منابع کتابخانه‌ای بود. در این مطالعه در ابتدا مروری بر منابع کتابخانه‌ای و تحقیقاتی در زمینه سنجش و ارزیابی در آموزش جغرافیا، به عمل آمد. تمامی منابع مورد استفاده در این مقاله از موتور جستجوی گوگل، پایگاه‌های علمی ساینس دایرکت، گوگل اسکولار، وایلی آنلاین و الزویر جستجو شد. برای جستجو از کلیدواژه‌های مختلفی از جمله «ارزیابی، سنجش، آزمون، جغرافیا، روش، مروری» استفاده شد. تمامی این کلیدواژه‌ها به دو زبان فارسی و انگلیسی بین سال‌های ۱۹۸۸ تا ۲۰۲۱ مورد جستجو قرار گرفت و از بین ۴۵ مقاله، کتاب و پایان‌نامه استخراج شده، ۳۰ منبع مورد استفاده قرار گرفت و اطلاعات لازم در مورد مبانی نظری و پژوهشی روش‌های ارزیابی و سنجش در جغرافیا گردآوری شد. معیارهای ورود، از نظر نوع، مقالات علمی-پژوهشی، ISI، رساله‌های دکتری، پایان‌نامه‌های کارشناسی ارشد، کنفرانس‌ها و کتاب‌ها، از نظر زمان بین سال‌های ۱۹۸۸ تا ۲۰۲۱، از نظر طرح پژوهش، تحقیقات کیفی، کمی و آمیخته، از نظر منطقه جغرافیایی، تمام جهان (به زبان‌های فارسی و انگلیسی)، از نظر زمینه نیز دانشگاهی و غیردانشگاهی بودند. معیارهای خروج نیز سرمقاله مجلات، آیین‌نامه‌ها و اسناد، از نظر زمان قبل از سال ۱۹۸۸، از نظر طرح پژوهش، مقالات و پژوهش‌هایی که فاقد روش‌های کمی، کیفی و آمیخته بودند، از نظر منطقه جغرافیایی مقالاتی غیر از زبان انگلیسی و فارسی بودند.

یافته‌های تحقیق

در سراسر مجموعه داده‌ها، هشت موضوع شناسایی شد: (۱) ارزیابی تکوینی؛ (۲) استدلال فضایی؛ (۳) استانداردهای عملکرد/ موفقیت؛ (۴) علامت‌گذاری؛ (۵) تجزیه و تحلیل محتوای موارد ارزیابی (همسویی با الزامات سیاست ملی)؛ (۶) ارزیابی بین‌المللی؛ (۷) سایر استراتژی‌های ارزیابی (نقشه‌برداری از طرح‌ها، ارزیابی خود) و (۸) مطالعات موردی و ارزیابی عملکرد ملی.

۱) ارزیابی تکوینی

پنج مقاله بازیابی شده منحصراً به ارزیابی برای یادگیری (ارزیابی برای یادگیری) (ارزیابی تکوینی) در جغرافیا متمرکز شده است. تمام مطالعات در مدارس متوسطه در جهان غرب در انگلیس انجام شده است. این مطالعات نقش بازخورد (بالدوین، ۲۰۰۳)، همکاری دانشگاه / مدرسه (کوپر و کاوی ۲۰۱۰)، ارزیابی خود (دیویس، دوربین، کلارک و دیل، ۲۰۰۴)، برنامه‌ریزی و اجرای ارزیابی برای یادگیری (لیت و نیکولز، ۲۰۰۰؛ تیکناز و ساتن، ۲۰۰۶) را مورد بررسی قرار داده‌اند. سه مورد از مطالعات به‌عنوان رویکرد تحلیلی از تجزیه و تحلیل موضوعی استفاده کرده‌اند، یک مورد مطالعه موردی و دیگری از تکنیک‌های مرتبط با قوم‌نگاری. انواع روش‌های جمع‌آوری داده‌ها متن‌آزمون، مصاحبه، مشاهدات کلاس، صدا یا ویدئوی ضبط شده، تأمل، نظرسنجی و صورت‌جلسات بود. قابل‌ذکر است که هیچ‌یک از این مطالعات در مدارس ابتدایی انجام نشده است و فقط یک مورد از اقدامات کمی استفاده کرده است. در مجموع، مطالعات نشان داد که ارزیابی برای یادگیری می‌تواند تأثیر قابل‌توجهی بر نتایج دانش‌آموزان در جغرافیا داشته باشد. به‌عنوان مثال، کوپر و کاوی (۲۰۱۰) دریافتند که ارزیابی برای یادگیری انگیزه دانش‌آموزان و استقلال یادگیری را افزایش می‌دهد. علاوه بر این، دانش‌آموزان با درک معیارهای ارزیابی می‌توانند به‌طور مؤثرتری کیفیت را بیان کنند و بازخورد معلم را برای بهبود نمرات خود تفسیر کنند (دیویس و همکاران، ۲۰۰۴؛ لیت و نیکولز، ۲۰۰۰). مشخص شد که فعالیت‌های ارزیابی برای یادگیری با پشتیبانی خارجی و آزمایش حرفه‌ای سیستماتیک پایدارتر هستند (کوپر و کووی، ۲۰۱۰). تیکناز و ساتن (۲۰۰۶) همچنین اهمیت الزامات قانونی و دانش حرفه‌ای معلمان را در اجرای ارزیابی برای یادگیری در کلاس‌های جغرافیایی برجسته

کردند. به‌عنوان مثال بالدوین (۲۰۰۳) دریافت باوجوداینکه معلمان در مطالعه وی (۶ نفر) از اهمیت ارزیابی تکوینی آگاه بودند، از نظرات کمی استفاده کردند و به‌طور کلی از رویکردهای «پرورش به‌پیش» مطابق با استانداردهای پیشرفت استفاده نمی‌کردند. بالدوین (۲۰۰۳) بر لزوم اسناد سیاست‌گذاری در نیویورک برای گنجاندن نمونه‌هایی از روش ارزیابی تکوینی تأکید کرد.

۲) استدلال فضایی

پنج مقاله در رابطه با استدلال فضایی وجود داشت: چهار مورد در مدارس متوسطه یا دانشگاه و یک مورد در مدرسه ابتدایی. چهار مورد از پنج مطالعه با استفاده از روش کمی شامل جمع‌آوری و تجزیه و تحلیل آماری داده‌های آزمون (بتزبی، گالچ و مارش، ۲۰۰۶؛ هیونا و شارپ، ۲۰۱۳؛ لی و بدنارز، ۲۰۱۲؛ تومازوفسکی، وودایک، پارودی و هلت، ۲۰۱۵) اجرا شد. این مقالات بر پیشرفت دانش فضایی اتفاقی (باترسی و همکاران، ۲۰۰۶) و اعتبارسنجی یک ابزار ارزیابی (STAT) متمرکز بودند (لی و بدنارز، ۲۰۱۲؛ تومازوفسکی و دیگران، ۲۰۱۵). مقاله دیگر از رویکرد کیفی شامل تحلیل اسناد برای بررسی میزان تمرین مفاهیم تفکر فضایی در مدارس ایالات متحده استفاده کرد (آنتهماتن، ۲۰۱۰). هر پنج مقاله مستقیماً با آمریکای شمالی در ارتباط بودند. باین‌حال، مطالعه نهایی شامل آزمایش یک ابزار آمریکایی در رواندا بود.

اولین مقاله در این موضوع (باترسی و همکاران، ۲۰۰۶) پیشرفت دانش جغرافیایی دانش‌آموزان را هنگام بلوغ از دوره راهنمایی تا دانشگاه بررسی کرد. باترسی و همکاران (۲۰۰۶) از نظر دستیابی به مفاهیم خاص فضایی یک سلسله‌مراتب مشخص کشف کردند. آن‌ها نتیجه گرفتند که بدون دستورالعمل‌های هدفمند، درک مفهوم هم‌پوشانی نقشه برای اکثر دانش‌آموزان دوره راهنمایی بسیار دشوار است. به همین دلیل، باترسی و همکاران (۲۰۰۶) در مورد پایه‌ریزی شالوده‌ای برای یادگیری مفاهیم مکانی در دوره ابتدایی بحث کردند. باین‌وجود، آنتامتن (۲۰۱۰) دریافت که بسیاری از مهارت‌های اصلی تفکر فضایی در مطالعات اجتماعی دولت و اسناد استاندارد

جغرافیایی وجود ندارد؛ بنابراین، راهنمایی معلمان در این زمینه هدایت نمی‌شدند. وی از مؤلفان استانداردها درخواست کرد تا از تحقیقات در این زمینه مطلع باشند. سه مطالعه دیگر در این زمینه، توسعه آزمون استاندارد برای توانایی تفکر فضایی هستند (هیونا و شارپ، ۲۰۱۳؛ لی و بدنارز، ۲۰۱۲؛ تومازوفسکی و همکاران، ۲۰۱۵). لی و بدنارز (۲۰۱۲) آزمون توانایی تفکر فضایی (STAT) را توسعه دادند و قابلیت اطمینان و روایی آن را در ارزیابی مهارت تفکر فضایی ۵۳۲ دانش‌آموز و دانشجو در ایالات متحده نشان دادند. تومازوفسکی و همکاران (۲۰۱۵) ابزار توانایی تفکر فضایی را برای زمینه فرهنگی رواندا اصلاح کردند و دریافتند که دانش‌آموزان و دانشجویان شهری و پسر به ترتیب در رابطه با مهارت تفکر فضایی از دانش‌آموزان و دانشجویان روستایی و زن بهتر عمل می‌کنند. هیونا و شارپ (۲۰۱۳) بر اساس کارهای لی و بدنارز (۲۰۱۲) و باترسی و همکاران (۲۰۰۶) با ایجاد یک آزمون معتبر و قابل اعتماد از تفکر فضایی، مربیان را قادر ساختند عملکرد دانش‌آموزان و سطح درک آن‌ها را در سطوح مختلف محک بزنند. یافته‌های آن‌ها تأیید کرد که در استفاده از مفاهیم فضایی یک درجه سختی وجود دارد و تفکر فضایی دارای ابعاد مختلفی از جمله تجزیه و تحلیل، درک مطلب، نمایش، کاربرد، مقیاس و رابطه فضایی است.

۳) استانداردهای عملکرد و موفقیت

دو مقاله سعی در ایجاد استانداردهای پیشرفت برای دانش‌آموزان متوسطه داشته است. با این حال، هر یک روش متفاوتی را در پیش گرفتند. در حالی که دیویس (۲۰۰۲) از متون امتحانی برای تعیین میزان پیچیدگی پاسخ دانش‌آموزان، کیفیت رتبه‌بندی با توجه به ویژگی، کامل بودن و توانایی قضاوت مستدل استفاده کرد، لاک و بارت (۲۰۰۲) قبل از پیشنهاد و اعتبارسنجی مقیاس موفقیت، نتایج یادگیری را در اسناد برنامه درسی موجود طبقه‌بندی کرد. هر دو مطالعه با روش ترکیبی در مدارس متوسطه انجام شده است. هر دو مقاله از توسعه و استفاده از استانداردهای موفقیت در جغرافیا حمایت می‌کنند اما معتقدند که تحقیقات بیشتری لازم است.

۴) نشانه‌گذاری

دو مقاله در این زمینه قرار گرفت: مقاله کریسپ (۲۰۰۸) و پانادرو، تاپیا و هورتاس (۲۰۱۲). در حالی که هر دو مطالعه در محیط‌های ثانویه چند جایگاهی انجام شده‌اند، هر دو بر ذی‌نفعان مختلف در فرآیند علامت‌گذاری، ارزیابان و دانش‌آموزان متمرکز شده‌اند. بعلاوه، هر دو مطالعه از روش‌های مختلفی استفاده کردند. کریسپ (۲۰۰۸) برای کشف روندهای روان‌شناختی اساسی در علامت‌گذاری متون آزمون جغرافیا سطح AS1 و A2 در انگلستان، از پروتکل‌های «بلندبلند فکر کردن» و «تجزیه و تحلیل آماری» استفاده کرد. یافته‌ها حاکی از آن بود که ارزیابی‌های مثبت، مقایسه و خواندن دقیق برای جلوگیری از شدت علامت‌گذاری و مشکلات احتمالی مهم بودند. از طرف دیگر، مطالعه پانادرو و همکاران (۲۰۱۲) بر رابطه بین خودارزیابی دانش‌آموزان (شامل استفاده از سرفصل‌ها و نمونه متون)، خودتنظیمی و خودکارآمدی متمرکز بود. یافته‌ها نشان داد که ابزارهای خودارزیابی، سطوح بالاتری از خودتنظیمی را به بار می‌آورند و در این میان، متون از سرفصل‌ها عملکرد بهتری داشتند. با این حال، وقتی نوبت به خودکارآمدی رسید، نویسندگان ادعا کردند که چنین ابزارهایی برای ارزیابی تسلط بر یادگیری کافی نیستند.

۵) تحلیل محتوای موارد ارزیابی: همسویی با الزامات سیاست ملی

یکی دیگر از موضوعات مشترک در تحقیقات تجربی، همسویی موارد ارزیابی با سیاست ملی از جمله استانداردهای ملی، برنامه درسی و مشخصات امتحان بود. سه مطالعه که در کشورهای چین (یانگ، ۲۰۱۳)، ترکیه (یاشار، ۲۰۰۹) و هند (میشرا، ۲۰۱۵) انجام شدند، ماهیت سنجش در کتاب‌های درسی را بررسی کردند. دو مقاله دیگر (بیتربوس، ون در شی و کوپر، ۲۰۱۷؛ ورتهایم و همکاران، ۲۰۱۳) در این موضوع بر آزمون‌های استاندارد مقیاس بزرگ تمرکز داشتند. هر پنج مقاله تقاضای شناختی و نوع دانش موردنظر را بررسی کردند.

دو مقاله اول در این زمینه تغییرات سؤالات کتاب درسی را در پاسخ به اصلاحات آموزشی ملی بررسی کرده است. در مورد چین، این امر به اجرای استانداردهای ملی ۲۰۰۱ و در ترکیه تغییر در برنامه درسی جغرافیای آموزش متوسطه (۲۰۰۵) اشاره داشت. این مطالعه چینی بر روی نوع سؤال، تقاضای شناختی و ماهیت سیاسی اقلام

متمرکز بود. این مطالعه ترکیبی بررسی کرد که آیا موارد ارزیابی اصلاح برنامه درسی را منعکس می‌کنند؛ یعنی تغییر به سمت ارزیابی فرایندهای یادگیری و نه صرفاً محصول. به‌طور کلی، هر دو مطالعه نتیجه گرفتند که موارد ارزیابی در کتاب‌های درسی مؤثرتر شده است. برای چینی‌ها، انعطاف‌پذیری بیشتری در تولید کتاب‌های درسی با تأکید بیشتر بر تفکر عالی و حرکت به سمت سؤالات کمتر سیاسی وجود دارد. این یک داستان مشابه در ترکیه بود که دارای موارد ارزیابی متنوع‌تر و با کیفیت بهتر طراحی تفکر مرتبه بالاتر است. مقاله سوم در این موضوع، پژوهش میشرا (۲۰۱۵) است که سؤالات کتاب درسی جغرافیای هند را در رابطه با چارچوب برنامه درسی ملی (NCF) با استفاده از طبقه‌بندی تفکر فضایی جو و بدنارز (۲۰۰۹) بررسی کرد. برخلاف مطالعات فوق، میشرا (۲۰۱۵) دریافت که سؤالات کتاب‌های جغرافیایی سال‌های نهم و دهم تمرکز بر حل واقعی مسئله ندارند، مهارت‌های تفکر مرتبه بالاتر را ارزیابی نمی‌کنند و رویکرد منطقه‌ای جغرافیا را از دهه ۱۹۶۰ که محدود بود، منعکس می‌کنند.

مطالعه بیجستربوش و همکاران (۲۰۱۷) یافته‌های مشابه پژوهش میشرا (۲۰۱۵) را نشان داد. باین‌حال، زمینه متفاوت بود. آن‌ها در تجزیه و تحلیل ۳۹ امتحان داخلی در ۱۳ مدرسه دریافتند که بیشتر موارد آزمون سطوح پایین‌تری از تقاضای شناختی (عمدتاً یادآوری و تا حدود کمتری فهم) و دانش مفهومی و واقعی را هدف قرار می‌دهند. سؤالات بیش از آنکه به دانش‌آموزان اجازه دهد کاربرد مهارت‌های جغرافیایی و دانش مدل‌ها و تئوری‌های کلیدی را نشان دهند، بیشتر بر روی یادگیری سطحی و خام متمرکز بود.

ورتهایم و همکاران (۲۰۱۳) ماهیت موارد ارزیابی را در آزمون‌های استاندارد مقیاس بزرگ و ارزیابی کلاس‌ها در کلاس‌های درسی مدارس ایالات متحده بررسی کردند. آن‌ها نتیجه گرفتند که شیوه‌های ارزیابی در ایالات متحده عمدتاً بر یادآوری دانش واقعی و مفهومی متمرکز است و تعداد کمی از موارد، توانایی دانش‌آموزان را برای نشان دادن شیوه‌های جغرافیایی از جمله استدلال مبتنی بر شواهد ارزیابی می‌کنند.

۶) ارزیابی بین‌المللی

در رابطه با ارزیابی بین‌المللی، ما سه مقاله را شناسایی کردیم، یعنی ون در شی و کلکمن (۲۰۱۰) و لین و بورک (۲۰۱۶ الف، ۲۰۱۶ ب). مقاله اول ماهیت ارزیابی‌های ملی موجود (المپیاد بین‌المللی جغرافیا) را توصیف کرده و نتایج آزمون چندرسانه‌ای را با استفاده از همبستگی پیرسون تجزیه و تحلیل کرده است. این مطالعه همبستگی شدیدی بین سه بخش المپیاد ۲۰۰۸ و نمره کل را نشان داد. نویسندگان نتیجه گرفتند که المپیاد بین‌المللی جغرافیا ارزیابی معتبر و قابل‌اعتمادی است که می‌تواند برای ایجاد معیار بین‌المللی سواد جغرافیایی مورد استفاده قرار گیرد.

دو مقاله کیفی دیگر یافته‌های یک نظرسنجی از ۷۴ ذی‌نفع بین‌المللی از ۳۶ کشور را گزارش و بررسی کردند: (۱) ادراکات انجمن آموزش بین‌المللی جغرافیا در مورد مزایا و چالش‌های آغاز و اجرای ارزیابی جغرافیا در TIMSS و (۲) آیا از نظر آنچه باید ارزیابی شود و نحوه اجرای آزمون می‌تواند اتفاق نظری وجود داشته باشد. در مقاله اول، محققان دریافتند که انجمن آموزش جغرافیا به دلیل ارزش داده‌های به‌دست‌آمده برای تحقیق، سیاست‌گذاری و شیوه‌های آموزشی، اکثراً موافق چنین اقدامی است. تعدادی از سؤالات در مورد توسعه و اجرای این ارزیابی در مقاله دوم مطرح و بررسی شده است. نتایج نشان داد که انجمن جهانی آموزش جغرافیا در نظرات خود در مورد سن هدف برای ارزیابی بین‌المللی و مفاهیم و مهارت‌هایی که باید هدف قرار گیرد، اختلاف نظر دارد. یافته‌ها حاکی از آن است که امکان اجرای ارزیابی بین‌المللی از این نوع بدون ایجاد مجموعه‌ای مشترک از اهداف یادگیری برای دانش‌آموزان پایه چهارم و هشتم وجود ندارد. این مقاله اظهار داشت که این فرایند احتمالاً منجر به تغییراتی در برنامه درسی محلی می‌شود تا اطمینان حاصل شود که یک چارچوب مشترک برای ارزیابی وجود دارد.

۷) سایر استراتژی‌های ارزیابی

این مضمون شامل ترکیبی از روش‌ها برای ارزیابی درک جغرافیایی، یعنی نقشه‌برداری مفهومی، نقشه‌برداری طرح، نوشتن استدلالی و تداعی آزاد کلمه است. دو مقاله در این موضوع، اندروز، ترسلر و مینتز (۲۰۰۸) و وهری، مونرو اوسی، کاب و فونت (۲۰۱۲)، هر دو نقشه‌برداری مفهوم را به‌عنوان ابزاری برای ارزیابی مفهومی در جغرافیا

بررسی کردند. اندروز و همکاران مقاله (۲۰۰۸) بر دانش‌آموزان سال‌های پایه ششم تا نهم (۳۲۵ نفر) متمرکز شدند و یک طرح آزمایشی را اتخاذ کردند، درحالی‌که وهری و همکاران (۲۰۱۲) رویکرد کیفی را بر روی دانش‌آموزان پایه هفتم اجرا کردند (۴۳ نفر). هر دو مطالعه در مورد ارزش نقشه‌های مفهومی به‌عنوان یک گزینه جایگزین یا الحاقی برای آزمایش‌های مداد و کاغذ سنتی نتیجه‌گیری مشابهی داشتند. این مطالعات نشان داد که چگونه تهیه نقشه‌های مفهومی می‌تواند اطلاعات کیفی و کمی ارزشمندی در مورد عمق درک دانش‌آموزان فراهم کند (اندروز و همکاران، ۲۰۰۸) و منابعی را به معلمان ارائه دادند (وهری و همکاران، ۲۰۱۲). علاوه بر این، وهری و همکاران (۲۰۱۲) استدلال کردند که تکنیک‌های نقشه‌برداری مفهومی مهارت تفکر فضایی را افزایش می‌دهد و علاقه را به موضوع مورد مطالعه، در این مورد، جغرافیای انسانی تحریک می‌کند.

تحقیقات مرتبط با بررسی نمایش‌های بصری به‌عنوان ابزار ارزیابی در جغرافیا، شامل کارهای هاروود و راولینگر (۲۰۰۱) بود. این مطالعه شامل بررسی ۲۶ طرح نقشه آزاد دانش‌آموزان ابتدایی جهان برای ارزیابی درک آن‌ها از آرایش فضایی، اندازه و شکل قاره‌ها بود. نتایج، ارزش این روش را برای شناسایی سوءبرداشت‌ها و نقاط کور در درک دانش‌آموزان برجسته کرد. علاوه بر این، نویسندگان استدلال کردند که روند ساخت طرح‌های آزاد و دریافت بازخورد از معلمان و همسالان می‌تواند دانش جهانی دانش‌آموزان ابتدایی را بهبود بخشد.

محققان مختلف استفاده از مقاله‌ها و نوشتن استدلال را به‌عنوان روش‌های ارزیابی درک جغرافیایی بررسی کرده‌اند. دو مقاله در این رابطه، شیفل و اولنویکل (۲۰۱۰) و مونونویو (۲۰۰۷) بود. هر دو مطالعه در مقطع متوسطه اجرا شدند اما از روش‌های مختلفی استفاده کردند. مطالعه بود که و همکاران (۲۰۱۰) آلمان تعداد و نوع استدلال‌های استفاده‌شده توسط دانش‌آموزان در یک کلاس را محاسبه کرد. ابتدا متن‌ها با استفاده از نمودار استدلال (ایجاد نمودارها برای نشان دادن چگونگی ارتباط قسمت‌های استدلال با یکدیگر) بررسی شد و سپس با بررسی ساختار استدلال‌ها با استفاده از مدل تولمین مورد تجزیه و تحلیل قرار گرفتند. مطالعه مونونویو (۲۰۰۷) شامل ۸۳ دانش‌آموز در سه مدرسه در زیمبابوه بود. مونونویو برای تعیین اثربخشی

نسبی کلاس در مقابل دستورالعمل‌های کار میدانی، ساختار طبقه‌بندی نتیجه یادگیری مشاهده‌شده (SOLO) و یک طرح مارک گذاری را ایجاد کرد که توسط لمبرت (۱۹۹۶) انجام شد. تجزیه و تحلیل نویسنده اثربخشی بررسی ساختار و پیچیدگی پاسخ‌های مقاله را به‌عنوان معیاری از درک نشان داد.

۸ مطالعات موردی و ارزیابی ملی عملکرد

دو مقاله در این موضوع، پینار (۲۰۱۱) و بات، ویدن، چاب و روکوز (۲۰۰۶) دیدگاه مربیان در مورد ارزیابی در کشورهای مربوطه خود (ترکیه و انگلیس) را ارائه دادند. مطالعه کیفی ترکی با حضور ۱۰ معلم در چندین سایت نشان داد که معلمان جغرافیا نسبت به تکالیف حل مساله و مبتنی بر فرآیند که در اصلاح برنامه درسی ۲۰۰۵ مهم بودند، مقاومت نشان می‌دهند. معلمان شرکت‌کننده از نظر دانش و مهارت محتوا احساس عدم کفایت می‌کردند و مرتباً به روش‌های معمول و سنتی خود روی می‌آوردند. این مطالعه در انگلیس بسیار گسترده‌تر بود و گزارشی از آموزش و ارزیابی جغرافیا در ۱۷ مدرسه متوسطه ایالتی انگلیس را ارائه می‌داد که «دریچه‌ای برای تمرین و عملکرد در ارزیابی در جغرافیا» ارائه می‌دهد (بات و همکاران، ۲۰۰۶، ص ۱۳۵) در میان طیف وسیعی از یافته‌ها در مورد مشخصات جغرافیایی گواهینامه عمومی در آموزش متوسطه (GCSE)، برنامه‌ریزی درسی و آموزش، نویسندگان نتیجه گرفتند که معلمان جغرافیا فهم جامعی از اهمیت ارزیابی تکوینی و تعیین هدف برای به حداکثر رساندن نتایج یادگیری دانش‌آموزان ندارند. هر دو مقاله بر نیاز ذی‌نفعان اصلی برای درک منطق اصلاحات آموزشی و توسعه حرفه‌ای با دانش و مهارت‌های مناسب برای اجرا تأکید داشت. این امر با نتیجه‌گیری کوپر و کاوی (۲۰۱۰) در مورد ویژگی‌های اصلاحات پایدار مطابقت دارد (نگاه کنید به موضوع ۱).

بحث و نتیجه‌گیری

دو سؤال مهم تحقیقاتی برای این مطالعه عبارت‌اند از: (۱) در مورد ارزیابی مبتنی بر شواهد در آموزش جغرافیا چه می‌دانیم؟ و (۲) چه شکاف‌هایی شناسایی شده است که به جهت‌های تحقیقاتی آتی اشاره می‌کند؟ در پاسخ به سؤال اول، این بررسی ۳۰ مقاله تجربی و بررسی‌شده را منتشر کرد که طی ۱۶ سال گذشته به زبان انگلیسی

منتشر شده‌اند و تمرکز آن‌ها بر ارزیابی در آموزش جغرافیای مدارس است. این بررسی نشان می‌دهد که ما در مورد هشت موضوع که در بالا ذکر شد، دانش ارزیابی در آموزش جغرافیا را داریم. در این مضامین چندین گرایش متمایز شناسایی شده و شایسته شرح بیشتر است.

در اکثر مطالعات، دوره‌های جمع‌آوری داده‌ها در این مطالعات ارزیابی آموزش جغرافیایی کوتاه بود و هیچ مطالعه طولی در این زمینه اجرا نشد. علاوه بر این، بیشتر مطالعات در مقیاس‌های کوچک با بافت محلی و ملی (عمدتاً مستقر در انگلستان و ایالات متحده) اجرا شده بود. علاوه بر این، بین محققانی که سؤالات تحقیق مشابهی را در مکان‌ها یا زمینه‌های مختلف کاوش می‌کردند، ارتباطی وجود نداشت. به‌عنوان مثال، از پنج مقاله موجود در موضوع ارزیابی برای یادگیری، فقط یک مقاله (دیویس و همکاران، ۲۰۰۴) به هر یک از مقالات قبلی در زمینه ارزیابی جغرافیایی استناد کرده است؛ بنابراین درحالی که تحقیقاتی به ارزیابی در آموزش جغرافیایی اختصاص داده شده بود، به نظر نمی‌رسید که یک جهت منسجم یا همکاری در سطح جهانی پیرامون این مقالات وجود داشته باشد.

روند بعدی که در پیشینه تحقیقاتی آشکار بود، فقدان تحقیق در حوزه توسعه و اعتبارسنجی ابزار برای ارزیابی ادراکات و مهارت‌های کلیدی جغرافیایی بود. استثنائات کلیدی پژوهش لی و بدنارز (۲۰۱۲) و هوینه و شارپ (۲۰۱۳) بود که ابزارهایی را برای ارزیابی استدلال فضایی تهیه و تأیید کرده بودند. این مطالعات نشان‌دهنده دو مقاله از تعداد محدود مقالات ارزیابی در آموزش جغرافیا است. مطالعه رواندا توسط تومازوفسکی و همکاران (۲۰۱۵) از یک روش ترکیبی استفاده کرده است. این بررسی همچنین برخی موارد همسویی با شیوه‌های ارزیابی را در آموزش جغرافیا برجسته کرد. این تحقیق حاکی از آن است درحالی که در استانداردهای آموزشی و اسناد برنامه درسی در سطح بین‌المللی تغییراتی صورت گرفته است، اما روش‌های ارزشیابی همیشه متناسب با این تغییرات نبوده‌اند. به‌عنوان مثال، در مطالعه میشر (۲۰۱۵) عدم تطابق بین الزامات چارچوب برنامه درسی ملی و ماهیت سؤالات کتاب درسی جغرافیا وجود داشت. این امر با تحقیقات انجام شده توسط بیتربوش و همکاران (۲۰۱۷) در هلند

مطابقت داشت که دریافت، ارزیابی‌های داخلی فاقد تمرکز بر دانش پیچیده و فرایندهای شناختی است که در برنامه‌های درسی ذکر شده است.

در پاسخ به سؤال دوم تحقیق، در مورد مؤلفه‌های سواد جغرافیایی و توسعه و اعتبارسنجی ابزار ارزیابی این موارد، نیاز به اتفاق نظر است. اگرچه مطالعات در مقیاس کوچک می‌توانند در زمینه‌های محلی آموزنده باشند، اما نیاز به مطالعات طولی در مقیاس بزرگ با محققان بین‌المللی است که به‌طور مشترک روی یک سؤال تحقیقاتی توافق شده کار کنند. منشور بین‌المللی آموزش جغرافیایی (۲۰۱۶) نشان می‌دهد که یکی از این سؤالات باید شامل چگونگی درک پیشرفت‌های یادگیری دانش‌آموزان در جغرافیا و نحوه کالیبره و ارزیابی بهتر این پیشرفت‌ها باشد. برای افزایش اعتبار استنباط از تحقیقات، طراحی این مطالعات باید شامل روش‌های کیفی، کمی و مختلط باشد. همچنین واضح است که تحقیق در مورد ارزیابی مبتنی بر شواهد از درک جغرافیایی در زمینه‌های دبستان موردنیاز است. جنبش‌های اخیر در فضای آموزش جغرافیایی، برخی از موارد مشخص شده در بالا را مرتفع می‌سازد. به‌عنوان مثال، مرکز ملی تحقیقات در آموزش جغرافیا (NCRGE) که توسط بنیاد ملی علوم در ایالات متحده پشتیبانی می‌شود، هماهنگی تحقیق، همکاری و اشتراک اطلاعات بین جغرافی‌دانان و محققان آموزشی در سایر رشته‌ها را تسهیل می‌کند.

به‌طور خلاصه، این بررسی سیستماتیک، کمبود تحقیق در مورد ارزیابی در آموزش جغرافیا را تأیید می‌کند و یک رویکرد ساختاری ارائه می‌دهد که می‌تواند برای شناسایی شکاف‌های موجود تکرار شود. ما به‌عنوان یک جامعه یادگیری و تحقیقاتی، باید در مورد دانش و مهارت‌های اساسی دانش‌آموزان، ماهیت پیشرفت‌های یادگیری مرتبط با این دانش و مهارت‌ها و انواع روش‌های ارزیابی که اقدامات معتبری را ارائه می‌دهند، بیشتر بدانیم. این امر یک راهنما و خط‌مشی تحقیقاتی در مسئله و موضوع سنجش در آموزش جغرافیا است که باید بدان توجه ویژه‌ای معطوف شود.

منابع

- Akerlind, G. (2002, November). *Principles and practice in phenomenographic research*. Paper presented at the Current Issues in Phenomenography Conference, Australian National University, Canberra.
- Andrews, K. E., Tressler, K. D., & Mintzes, J. J. (2008). *Assessing environmental understanding: An application of the concept mapping strategy*. *Environmental Education Research*, 14(5), 519–536.
- Anthamatten, P. (2010). *Spatial thinking concepts in early grade-level geography standards*. *Journal of Geography*, 109(5), 169–180.
- Atasoy, E. (2009). *Eighth grade students' associations of geographical and environmental concepts*. *New Educational Review*, 17(1), 102–124.
- Atherton, R. (2011). *Controlled assessment: Getting it right?* *Teaching Geography*, 36(1), 26–27.
- Baldwin, R. (2003). *Teachers' use of feedback comments to improve student achievement in standards-based assessment*. *International Journal of Learning*, 10, 1197–1208.
- Battersby, S. E., Golledge, R. G., & Marsh, M. J. (2006). *Incidental learning of geospatial concepts across grade levels: Map overlay*. *Journal of Geography*, 105(4), 139–146.
- Bennetts, T. (2005a). *Progression in geographical understanding*. *International Research in Geographical and Environmental Education*, 14(2), 112–132.
- Bennetts, T. (2005b). *The links between understanding, progression and assessment in the secondary geography curriculum*. *Geography*, 90(2), 152–170.
- Bettis, N. C. (2001). *Assessment issues in geographic education for the twenty-first century*. *Journal of Geography*, 100(4), 172.
- Bijsterbosch, E., van der Schee, J., & Kuiper, W. (2017). *Meaningful learning and summative assessment in geography education: An analysis in secondary education in the Netherlands*. *International Research in Geographical and Environmental Education*, 26(1), 17–35.

- Budke, A., Schiefele, U., & Uhlenwinkel, A. (2010). 'I think it's stupid' is no argument: Investigating how students argue in writing. *Teaching Geography*, 35(2), 66–69.
- Butt, G., Weeden, P., Chubb, S., & Srokosz, A. (2006). *The state of geography education in English secondary schools: An insight into practice and performance in assessment*. *International Research in Geographical and Environmental Education*, 15(2), 134–148.
- Cooper, B., & Cowie, B. (2010). *Collaborative research for assessment for learning*. *Teaching and Teacher Education*, 26(4), 979–986.
- Crisp, V. (2008). *Exploring the nature of examiner thinking during the process of examination marking*. *Cambridge Journal of Education*, 38(2), 247–264.
- Davies, P. (2002). *Levels of attainment in geography*. *Assessment in Education: Principles, Policy & Practice*, 9(2), 185–204.
- Edelson, D. C., Shavelson, R. J., & Wertheim, J. A. (2013). *A road map for 21st century geography education: Assessment*. Washington, DC: National Geographic Society.
- Ford, J. D., & Pearce, T. (2010). *What we know, do not know, and need to know about climate change vulnerability in the western Canadian Arctic: A systematic literature review*. *Environmental Research Letters*, 5(1), 1–9. doi:10.1088/1748-9326/5/1/014008
- George, J., Clarke, J., Davies, P., & Durbin, C. (2002). *Helping students to get better at geographical writing*. *Teaching Geography*, 27(4), 156–160.
- Harwood, D., & Rawlings, K. (2001). *Assessing young children's freehand sketch maps of the world*. *International Research in Geographical and Environmental Education*, 10(1), 20–45.
- Huynh, N. T., & Sharpe, B. (2013). *An assessment instrument to measure geospatial thinking expertise*. *Journal of Geography*, 112(1), 3–17.
- International Geographical Union Commission on Geographical Education (IGU-CGE)*. (2016). 2016 International charter on geographical education. Beijing: Author.
- Jo, I., & Bednarz, S. (2009). *Evaluating geography textbook questions from a spatial perspective: Using concepts of space*,

- tools of representation, and cognitive processes to evaluate spatiality.* Journal of Geography, 108(1), 4–13.
- Lambert, D. (2011). *The lie of the land (revisited).* Teaching Geography, 36(1), 24–25.
- Lane, R., & Bourke, T. (2016a). *Possibilities for an international assessment in geography.* International Research in Geographical and Environmental Education, 26(1), 71–85. doi:10.1080/10382046.2016.1165920
- Lane, R., & Bourke, T. (2016b). *The inclusion of geography in TIMSS: Can consensus be reached ?* International Research in Geographical and Environmental Education, 26(2), 166–176. doi:10.1080/10382046.2016.1178939
- Lee, J., & Bednarz, R. (2012). *Components of spatial thinking: Evidence from a spatial thinking ability test.* Journal of Geography, 111(1), 15–26.
- Lock, G., & Barrett, R. (2002). *Standards framework: Developing scales of achievement in postcompulsory education: A case study.* Issues in Educational Research, 12(1), 35–48.
- Marcello, J. S. (2009). *A proposal for assessment in geography education.* Journal of Geography, 108 (4–5), 226–232.
- Mishra, R. K. (2015). *Mapping the knowledge topography: A critical appraisal of geography textbook questions.* International Research in Geographical and Environmental Education, 24(2), 118–130.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M. ... PRISMA-P Group. (2015). *Preferred reporting items for systematic review and meta-analysis protocols (PRISMA) 2015 statement.* Systematic Reviews, 4(1). 1–9. doi:10.1186/2046-4053-4-1
- Munowenyu, E. (2007). *Assessing the quality of essays using the SOLO taxonomy: Effects of field and classroom-based experiences by 'A' level geography students.* International Research in Geographical and Environmental Education, 16(1), 21–43.
- Murray, S., Nuttall, J., & Mitchell, J. (2008). *Research into initial teacher education in Australia: A survey of the literature 1995–2004.* Teaching and Teacher Education, 24(1), 225–239.

- Newton, P. E. (2007). *Clarifying the purposes of educational assessment*. *Assessment in Education*, 14(2), 149–170.
- Oxman, A. D., & Guyatt, G. H. (1993). *The science of reviewing research*. *Annals of the New York Academy of Sciences*, 703(1), 125–134.
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). *Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education*. *Learning and Individual Differences*, 22(6), 806–813.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell.
- Pinar, A. (2011). *Geography teachers' views on the assessment and evaluation instruments and methods used in the renewed geography curriculum*. *Educational Research and Reviews*, 6(3), 334–341.
- Plummer, R., de Loëe, R., & Armitage, D. (2012). *A systematic review of water vulnerability assessment instruments*. *Water Resources Management*, 26, 4327–4346.
- Rudel, T. K. (2008). *Meta-analyses of case studies: A method for studying regional and global environmental change*. *Global Environmental Change-Human and Policy Dimensions*, 18, 18–25.
- Smothers, J. (2002). *Geography standards: Instruction and assessment in action*. *Geographical Education*, 15, 37–42.
- Tanago Gonzales, I., Urquijo, J., Blauhut, V., Villarroya, F., & De Stefano, L. (2016). *Learning from experience: A systematic review of assessments of vulnerability to drought*. *Natural Hazards*, 80 (2), 951–973.
- Tiknaz, Y., & Sutton, A. (2006). *Exploring the role of assessment tasks to promote formative assessment in Key Stage 3 Geography: Evidence from twelve teachers*. *Assessment in Education: Principles, Policy & Practice*, 13(3), 327–343.
- Tomaszewski, B., Vodacek, A., Parody, R., & Holt, N. (2015). *Spatial thinking ability assessment in Rwandan secondary schools: Baseline results*. *Journal of Geography*, 114(2), 39–48.
- van der Schee, J., & Kolkman, R. (2010). *Multimedia tests and geographical education: The 2008 International Geography*

- Olympiad*. International Research in Geographical and Environmental Education, 19(4), 283–293.
- Weeden, P. (2013). *How do we link assessment to making progress in geography?* In D. Lambert & M. Jones (Eds.), *Debates in geography education* (pp.143–154). London: Routledge.
- Wehry, S., Monroe-Ossi, H., Cobb, S., & Fountain, C. (2012). *Concept mapping strategies: Content, tools and assessment for human geography*. Journal of Geography, 111(3), 83–92.
- Wertheim, J. A., Edelson, D. C., Hildebrant, B., Hinde, E., Kenney, M., Kolvoord, R., ... Shavelson, R. (2013). *A road map for improving geography assessment*. Geography Teacher, 10(1), 15–21.
- Yang, D. (2013). *Comparing assessments within junior geography textbooks used in mainland China*. Journal of Geography, 112(2), 58–67.
- Yasar, O. (2009). *A comparative analysis of assessment and evaluation exercises included in geography textbooks written according to the 2005 secondary education geography curriculum and textbooks of the former curriculum in Turkey*. International Journal of Progressive Education, 5 (1), 45–68.