

تحلیل سؤال‌های آزمون علوم تجربی پایه نهم استان کرمانشاه و هم‌ترازی آن‌ها با ابعاد طبقه‌بندی اندرسون و کراتول

فرامرز ملکیان^۱

سیدمحمدرضا مصطفائی^۲

فتانه خاکره^۳

زینب فرجی^۴

تاریخ دریافت: ۹۶/۰۳/۲۰

تاریخ پذیرش: ۹۷/۰۲/۲۸

چکیده

هدف پژوهش حاضر تحلیل شاخص‌های روان‌سنجی و سطوح حیطة شناختی سؤال‌های امتحان علوم تجربی پایه نهم استان کرمانشاه در خرداد ۱۳۹۵ بر اساس طبقه‌بندی اندرسون و کراتول بود. این پژوهش از لحاظ هدف کاربردی، در گروه تحقیقات توصیفی و از نوع ارزشیابی بود. جامعه آماری شامل برگه‌های امتحانی درس علوم تجربی پایه نهم مدارس پسرانه ناحیه ۱ کرمانشاه بوده که تعداد آن‌ها ۲۰۰۷ برگه بود که به روش نمونه‌گیری تصادفی خوشه‌ای نسبی تعداد ۳۲۳ برگه از آن‌ها به‌عنوان حجم نمونه انتخاب شدند. ابزار اندازه‌گیری عبارت بود از فرمول‌های ضریب دشواری، ضریب تمیز، ضریب روایی محتوایی لاشه، ضریب پایایی کودر-ریچاردسون، ضریب آلفای کرانباخ و چک‌لیست اهداف شناختی طبقه‌بندی اندرسون و کراتول. یافته‌های پژوهش نشان داد سؤالات از ضریب دشواری مناسب و ضریب تمیز بالایی برخوردار بوده‌اند. ضریب روایی سؤالات پایین بود. هم‌چنین ضریب پایایی سؤالات عینی، پایین، اما ضریب پایایی سؤالات غیر عینی بالا بود. در مطابقت سؤالات با ابعاد فرایند شناختی و دانش طبقه‌بندی شناختی اندرسون و کراتول، بیش‌تر سؤالات در سطح به‌یادآوردن و در بعد دانش امور واقعی بوده و هیچ سؤالی در حیطة فرایندهای شناختی ارزشیابی کردن و آفریدن و بعد دانش فراشناختی طرح نشده بود.

کلیدواژه‌ها: تحلیل سؤال، شاخص‌های روان‌سنجی، طبقه‌بندی شناختی اندرسون و کراتول، علوم تجربی

^۱ استادیار گروه علوم تربیتی، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران.

^۲ دانشجوی دکتری مدیریت آموزشی گروه علوم تربیتی، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران. (نویسنده مسئول) Mostafaei92@yahoo.com

^۳ دانشجوی دکتری مدیریت آموزشی گروه علوم تربیتی، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران.

^۴ دانشجوی دکتری مدیریت آموزشی گروه علوم تربیتی، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران.

مقدمه

«ارزشیابی به سبب ماهیت و پیچیدگی‌های خاص خود از گسترده‌ترین و جنجالی‌ترین موضوعات آموزشی است و در صورتی که به نحو صحیحی اجرا شود می‌توان بر اساس نتایج آن نقاط قوت و ضعف آموزشی را مشخص نمود» (بقایی و دیگران، ۱۳۹۵: ۲۹۲). ارزشیابی تحصیلی از هدف جمع‌آوری اطلاعات برای تصمیم‌گیری درباره‌ی ارتقا پایه دانش‌آموزان دور شده است و به‌سوی گونه‌ای ارزشیابی روی آورده است که می‌خواهد در فرآیند یادگیری به دانش‌آموز کمک کند (حسینی و کیامنش، ۱۳۸۸: ۸۶). بیبای^۱ (۱۹۷۸) ارزشیابی پیشرفت تحصیلی را این‌چنین تعریف کرده است: «فرایند جمع‌آوری و تفسیر نظام‌دار شواهدی است که در نهایت به قضاوت ارزشی با چشم‌داشت به اقدامی معین بینجامد» (ولف، ۱۳۹۵: ۱۰). تهیه آزمون برای ارزش‌یاب از محتوای یک درس، از مشکل‌ترین کارهایی است که معلمین در جریان فعالیت‌های آموزشی بر عهده دارند. «در صورتی که معلم در انجام فرایند ارزشیابی، از توانمندی لازم برخوردار نباشد، هم به فرایند آموزش و هم به دانش‌آموزان ضرر جدی می‌رسد. چون در جدیدترین تعریفی که از ارزشیابی صورت پذیرفته است، ارزشیابی را کاملاً در خدمت آموزش و یادگیری دانسته‌اند. بعد از این‌که هر کدام از معلمین، آزمون‌های را تهیه و تدارک دیدند، وظیفه مهم‌تری هم در جریان ارزشیابی بر عهده دارند که شامل تحلیل آزمون‌های مختلف است» (امیری‌مجد، ۱۳۹۴: ۱۳). «آزمون‌های پیشرفت تحصیلی آزمون‌هایی است که میزان تسلط فرد بر یک محتوای درسی یا آموخته‌های وی را در زمینه‌های مختلف اندازه‌گیری می‌کند» (سرمد، بازرگان و حجازی، ۱۳۹۳: ۱۰۵). در محیط‌های آموزشی، بازخوردهای حاصل از ارزشیابی تحصیلی، نقش بسیار مهمی را در ارتقای سطح کیفی و کمی یادگیری دانش‌آموزان و عملکرد صحیح معلمان ایفا می‌کند. تحلیل سؤال‌های امتحانی و بررسی میزان سازگاری آن‌ها با هدف‌های آموزشی یکی از اصلی‌ترین بازخوردها در این زمینه هست. در تحلیل سؤال‌های آزمون، نقاط قوت و ضعف یک آزمون و کیفیت همه‌ی سؤال‌های آن تعیین می‌شود؛ بنابراین لازم است که معلمان پس از اجرای هر آزمون، سؤال‌های آن را تحلیل کنند و با استفاده از نتایج حاصل به تجدیدنظر در آزمون و بهبود کیفیت سؤال‌ها برای استفاده‌های بعدی اقدام نمایند. بسیاری از نظام‌های آموزشی جهان بر اساس استانداردهای آموزشی بنا شده‌اند. لذا در این نظام‌ها تمامی اجزاء

^۱ Beeby

برنامه‌ی آموزشی از هماهنگی و هم‌سویی خاصی بهره‌مند است. همان‌طور که کراتول^۱ بیان می‌کند: «ابزارهای آموزشی در هر رشته باید در راستای اهداف تعیین‌شده در استانداردهای آموزشی همان رشته و یا درس خاص باشد» (کراتول، ۲۰۰۲: ۲۱۳). مطالعه در خصوص هم‌ترازی^۲ و هماهنگی بین اجزاء برنامه‌ی درسی ضرورتی انکارناپذیر است. بررسی سؤالات امتحانی به‌عنوان ابزاری برای ارزشیابی پیشرفت تحصیلی و بررسی میزان هم‌ترازی آن‌ها با اهداف حوزه‌ی شناختی یکی از اولویت‌ها و ضرورت‌هاست. «وقتی که از یک آزمون به‌عنوان شاخص پیشرفت در یک حیطه‌ی آموزشی استفاده می‌شود، لازم است شواهد مربوط به این که آن آزمون نمونه‌ای از مجموع دانش‌های موردنظر فراهم می‌آورد و فرایندهای موجود در بعد هدف‌ها را فرامی‌خواند ارائه شود. چنین تحلیلی باید آن جنبه‌هایی را که آزمون معرف آن‌هاست و آن‌هایی را که نمی‌تواند معرف باشد آشکار سازد» (سیف، ۱۳۹۴: ۴۱۸). می‌توان برای پی بردن به هماهنگی بین سؤال‌های امتحانی و اهداف حوزه‌ی شناختی از روش‌های گوناگونی از قبیل مدل وب^۳ (۱۹۷۷)، مدل اچیو^۴ (۲۰۰۱)، مدل پرتو و اسمیت سن^۵ (۲۰۰۱) و مدل اندرسون و کراتول^۶ (۲۰۰۱) استفاده نمود. «از این میان چون مدل اندرسون و کراتول برای تمامی رشته‌ها و موضوعات درسی کاربرد داشته و تنها مدلی است که محتوا را بر اساس انواع دانش طبقه‌بندی می‌کند، این مدل بر مدل‌های قبلی رجحان دارد» (رضوانی و حق‌شناس، ۱۳۹۳: ۹۶). پین^۷ (۲۰۰۳) در رابطه با طبقه‌بندی اندرسون و کراتول گفته است: «دلیل عمده ایجاد طبقه‌بندی تازه آسان ساختن ارتباط در میان معلمان، تولیدکنندگان برنامه‌های درسی و متخصصان سنجش و ارزشیابی بوده است». «تحلیل سؤال مجموعه‌ای از روش‌هایی است که از طریق آن‌ها اثربخشی و کارایی هر یک از سؤالات امتحان و کل امتحان اندازه گرفته می‌شود؛ و این امکان را فراهم می‌سازد تا ویژگی‌های هریک از سؤالات امتحان بررسی شود و از این حیث اطمینان حاصل شود که سؤالات استاندارد بوده و از کیفیت مناسب برای سنجش ویژگی یا توانایی مورد انتظار برخوردار هستند» (کاظمی و احسان پور، ۱۳۸۹: ۶۴۴).

1. Krathwohl

2. Alignment

3. Webb

4. Achieve

5. Porter & Smithson

6. Anderson & Krathwohl

7. Payne

اهمیت و جایگاه کتاب علوم تجربی پایه‌ی نهم در انتخاب رشته‌ی تحصیلی دانش‌آموزان در دوره‌ی متوسطه‌ی دوم به این دلیل که این کتاب در دوره‌ی متوسطه دوم به چهار درس تخصصی و مهم زیست‌شناسی، فیزیک، شیمی و زمین‌شناسی تقسیم می‌شود و یادگیری آن برای دانش‌آموزان بسیار بااهمیت هست، پژوهش‌گران را بر آن داشت تا با تحلیل - به دست آوردن ضریب دشواری، ضریب تمیز، ضریب روایی، ضریب پایایی - سؤال‌های درس علوم تجربی استان کرمانشاه در خرداد ۹۵ به میزان دقت طراحان در طراحی سؤال‌های استاندارد و هم‌ترازی سؤال‌ها با اهداف حوزه‌ی شناختی اندرسون و کراتول، پی ببرند. درنهایت صلاحیت این آزمون را ارزیابی کرده و با شناخت نقاط ضعف و قوت، گام‌های مؤثرتری در جهت اصلاح و بهبود نظام ارزشیابی این درس در پایه‌ی نهم فراهم و پشتوانه‌ی پژوهشی قوی برای طراحی سؤال‌های درس مذکور در آزمون‌های بعد ایجاد شود. بر این اساس لازم و ضروری است طراح سؤال به هنگام تهیه آزمون در راستای اهداف آن درس قدم بردارد؛ و در پی اجرای آزمون، نتایج آن را تجزیه و تحلیل نماید تا با گرفتن بازخورد و مشخص نمودن نتایج به ارزش کمی و کیفی کار خویش پی برده و نقاط قوت و ضعف آزمون را در آزمون‌های آتی مدنظر داشته باشد. بنابراین این پژوهش به سؤالات ذیل پاسخ می‌دهد:

- ۱) ضریب دشواری سؤال‌های آزمون درس علوم تجربی چقدر است؟
- ۲) ضریب تمیز سؤال‌های آزمون درس علوم تجربی چقدر است؟
- ۳) ضریب روایی سؤال‌های آزمون درس علوم تجربی چقدر است؟
- ۴) ضریب پایایی سؤال‌های آزمون درس علوم تجربی چقدر است؟
- ۵) این آزمون کدامیک از ابعاد فرایندهای شناختی اندرسون و کراتول را پوشش می‌دهد و به چه نسبتی؟
- ۶) این آزمون کدامیک از ابعاد دانش اندرسون و کراتول را پوشش می‌دهد و به چه نسبتی؟

مفهوم شناسی

الف) شاخص‌های روان‌سنجی

متخصصان آزمون‌سازی برای ارزشیابی سؤالات آزمون شاخص‌هایی تحت عنوان شاخص‌های روان‌سنجی پیشنهاد کرده‌اند که از طریق آن‌ها سازندگان آزمون می‌توانند با شناسایی کردن نقایص و

برطرف کردن آن‌ها مهارت‌های خود را در نوشتن سؤالات امتحانی افزایش دهند و در کیفیت سؤالات آزمونی که در آینده مورد استفاده قرار خواهند داد، تجدیدنظر کنند. متخصصان و صاحب‌نظران ضریب دشواری، ضریب تمیز، روایی و پایایی سؤال‌ها را از موارد تحلیل آزمون می‌داند. «بنا به تعریف، درصد کل آزمون‌شوندگانی که به یک سؤال پاسخ درست می‌دهند ضریب دشواری^۱ آن سؤال است» (حسنی، ۱۳۸۶: ۴۴). هراندازه ضریب دشواری یک سؤال بزرگ‌تر باشد آن سؤال آسان‌تر و هراندازه این ضریب کوچک‌تر باشد، سؤال دشوارتر است. نوع سؤال می‌تواند در ضریب دشواری تأثیر داشته باشد. در یک حالت عمومی می‌توان گفت سؤال‌هایی بهتر هستند که ضریب دشواری آن‌ها از یک کمتر و از صفر بیشتر و به ۰/۵ نزدیک باشند. چنانچه بخواهیم به توان و قدرت سؤال برای جداسازی گروه قوی از ضعیف پی ببریم، از مفهومی به نام ضریب تمیز^۲ استفاده می‌کنیم. به دست آوردن این ضریب مشخص می‌کند یک سؤال تا چه اندازه می‌تواند گروه قوی را از گروه ضعیف جدا سازد. دامنه تغییرات ضریب تمیز از -۱ تا +۱ است. «هرقدر ضریب تمیز بزرگ‌تر باشد، قوه تمیز آن پرسش بیشتر و هرقدر ضریب تمیز کوچک‌تر باشد، قوه تمیز آن کم‌تر است. ضریب تمیز صفر حاکی از آن است که پرسش به‌هیچ‌وجه نتوانسته است بین گروه ضعیف و قوی تمایز قائل شود» (پاشا شریفی، ۱۳۹۵). از دیگر موارد تحلیل سؤال روایی^۳ هست. «مقصود از روایی آن است که وسیله‌ی اندازه‌گیری، بتواند خصیصه و ویژگی موردنظر را اندازه بگیرد» (خاکی، ۱۳۹۱: ۲۸۸). «منظور از روایی تست یا هر وسیله‌ی اندازه‌گیری روشن کردن و صراحت بخشیدن به این نکته است؛ که آیا وسیله موردنظر برای اندازه‌گیری آن خصیصه‌ای که خواسته شده است تا چه حد مناسب است؟ تا چه اندازه قادر است آنچه را باید بسنجد اندازه بگیرد؟ تمامی آن را اندازه بگیرد و جز آن هیچ صفت یا ویژگی دیگری را اندازه‌گیری نکند» (نوروزی، ۱۳۸۶: ۸-۱۵۷). برای تعیین روایی محتوایی از قضاوت متخصصان در مورد میزان مطابقت سؤال‌ها با محتوا و اهداف آن درس استفاده می‌شود. چهارمین مورد تحلیل آزمون به پایایی^۴ آن مربوط می‌شود. پایایی یا قابلیت اعتماد یکی از ویژگی‌های مهم سؤال‌های آزمون‌ها هست. «ازجمله تعریف‌هایی که برای قابلیت اعتماد ارائه شده است می‌توان به تعریف ایبل و فریسیبی (Ebel & Frisbie, ۱۹۸۹) به نقل از (Jurs, ۱۹۹۰ Wiersma & Frisbie, ۱۹۸۹) اشاره کرد:

1. Difficulty index

2. Discriminative index

3. Validity

4. Reliability

همبستگی میان یک مجموعه از نمرات و مجموعه دیگری از نمرات در یک آزمون معادل که به صورت مستقل بر یک گروه آزمودنی به دست آمده است» (سرمد و همکاران، ۱۳۹۳: ۱۶۶). «پایایی یک آزمون بیانگر ثبات در اندازه‌گیری چیزی است که آزمون اندازه می‌گیرد. در آزمون‌های هنجاری مفهوم پایایی با پراکندگی در نمره‌ی آزمون ارتباط دارد. اگر نتایج حاصل از دو بار اندازه‌گیری یک صفت یا یک آزمون (نظیر دو بار تصحیح اوراق توسط یک نفر) یا دو بار اندازه‌گیری یک صفت با دو وسیله‌ی همتا (نظیر اوراق توسط دو نفر به صورت مستقل) کم‌وبیش یکسان باشد، ابزار اندازه‌گیری را پایا می‌گویند» (کیامنش، ۱۳۹۳: ۹۶).

ب) طبقه‌بندی اندرسون و کراتول

یکی دیگر از هدف‌های ارزشیابی و گرفتن آزمون‌ها، تعیین کردن میزان دستیابی به هدف‌های آموزشی و یادگیری هست. هدف‌های یادگیری خود برخاسته از هدف‌های غایی و کلی هستند. در ارزشیابی آموزشی عملکرد یادگیرندگان با توجه به هدف‌های یادگیری سنجش می‌شوند. هدف‌های آموزشی متنوع هستند و طبقه‌بندی‌های مختلفی از آن‌ها صورت گرفته است. یکی از این طبقه‌بندی‌ها موسوم به طبقه‌بندی تجدیدنظر شده حوزه شناختی هست که توسط اندرسون، کراتول و همکاران (۲۰۰۱) انجام گرفته است. «طبقه‌بندی اندرسون و کراتول، طبقه‌بندی تجدیدنظر شده بلوم بود که دارای یک بعد دانش و یک بعد فرایند شناختی هست. بعد فرایند شناختی دربرگیرنده سطوح «به یادآوردن، فهمیدن، به کار بستن، تحلیل کردن، ارزشیابی و آفریدن است. طبقه‌بندی هر دو بعد، به صورت سلسله‌مراتبی، یعنی از عینی به انتزاعی و از ساده به مشکل است» (کراتول، ۲۰۰۲: ۲۱۲). «بعد دانش شامل، دانش امور واقعی^۱ که دربردارنده دانش اصطلاحات، دانش اجزا و عناصر خاص هست؛ دانش مفهومی^۲ شامل دانش مقوله‌ها، طبقه‌ها و روابط میان آن‌هاست؛ دانش روندی^۳ دربردارنده دانش انجام دادن کارها هست و بالاخره دانش فراشناختی^۴، دانش فرد درباره شناخت خودش است» (پین، ۲۰۰۳: صص ۱۱۶-۱۱۷). «اندرسون و کراتول فرایندهای شناختی «به یادآوردن، فهمیدن، به کار بستن» را جزء سطوح پایین یادگیری و «تحلیل کردن، ارزشیابی و

1. Factual knowledge

2. Conceptual knowledge

3. Procedural knowledge

4. Metacognitive knowledge

آفریدن» را در بالاترین مرحله یادگیری (سطوح بالایی) در نظر گرفته‌اند که در هر برنامه آموزشی می‌بایست موردتوجه بیشتر قرار گیرند» (رضوانی، ۱۳۹۳: ۹۷).

پژوهش‌های متعددی مشابه این پژوهش انجام گرفته است. در تحقیقی که توسط میرآقایی و همکاران (۱۳۹۴) انجام گرفت شاخص‌های روان‌سنجی و سطوح حیطه شناختی سؤال‌های امتحانات نهایی ریاضیات و علوم پایه سوم راهنمایی شهرستان خرم‌آباد در خرداد ۱۳۸۶ تحلیل شدند. نتایج محاسبات نشان داد ضریب دشواری سؤال‌ها مناسب نبوده ولی ضریب تمیز آن‌ها رضایت‌بخش بود. پایایی سؤال‌ها (۰/۹۰) نشان داد که آزمون از ثبات خوبی برخوردار بوده است. بیشتر سؤال‌های ریاضی مربوط به سطح کاربرد (۸۶٪)، فهم (۶٪) و دانش و تحلیل (هرکدام ۴٪) مربوط بودند. در درس علوم تجربی هیچ سؤال‌ی به سطوح تحلیل، ترکیب و ارزشیابی اختصاص نداشتند. امیری مجد (۱۳۹۴) در مطالعه‌ای با عنوان تحلیل مواد آزمون پایان سال دروس ریاضی، فیزیک، شیمی و علوم اجتماعی سال اول دبیرستان شهرستان سرابله با توجه به طبقه‌بندی اهداف شناختی اندرسون و کراتول انجام داد، نتایج نشان داد که ضریب دشواری و تمیز دروس ریاضی و فیزیک در سطح متوسط و دروس شیمی و مطالعات اجتماعی نیز با توجه به ضریب دشواری سؤالات آسانی که به جنبه تمایز گذاری بین دو گروه از دانش‌آموزان ضعیف بودند. از جنبه شاخص پایایی آزمون‌ها پایایی در سطح مطلوبی داشتند. درزمینه‌ی تحلیل محتوای سؤالات بر اساس طبقه‌بندی اندرسون و کراتول در سطح شناختی بیشترین توجه طراحان سؤالات در دروس ریاضی و فیزیک بر سطح تحلیل و در دروس شیمی و علوم اجتماعی بیشترین تأکید بر سطح یادآوردن بود. همچنین درزمینه‌ی تحلیل محتوای سؤالات بر اساس طبقه‌بندی اندرسون و کراتول در سطح دانشی بیشترین توجه طراحان سؤالات در دروس ریاضی، فیزیک و شیمی بر دانش روندی و در دروس علوم اجتماعی بیشترین تأکید بر دانش امور واقعی، مفهومی و روندی به‌طور هم‌زمان توجه شده بود. در نتیجه می‌توان گفت در این آزمون‌ها درزمینه‌ی سطوح شناختی به سطح ارزشیابی و آفریدن و درزمینه‌ی سطح دانشی به دانش فراشناختی کم‌ترین توجه شده بود. رضوانی و حق‌شناس (۱۳۹۳) در پژوهشی با عنوان ارزیابی هم‌ترازی کتاب‌های کارشناسی زبان انگلیسی تخصصی رشته‌های علوم انسانی انتشارات سمت و استانداردهای آموزشی مربوطه بر اساس اهداف آموزشی اندرسون و کراتول به این نتایج دست یافتند که اهداف در نظر گرفته‌شده در کتاب‌ها (۷۱/۰۶٪) و استانداردهای (۸۱/۱۴٪) مربوط به آن‌ها را سطوح پایینی (به یادآوردن، فهمیدن، به کار بستن) طبقه‌بندی اهداف آموزشی تشکیل می‌دهند و تنها ۱۲/۶ درصد

استانداردهای آموزشی و ۲۸/۹۴ درصد کتاب‌ها را اهداف سطوح بالایی اهداف آموزشی تحلیل کردن، ارزشیابی کردن، آفریدن به خود اختصاص داده‌اند. همچنین نتایج محاسبه هم‌ترازی (۰/۴۱) استانداردهای آموزشی و کتاب‌ها نشان داد که این دو جزء از هم‌ترازی قابل قبولی برخوردار نیستند. گواهی، علیزاده و اورانوس بخت (۱۳۹۱) در پژوهشی با عنوان تحلیل آزمون هماهنگ کشوری درس فیزیک ۱ و آزمایشگاه به بررسی نقاط قوت و ضعف آزمون نامبرده پرداختند. پژوهش‌گران به این نتیجه دست یافتند که پرسش‌هایی که مربوط به طبقه دانش می‌باشند، ضریب تمیز خوبی ندارند؛ زیرا بین آزمون‌شوندگان دارای نمره‌های متفاوت در آزمون تمیز قائل نمی‌شوند (ص: ۴۰). زمانی (۱۳۹۱) به بررسی هم‌ترازی آزمون کارشناسی ارشد رشته‌های ترجمه و آموزش زبان انگلیسی در ایران و سرفصل‌های آموزشی و کتاب‌های رسمی کشور با استفاده از طبقه‌بندی اهداف آموزشی اندرسون و کراتول پرداختند. نتایج نشان داد که در رشته بر سطوح پایینی طبقه‌بندی اهداف آموزشی اندرسون و کراتول (به یادآوردن، فهمیدن، به کار بستن) تأکید شده است. در رشته‌ی آموزش زبان انگلیسی، سرفصل‌ها و آزمون‌ها بیش‌ترین شاخص هم‌ترازی را داشتند (۰/۷۰) درحالی‌که سرفصل‌ها و کتاب‌ها کم‌ترین شاخص را داشتند (۰/۵۱) در رشته‌ی مترجمی زبان انگلیسی سرفصل‌ها و آزمون‌ها و همچنین کتاب‌ها و آزمون‌ها از بیشترین شاخص هم‌ترازی (۰/۶۰) برخوردار بودند. همچنین نتایج تحقیق نشان داد از بین کتاب‌های آموزش زبان انگلیسی کتاب آزمون‌سازی زبان انگلیسی بیشترین تأکید را بر سطوح بالایی طبقه‌بندی اهداف و کتاب زبان‌شناسی بیشترین تأکید را بر سطوح پایینی طبقه‌بندی اهداف داشته‌اند؛ و در رشته‌ی مترجمی زبان انگلیسی کتب مربوط به مهارت ترجمه بیشترین تأکید را بر سطوح بالایی طبقه‌بندی اهداف داشته‌اند. به‌علاوه رشته‌ی مترجمی زبان انگلیسی نسبت به رشته آموزش زبان انگلیسی بیشتر بر سطوح بالایی طبقه‌بندی اهداف آموزشی اندرسون و کراتول تأکید کرده است. شوندر غربی (۱۳۸۹) در پژوهش خود به ساخت، اعتبار یابی و روا سازی آزمون سنجش عملکردی در دروس علوم تجربی و ریاضیات پایه سوم ابتدایی شهر تهران پرداخت. نتایج نشان داد که بیشتر سؤالات دارای ضریب دشواری بالا و ضریب تمیز پایینی هستند؛ و آزمون‌ها به‌صورت کارآمدی نتوانسته‌اند بین دو گروه تفاوت را نشان دهند (چکیده). ربی‌هاوی (۱۳۸۹) نیز طی پژوهشی به بررسی میزان انطباق سؤال‌های امتحانی دروس مراکز تربیت‌معلم استان خوزستان با اصول روان‌سنجی پرداخت. نتایج نشان داد ضریب دشواری سؤال‌های امتحانی هر دو درس در مقایسه با ملاک پیشنهادی (۰/۳۰ تا ۰/۷۰) مناسب نبودند ولی

ضریب تمیز سؤال‌های امتحانی به‌استثنای تعداد کمی از سؤال‌ها در مقایسه با ملاک پیشنهادی (۰/۲۰) یا بالاتر) در حد مناسب و قابل قبولی بودند. با تعیین ضریب پایایی مشخص شد آزمون از دقت و ثبات مناسبی برخوردار نبوده است. همچنین مشخص گردید که سؤال‌ها بیشتر در سطح دانش و درک و فهم بوده و به اندازه‌گیری سطوح بالای شناختی (ترکیب و ارزشیابی) اختصاص نداشتند (صص: ۸۱-۶۸). جعفری (۱۳۸۸) به تحلیل محتوای سؤال‌های کتابهای ریاضی دوم و سوم دبیرستان در رشته ادبیات و علوم انسانی بر اساس طبقه‌بندی تجدیدنظر شده بلوم در حیطه شناختی پرداخت. نتایج نشان داد در کتاب ریاضی دوم دانش روندی بیشترین درصد سؤال‌ها (۶۵/۷۳) را به خود اختصاص داده است؛ اما هیچ سؤالی در سطح دانش فراشناختی تدوین نشده بود. در بعد فرایند شناختی در کتاب ریاضی دوم فهمیدن بیشترین درصد سؤال‌ها (۴۳/۸۲) و تحلیل کم‌ترین درصد (۱/۹۷) را به خود اختصاص داده‌اند. در کتاب ریاضی سوم در بعد دانش، دانش روندی بیشترین درصد سؤال‌ها (۸۱/۴۳) و دانش فراشناختی کم‌ترین درصد سؤال‌ها (۰) را به خود اختصاص داده‌اند. مقایسه فراوانی و درصد طبقات مختلف بعد فرایند شناختی، در کتاب ریاضی سوم نشان داد که طبقه کاربرد بیشترین درصد سؤال‌ها (۵۲/۷۴) را به خود اختصاص داده و فاصله زیادی با طبقه ارزشیابی دارد که کم‌ترین درصد سؤال‌ها را به خود اختصاص داده است (صص: ۱۸۲-۱۸۰). کارا و ژپنی^۱ (۲۰۱۱) در ترکیه میزان هم‌ترازی استانداردها و امتحانات زیست‌شناسی مقطع راهنمایی را بررسی کردند که نتیجه حاکی از پایین بودن هم‌ترازی میان اجزای برنامه‌ی آموزشی بود (ص: ۸۳). ادواردز^۲ (۲۰۱۰) با استفاده از روش اندرسون و کراتول و شاخص پورتر، هم‌ترازی استانداردها و امتحانات پایانی درس فیزیک و شیمی مقطع متوسطه را در آفریقای جنوبی بررسی کردند. میزان هم‌ترازی به‌دست‌آمده در این پژوهش برای درس فیزیک و استانداردها ۰/۶ و برای درس شیمی و استانداردهای مربوطه ۰/۸ بود. به‌علاوه، در هر دو درس فرایند شناخت به‌یادآوردن کم‌تر از سایر روش فرایندها موردتوجه قرار گرفته بود. این در حالی است که فرایندهای فهمیدن و به کار بستن بیشترین نوع شناختی در هر دو درس بودند. درواقع سطوح پایینی طبقه‌بندی اهداف آموزشی اندرسون و کراتول بیشتر از سطوح بالایی موردتوجه بوده است (ص: ۵۸۶). مارتن و سیرسی^۳ (۲۰۰۹) سه مدل هماهنگی برنامه آموزشی

1. Kara and Çepni

2. Edwards

3. Martone & Sireci

پرتر، وب و آچیو برای ارزیابی هم‌ترازی اهداف، آزمون‌ها و فعالیت‌های آموزشی را مورد بررسی قرار دادند. نتایج به‌دست‌آمده نشان داد که هر سه مدل به ارزشیابی محتوا و فرایندهای شناختی اهداف، آزمون‌ها و فعالیت‌های آموزشی با استفاده از متد پرتر پرداختند. مدل‌های وب و آچیو بیش‌تر برای ارزیابی و مقایسه گستره و تعادل میان اهداف و آزمون‌ها استفاده می‌شوند. برد میزان پوشش اهداف آموزشی در آزمون‌ها را بررسی می‌کند و به مقایسه تعداد پرسش‌ها و اهداف آموزشی می‌پردازد (ص: ۱۳۵۱). در پژوهشی دیگر وب^۱ (۲۰۰۷) هم‌ترازی استانداردها و امتحانات درس ریاضی هفت مقطع تحصیلی در ایالت ویسکانسین^۲ آمریکا را بررسی کرد. عدم وجود هم‌ترازی کافی میان استانداردها و امتحانات، نتیجه‌ی این مطالعه بود (ص: ۲۴). هانگ، چانگ و لین^۳ (۲۰۰۶) بر اساس اهداف آموزشی اندرسون و کراتول بعد دانش استانداردهای درس مطالعه فناوری را در تمامی کلاس‌های ابتدایی بررسی کردند. در استانداردها، دانش فراشناختی ۰/۴۰ و دانش مفهومی ۰/۳۶ بیشترین میزان را به خود اختصاص دادند. در این مطالعه محققان دریافتند که استانداردها دانش امور واقعی ۰/۷ و دانش روندی ۰/۱۷ را کمتر در برداشتند (ص: ۲۹۷). گادو^۴ (۲۰۰۴)، در پژوهش خود، به بررسی شاخص‌های روان‌سنجی آزمون‌های ملی معلم ساخته (آزمون‌های نهایی) فیزیک و شیمی سال‌های اول، دوم و سوم دبیرستان در بنین^۵ پرداخت. در این تحقیق برای بررسی روایی محتوایی یک نمونه ۲۵ نفری از معلمان دروس مربوطه و برای محاسبه ضرایب پایایی، دشواری و تمیز، از هر درس در هر پایه ۶۳۰ برگه‌ی نمره‌گذاری شده انتخاب شد. نتایج نشان دادند که آزمون‌ها بسیار پایا بودند و شاخص‌های دشواری و تمیز سؤالات از سطح رضایت‌بخش برخوردار بودند، اما فاقد روایی محتوایی بودند. با کاوش و نگاه به پیشینه پژوهش متوجه اهمیت موضوع و کاربرد آن هستیم و پژوهش‌های متعددی در این زمینه به اجرا درآمده است؛ اما آنچه این پژوهش را از بقیه متمایز می‌کند اجرای آن برای بار اول در سؤال‌های نهایی درس علوم تجربی پایه نهم استان کرمانشاه هست. ضرورت این امر زمانی محرز می‌شود که آموزش و پرورش این استان در زمینه‌ی طراحی

1. Webb

2. Wisconsin

3. Huang, Chang & Yi Lin

4. Gado

5. Benin

سؤالات رتبه خوبی در بین استان‌های کشور ندارد. در ضمن پژوهشی که تمام متغیرهای موجود در پژوهش جاری را پوشش دهد یافت نشد.

روش تحقیق

این پژوهش چون به دنبال استفاده از نتایج پژوهش برای یافتن مشکلات و حل آن‌ها بود، از لحاظ هدف از نوع کاربردی و چون در پی توصیف «آنچه هست» بود، از نظر ماهیت در گروه تحقیقات توصیفی و از نوع ارزشیابی هست. پژوهش ارزشیابی به فرایند قضاوت درباره ارزش، شایستگی و سودمندی اهداف و برنامه‌های آموزشی گفته می‌شود. «هدف روش تحقیق ارزشیابی، جمع‌آوری اطلاعات تفصیلی به‌عنوان پایه‌ای برای قضاوت درباره مؤثر بودن و ثمربخشی محصول، فرایند یا برنامه برحسب هدف‌ها یا ارزش‌های پذیرفته‌شده می‌باشد» (بست^۱، ۱۳۹۰). جامعه آماری عبارت است از برگه‌های امتحانی درس علوم تجربی-آزمون درس علوم تجربی پایه‌ی نهم استان کرمانشاه در خرداد ۱۳۹۵ شامل ۱۹ سؤال کلی بود که با توجه به این که سؤالات ۱، ۲، ۳، ۴ و ۸ هر کدام چند سؤال فرعی دیگر را پوشش می‌دادند، در مجموع ۳۴ سؤال مورد تحلیل قرار گرفت - تمامی دانش‌آموزان پسر پایه‌ی نهم مدارس متوسطه‌ی اول آموزش و پرورش ناحیه‌ی یک کرمانشاه که در امتحانات خرداد سال ۱۳۹۵ شرکت داشته‌اند و تعداد آن‌ها ۲۰۰۷ نفر است. برای انتخاب نمونه‌ای مکفی و معرف جامعه مدارس مذکور به سه دسته‌ی برخوردار (۶ مدرسه-۲۳٪)، نیمه برخوردار (۱۳ مدرسه-۵۰٪) و محروم (۷ مدرسه-۲۷٪) تقسیم‌بندی شدند. (این تقسیم‌بندی توسط اداره آموزش و پرورش ناحیه یک شهر کرمانشاه و بر اساس منطقه جغرافیایی مدارس صورت گرفته است). حجم نمونه بر اساس فرمول کوکران ۳۲۳ نفر برآورد گردید. برای انتخاب نمونه‌ها با توجه به درصد توزیع مدارس از نمونه‌گیری خوشه‌ای نسبی استفاده شد و تعداد ۷۵ برگه امتحانی از مدارس برخوردار، ۱۶۱ برگه امتحانی از مدارس نیمه برخوردار و ۸۷ برگه امتحانی از مدارس محروم انتخاب شدند. در این پژوهش برای رسیدن به جواب سؤالات پژوهش از چندین ابزار مختلف، به شرح ذیل استفاده شد: برای تحلیل سؤال‌های آزمون از روش‌های کمی (ضریب دشواری، ضریب تمیز، روایی و پایایی) و برای پی بردن به میزان هم‌ترازی و همگنی سؤال‌ها با ابعاد طبقه‌بندی شناختی اندرسون و کراتول، از روش کیفی - تفسیر و بررسی محتوای سؤال‌ها - استفاده گردید. بدین منظور تعداد ۳۲۳

^۱. Best

نمونه از برگه‌های امتحانی درس علوم تجربی پایه‌ی نهم دانش آموزان پسر ناحیه یک کرمانشاه در خرداد ۱۳۹۵ مورد تجزیه و تحلیل قرار گرفتند. ابتدا برگه‌های امتحانی از پایین‌ترین نمره تا بالاترین نمره مرتب شدند. سپس ۲۷٪ برگه‌ها از پایین به عنوان گروه پایین و ۲۷٪ برگه‌ها از بالا به عنوان گروه بالا انتخاب شدند. متخصصان آزمون‌سازی (ویتنی و سیبرز، ۱۹۷۰؛ نیتکو ۱۹۸۳) پیشنهاد کرده‌اند: در شرایطی که تعداد کل آزمون‌شوندگان بیش از ۴۰ نفر است بهترین رقم برای گروه بالا و پایین ۲۷٪ کل برگه‌هاست.

بر این اساس تعداد برگه‌های امتحانی هر گروه ۹۰ برگه هست. برای به دست آوردن ضریب دشواری از فرمول

$$\text{ضریب دشواری} = \frac{\text{انتخاب‌های درست گروه پایین} + \text{انتخاب‌های درست گروه بالا}}{\text{تعداد افراد گروه بالا} + \text{تعداد افراد گروه پایین}} \times 100$$

و برای محاسبه ضریب تمیز از فرمول:

$$\text{ضریب تمیز سؤال} = \frac{\text{انتخاب‌های درست گروه پایین} - \text{انتخاب‌های درست گروه بالا}}{\text{تعداد افراد یک گروه (بالا یا پایین)}}$$

استفاده شد (خورشیدی و ملک‌شاهی، ۱۳۸۵: ۱۲۴). «برای ارزیابی روایی محتوایی به شکل کمی، از ضریب روایی محتوایی^۱ استفاده شد. این روش به روش لاشه شهرت دارد» (شولتز و ویتنی^۲، ۲۰۰۵). برای این منظور از نه متخصص درس علوم تجربی - مدرسین روش تدریس علوم تجربی در دانشگاه فرهنگیان، سرگروه‌های آموزشی و دبیران مجرب - درخواست شد میزان هماهنگی محتوایی هر یک از سؤالات با اهداف درس مذکور را بر اساس طیف سه‌قسمتی «ضروری است»، «مفید است ولی ضرورتی ندارد» و «ضرورتی ندارد» تعیین نمایند. پس از آن پاسخ‌ها مطابق فرمول زیر محاسبه گردید.

$$CVR = \frac{n_E - \frac{N}{2}}{\frac{N}{2}}$$

^۱. Content validity Ratio

^۲. SHultz & Whitney

nE تعداد متخصصانی است که گزینه‌ی «ضروری» را انتخاب نموده‌اند؛ و N تعداد کل متخصصان است. ضریب روایی محاسبه شده با فرمول بالا برای هر سؤال دامنه‌ای از $+1$ تا -1 است که هر چه رقم بزرگ‌تر باشد روایی سؤال بیشتر است. برای برآورد روایی هر سؤال، روایی به دست آمده از فرمول بالا با حداقل مقدار روایی جدول ۱ مقایسه می‌شود. اگر مقدار به دست آمده از مقدار جدول بیشتر بود آن سؤال دارای روایی و اگر کم‌تر بود آن سؤال فاقد روایی است. «برای تعیین ضریب روایی کل آزمون، لازم است میانگین ضریب‌های روایی همه‌ی سؤال‌ها محاسبه شود که همان شاخص روایی محتوایی^۱ است» (سیف، ۱۳۹۴: ۵۴۴).

جدول ۱: تصمیم‌گیری در مورد ضریب روایی محتوایی

تعداد متخصصان	۵	۶	۷	۸	۹	۱۰	۱۵	۲۰	۲۵	۳۰	۴۰
حداقل مقدار روایی	۰/۹۹	۰/۹۹	۰/۹۹	۰/۸۵	۰/۷۸	۰/۶۲	۰/۴۹	۰/۴۲	۰/۳۷	۰/۳۳	۰/۲۹

منبع: Lawshe (1975)

از آنجاکه سؤال‌های این آزمون شامل سؤال‌های عینی (چهار جوابی، جاخالی، صحیح-غلط و جور کردنی) و سؤال‌های غیر عینی (تشریحی و مسئله) است، برای به دست آوردن پایایی دسته اول از روش کودر-ریچاردسون^۲ و فرمول KR20

$$r_{tt} = \frac{n}{n-1} \left(1 - \frac{\sum pq}{\sigma^2} \right)$$

استفاده شد زیرا «وقتی هر سوال در الگوی همه‌یاهیچ، و به صورت ۱ یا صفر نمره‌گذاری شود، فرمول بالا مورد استفاده قرار می‌گیرد» (هومن، ۱۳۸۹: ۲۳۵). برای به دست آوردن پایایی سؤال‌های غیر عینی ابتدا هر سؤال با توجه به بارم در یک طیف ۵ رشته‌ای قرار داده شد و سپس با استفاده از ضریب آلفا^۳ در نرم‌افزار SPSS نسخه ۲۲ پایایی آن‌ها به دست آمد. ابزار مورد استفاده برای بررسی هم‌ترازی سؤال‌ها با ابعاد طبقه‌بندی حوزه‌ی شناختی اندرسون و کراتول چک فهرستی بود که توسط رضوانی و زمانی (۱۳۹۱) طراحی شده که نمونه‌ی آن در شکل ذیل آمده است (جدول ۲). البته این چک‌لیست در این پژوهش به دو چک‌لیست مجزا در بعد فرایندهای شناختی و بعد فرایند دانشی تقسیم شده است؛ که در بخش تجزیه و تحلیل نشان داده خواهد شد.

1. Content validity Index

2. Kuder-Richardson

3. Coefficient alpha

جدول ۲: چک لیست اهداف شناختی طبقه‌بندی اندرسون و کراتول

بعد دانش			
فرایندهای شناختی	دانش امور واقعی	دانش امور مفهومی	دانش امور روندی
به یاد آوردن: ۱- تشخیص دانش داده‌ها			
۲- فراخوان دانش‌های گذشته			
فهمیدن: ۱- تفسیر داده‌ها ۲- مثال آوردن ۳- طبقه‌بندی کردن داده‌ها ۴- استنباط کردن ۵- مقایسه کردن ۶- شرح دادن			
به کار بستن: ۱- اجرای یک‌رویه برای کاری که آشنا است ۲- اجرای یک‌رویه برای انجام کاری که ناآشنا است			
تحلیل کردن: ۱- تفکیک داده‌ها به لحاظ میزان اهمیت و ارتباط اجزا باهم ۲- سازمان‌دهی داده‌ها ۳- پی بردن به مقصود و منظور نویسنده			
ارزشیابی کردن: آزمایش و بررسی داده‌ها ۲- انتقاد و داوری داده‌ها			
آفریدن: ۱- دادن پیشنهادها برای حل مسئله ۲- برنامه‌ریزی برای حل مسئله ۳- انجام طرح برای حل مسئله			

برای مقایسه و برآورد میزان هماهنگی این سؤالات با اهداف حیطه‌ی شناختی اندرسون و کراتول از متخصصین و اساتید دانشگاه در رشته‌های برنامه‌ریزی درسی و مدیریت آموزشی به روش فن دلفی^۱ استفاده شد. برای انجام این کار ابتدا سؤالات و چک‌لیست اهداف شناختی اندرسون و کراتول (جدول ۱) برای متخصصین ارسال شد. پس از بازگرداندن چک‌لیست‌ها، پژوهشگران آن‌ها را تجزیه و تحلیل کرده و بر اساس فراوانی پاسخ‌ها، نحوه همپوشانی سؤالاتی که آزمون با حیطه‌های فراشناختی و بعد دانش را تنظیم نموده و برای بار دوم به متخصصان ارسال نمودند. متخصصان با موافقت با ترکیب جدید بر درست بودن آن صحنه گذاشتند. به منظور بررسی هدف‌های شماره ۱ الی ۳ از فرمول‌ها و روابطی که در بخش ابزار پژوهش به آن‌ها اشاره شد استفاده گردید. همچنین پایایی سؤالات با روش کودر- ریچاردسون و آلفای کرانباخ و به کمک نرم‌افزار SPSS نسخه‌ی ۲۲ انجام پذیرفت؛ و برای بررسی هدف‌های شماره ۵ و ۶ افعال و اسامی موجود در محتوای سؤالات با افعال و اسامی مندرج در چک‌لیست اهداف شناختی طبقه‌بندی اندرسون و کراتول مطابقت داده شد.

^۱. Delphi Tecnique

یافته‌ها

سؤال اول پژوهش: ضریب دشواری سؤال‌های آزمون درس علوم تجربی پایه‌ی نهم استان کرمانشاه در خرداد ۱۳۹۵ چقدر است؟

جدول ۳: تحلیل ضریب دشواری سؤالات پایه نهم علوم تجربی استان کرمانشاه در خرداد ۱۳۹۵

ضریب دشواری	ردیف سؤال	
۰/۹۶	A	۱
۰/۴۶	B	
۰/۷۵	C	
۰/۶۸	D	
۰/۹۶	A	۲
۰/۹۳	B	
۰/۷۳	C	
۰/۷۳	D	
۰/۵۸	A	۳
۰/۵۵	B	
۰/۶۸	C	
۰/۳۶	D	
۰/۷۱	E	
۰/۷۳	F	
۰/۶۰	G	
۰/۸۳	A	۴
۰/۶۱	B	
۰/۸۰	C	
۰/۷۰		۵
۰/۷۰		۶
۰/۵۴		۷
۰/۶۶	الف	۸
۰/۴	ب	
۰/۵۱		۹
۰/۷۳		۱۰
۰/۷۷		۱۱
۰/۶۵		۱۲
۰/۴۰		۱۳
۰/۸۵		۱۴
۰/۶۳		۱۵
۰/۵۸		۱۶

۰/۴۰	۱۷
۰/۶۴	۱۸
۰/۶۰	۱۹

با توجه به مقادیر جدول ۳ سؤال‌های شماره ۱(A) و ۲(A) با ضریب دشواری (۰/۹۶) آسان‌ترین سؤال‌ها و سؤال شماره ۳(D) با ضریب دشواری (۰/۳۶) سخت‌ترین سؤال است. از این سؤال‌ها تعداد ۱۳ سؤال که ضریب دشواری آن‌ها بیشتر از ۰/۷ است در دسته سؤال‌های آسان هستند. همچنین تعداد ۲۱ سؤال که ضریب دشواری آن‌ها بین ۰/۳ تا ۰/۷ است، می‌توانند حداکثر اطلاع را درباره تفاوت بین آزمودنی‌ها به دست دهند.

سؤال دوم پژوهش: ضریب تمیز سؤال‌های آزمون درس علوم تجربی پایه‌ی نهم استان کرمانشاه در خرداد ۱۳۹۵ چقدر است؟

جدول ۴: تحلیل ضریب تمیز سؤالات پایه نهم علوم تجربی استان کرمانشاه در خرداد ۱۳۹۵

ضریب تمیز	ردیف سؤال	
۰/۰۶	A	۱
۰/۲۶	B	
۰/۵۳	C	
۰/۴۳	D	
۰/۰۶	A	۲
۰/۱۳	B	
۰/۱۶	C	
۰/۳۳	D	
۰/۸۳	A	۳
۰/۷۰	B	
۰/۵۰	C	
۰/۷۳	D	
۰/۷۳	E	
۰/۳۰	F	
۰/۷۳	G	
۰	A	۴
۰/۴۳	B	
۰/۴۰	C	
۰/۴۰		۵
۰/۳۸		۶
۰/۵۸		۷

۰/۶۶	الف	۸
۰/۷۳	ب	۹
۰/۹۳		۱۰
۰/۵۰		۱۱
۰/۳۱		۱۲
۰/۵۱		۱۳
۰/۴۰		۱۴
۰/۳۰		۱۵
۰/۵۰		۱۶
۰/۷۵		۱۷
۰/۴۱		۱۸
۰/۶۵		۱۹
۰/۳۴		

دامنه تغییرات ضریب تمیز از ۱- تا ۱+ است. «هرقدر ضریب تمیز بزرگ‌تر باشد، قوه تمیز آن پرسش بیشتر و هرقدر ضریب تمیز کوچک‌تر باشد، قوه تمیز آن کم‌تر است. ضریب تمیز صفر حاکی از آن است که پرسش به‌هیچ‌وجه نتوانسته است بین گروه ضعیف و قوی تمایز قائل شود» (پاشا شریفی، ۱۳۹۵). سؤالاتی که ضریب تمیز بالاتر از ۰/۴ دارند بسیار خوب، سؤالاتی که ضریب تمیز بین ۰/۳ و ۰/۳۹ دارند خوب، سؤالاتی که ضریب تمیز بین ۰/۲ تا ۰/۲۹ دارند نیاز به تجدیدنظر دارند. پرسش‌هایی که ضریب تمیز کم‌تر از ۰/۱۹ دارند ضعیف هستند. با استناد به جدول ۴ سؤال شماره ۹ با ضریب تمیز ۰/۹۳ در جدا کردن آزمون‌شوندگان قوی و ضعیف دارای بیش‌ترین توان تمیز است؛ و سؤال شماره ۴(A) با ضریب تمیز صفر بیان‌گر آن است که این سؤال به‌هیچ‌وجه نتوانسته بین گروه قوی و ضعیف تمایزی قائل شود. تمامی سؤالاتی که ضریب تمیز پایینی دارند جزو سؤالات عینی می‌باشند.

سؤال سوم پژوهش: ضریب روایی سؤال‌های آزمون درس علوم تجربی پایه ی نهم استان کرمانشاه در خرداد ۱۳۹۵ چقدر است؟

جدول ۵: روایی سؤالات امتحان علوم تجربی پایه‌ی نهم استان کرمانشاه در خرداد ۱۳۹۵

ردیف	سؤال	تعداد متخصصین	حداقل ضریب لازم برابر جدول	ضریب روایی به دست آمده برابر فرمول	روایی دارد/ ندارد
۱	A	۸	۰/۸۵	۰/۵۰	ندارد
	B	۸	۰/۸۵	-۰/۷۵	ندارد
	C	۹	۰/۷۸	۰/۷۷	ندارد
	D	۹	۰/۷۸	۱	دارد
۲	A	۸	۰/۸۵	۰/۵۰	ندارد
	B	۸	۰/۸۵	۱	دارد
	C	۹	۰/۷۸	۰/۷۷	ندارد
	D	۸	۰/۸۵	۰/۵۰	ندارد
۳	A	۹	۰/۷۸	۰/۳۳	ندارد
	B	۹	۰/۷۸	۱	دارد
	C	۹	۰/۷۸	۰/۵۵	ندارد
	D	۹	۰/۷۸	-۰/۱۱	ندارد
	E	۹	۰/۷۸	۰/۵۵	ندارد
	F	۹	۰/۷۸	۰/۳۳	ندارد
	G	۹	۰/۷۸	۰/۵۵	ندارد
۴	A	۸	۰/۸۵	۰/۵۰	ندارد
	B	۸	۰/۸۵	۰/۵۰	ندارد
	C	۸	۰/۸۵	۱	دارد
۵		۹	۰/۷۸	-۰/۱۱	ندارد
۶		۹	۰/۷۸	۱	دارد
۷		۸	۰/۸۵	-۰/۲۵	ندارد
۸	الف	۹	۰/۷۸	۰/۳۳	ندارد
	ب	۹	۰/۷۸	۰/۷۷	ندارد
۹		۹	۰/۷۸	۱	دارد
۱۰		۹	۰/۷۸	۰/۳۳	ندارد
۱۱		۸	۰/۸۵	۱	دارد
۱۲		۸	۰/۸۵	۱	دارد
۱۳		۸	۰/۸۵	-۰/۵۰	ندارد
۱۴		۹	۰/۷۸	۱	دارد
۱۵		۸	۰/۸۵	۰/۵۰	ندارد
۱۶		۸	۰/۸۵	۱	دارد
۱۷		۸	۰/۸۵	۱	دارد
۱۸		۸	۰/۸۵	۱	دارد
۱۹		۸	۰/۸۵	۰/۵۰	ندارد
	شاخص روایی محتوایی آزمون		۰/۸۲	۰/۵۶	ندارد

برابر اطلاعاتی که از جدول ۵ به دست آمد، سؤال‌های شماره ۱(D)، ۲(B)، ۳(B)، ۴(C)، ۶، ۹، ۱۱، ۱۲، ۱۴، ۱۶، ۱۷ و ۱۸ روایی محتوایی دارند. به عبارتی این سؤال‌ها می‌توانند نمونه‌ای دقیق و درست از محتوا و مطالب کتاب علوم تجربی پایه نهم باشند؛ اما سؤال‌های شماره ۱(A)، ۱(B)، ۱(C)، ۲(A)، ۲(C)، ۲(D)، ۳(A)، ۳(C)، ۳(D)، ۳(E)، ۳(F)، ۳(G)، ۴(A)، ۴(B)، ۵، ۷، ۸ (الف)، ۸ (ب)، ۱۰، ۱۳، ۱۵ و ۱۹ روایی محتوایی ندارند؛ یعنی این سؤال‌ها نمی‌توانند به‌خوبی بیانگر محتوا و اهداف کتاب علوم پایه‌ی نهم باشند. همچنین شاخص روایی محتوایی آزمون نشان می‌دهد که این آزمون از روایی محتوایی کافی برخوردار نیست.

سؤال چهارم پژوهش: ضریب پایایی سؤال‌های آزمون درس علوم تجربی پایه‌ی نهم استان کرمانشاه در خرداد ۱۳۹۵ چقدر است؟

جدول ۶: پایایی سوالات امتحان علوم تجربی پایه‌ی نهم استان کرمانشاه در خرداد ۱۳۹۵

نوع سؤال‌ها	ردیف سؤال‌ها	پایایی با روش کودر- ریچاردسون	پایایی با روش آلفای کرانباخ
عینی	۱ (A,B,C,D) - ۲ (A,B,C,D)	۰/۱۲۶	-
	۳(A,B,C,D,E,F,G) - ۴(A,B,C) - ۸ (الف) و (ب)		
غیر عینی	۵، ۶، ۷، ۹، ۱۰، ۱۱، ۱۲، ۱۳، ۱۴، ۱۵، ۱۶، ۱۷، ۱۸ و ۱۹	-	۰/۸۹۲

با توجه به جدول ۶ پایایی سؤال‌های عینی بسیار پایین (۰/۱۲۶)، اما ضریب پایایی سؤال‌های غیر عینی بالا (۰/۸۹۲) بود.

سؤال پنجم پژوهش: سؤال‌های آزمون درس علوم تجربی پایه‌ی نهم استان کرمانشاه در خرداد ۱۳۹۵ ابعاد فرایندهای شناختی اندرسون و کراتول را چگونه پوشش داده‌اند؟

جدول ۷: تحلیل بعد فرایندهای شناختی سؤالات پایه نهم علوم تجربی استان کرمانشاه در خرداد ۱۳۹۵ بر اساس طبقه‌بندی شناختی اندرسون و کراتول

محتوای سؤالات	به یادآوردن	فهمیدن	به کار بستن	تحلیل کردن	ارزشیابی کردن	آفریدن
گاز اتن، سیاره ناهید، باکتری‌ها، خزه‌ها، گاز اوزون، سال نوری، طبقه‌بندی جانداران، رگبرگ‌ها، ورقه‌های سنگ‌کره، آب‌وهوای گذشته، ویروس ایدز، جانوران بی‌مه‌ره، بوم‌سازگان، نقش گیاهان، حرکت ورقه‌های سنگ‌کره، شهاب‌سنگ، سیارک، قمر، گسل، فسیل، صورت‌های فلکی، جلبک‌ها، غذای جلبک‌ها، حرکت رو به بالای آب، حشرات، پرواز پرندگان، پستانداران، لاک‌لاک‌پشت، زنجیره غذایی، الکترولیت، نیرو، فشار آب، مزیت مکانیکی، تندی	۵۸/۸۲٪	۲۹/۴۱٪	۸/۸۲٪	۲/۹۴٪	۰٪	۰٪

مطابق جدول ۷، فرایند شناختی به یادآوردن با ۲۰ سؤال (۵۸/۸۲٪) بیشترین سهم را در بین دیگر فرایندها به دست آورده است. ۱۰ سؤال (۲۹/۴۱٪) در حوزه فرایند شناختی فهمیدن، ۳ سؤال (۸/۸۲٪) مربوط به فرایند شناختی به‌کار بستن و ۱ سؤال (۲/۹۴٪) در حوزه فرایند شناختی تحلیل کردن طراحی شده‌اند. هیچ‌یک از سؤالات در حیطه فرایندهای شناختی ارزشیابی کردن و آفریدن طراحی نشده است.

سؤال ششم پژوهش: سؤال‌های آزمون درس علوم تجربی پایه‌ی نهم استان کرمانشاه در خرداد ۱۳۹۵ ابعاد دانش طبقه‌بندی شناختی اندرسون و کراتول را چگونه پوشش داده‌اند؟

جدول ۸: تحلیل بعد دانش سؤالات پایه نهم علوم تجربی استان کرمانشاه در خرداد ۱۳۹۵ بر اساس طبقه‌بندی شناختی اندرسون و کراتول

محتوای سؤالات	دانش امور واقعی	دانش مفهومی	دانش روندی	دانش فراشناختی
گاز اتن، سیاره ناهید، باکتری‌ها، خزه‌ها، گاز اوزون، سال نوری، طبقه‌بندی جانداران، رگبرگ‌ها، ورقه‌های سنگ‌کره، آب‌وهوای گذشته، ویروس ایدز، جانوران بی‌مه‌ره، بوم‌سازگان، نقش گیاهان، حرکت ورقه‌های سنگ‌کره، شهاب‌سنگ، سیارک، قمر، گسل، فسیل، صورت‌های فلکی، جلبک‌ها، غذای جلبک‌ها، حرکت رو به بالای آب، حشرات، پرواز پرندگان، پستانداران، لاک‌لاک‌پشت، زنجیره غذایی، الکترولیت، نیرو، فشار آب، مزیت مکانیکی، تندی	۶۷/۶۴٪	۱۱/۷۶٪	۲۰/۵۸٪	۰٪

با توجه به جدول ۸، تعداد ۲۳ سؤال (۶۷/۶۴٪) در بعد دانش امور واقعی، تعداد ۴ سؤال (۱۱/۷۶٪) در بعد دانش مفهومی، تعداد ۷ سؤال (۲۰/۵۸٪) در بعد دانش روندی طراحی شده‌اند؛ اما هیچ سؤالی در بعد دانش فراشناختی جای نگرفته است.

بحث و نتیجه‌گیری

در بررسی یافته‌های ویژگی‌های روان‌سنجی - ضریب دشواری، ضریب تمیز، ضریب روایی و پایایی - و میزان پوشش سطوح مختلف حیطه شناختی طبقه‌بندی اندرسون و کراتول در این پژوهش، مشخص گردید که سؤال‌های آزمون دارای ضریب دشواری متوسط و ضریب تمیز بالایی بودند. همچنین بیشتر سؤالات در طبقه‌بندی بعد فرایند شناختی در سطح به یادآوردن و از لحاظ تقسیم‌بندی بعد دانش، بیشترین سؤالات در بعد دانش امور واقعی انتخاب شده بودند. در خصوص یافته‌های حاصل از پرسش اول باید گفت: «متخصصان آزمون‌سازی (مهرنر و لمان^۱، ۱۹۸۴، ثورانیدیک و هیگن^۲، ۱۹۸۶) ضریب دشواری بین ۰/۳۰ تا ۰/۷۰ برای آزمون‌های معلم ساخته پیشنهاد می‌کنند. آن‌ها به تجربه دریافتند سؤالاتی که دشواری آن‌ها بالاتر از ۰/۷۰ است بسیار آسان و سؤالاتی که ضریب دشواری آن‌ها پایین‌تر از ۰/۳۰ است برای دانش آموزان بسیار دشوارند» (میرآقایی و همکاران، ۱۳۹۳: ۱۱۴). در بررسی سؤال‌های این آزمون ضریب دشواری تعداد ۲۱ سؤال که ۱۰ نمره از بارم ۱۵ نمره‌ای کل آزمون را به خود اختصاص داده‌اند، ۰/۳۰ تا ۰/۷۰ بود، ضریب دشواری مناسبی داشتند و تعداد ۱۳ سؤال که ۵ نمره از ۱۵ نمره کل آزمون را داشتند ضریب دشواری بالای ۰/۷۰ داشتند که می‌توان آن‌ها را در گروه سؤال‌های آسان گذاشت. با توجه به این که حدود ۶۶٪ سؤال‌های این آزمون ضریب دشواری مناسب داشته‌اند می‌توان گفت ضریب دشواری این آزمون مناسب هست؛ زیرا سؤال‌های خوب سؤال‌هایی هستند که ضریب دشواری متوسطی داشته باشند. این نتیجه با یافته‌های امیری‌مجد (۱۳۹۴) و شوند غربی (۱۳۸۹) هم‌سو بوده و با نتایج پژوهش‌های ریهاوی (۱۳۸۹) و میر آقایی و همکاران (۱۳۹۳) ناهم‌سو است. «در مورد ضریب تمیز سؤالات، متخصصان، ضریب تمیز ۰/۲۰ را برای سؤالات آزمون‌های معلم ساخته پیشنهاد می‌کنند» (میرآقایی و همکاران، ۱۳۹۳: ۱۱۴). هرچه ضریب تمیز سؤالی بالاتر و به عدد یک نزدیک‌تر باشد، آن سؤال قدرت بیشتری در تمایز بین گروه قوی و ضعیف دارد. در بررسی این آزمون سؤال ۹ با ضریب تمیز ۰/۹۳ بهترین

¹. Mehrens & Lehman

². Thorandike & Hagen

سؤال و سؤال (A) ۴ با ضریب تمیز صفر ضعیف‌ترین سؤال برای تمایز بین گروه‌های قوی و ضعیف آزمون‌شوندگان است. همچنین تعداد ۵ سؤال از مجموع ۳۴ سؤال با بارم ۱/۲۵ از بارم ۱۵ نمره‌ای کل آزمون دارای ضریب تمیز کم‌تر از ۰/۲۰ بودند که می‌توان گفت کمی بیش‌تر از ۸٪ سؤال‌ها ضریب تمیز ضعیفی داشته و حدود ۹۲٪ آن‌ها دارای ضریب تمیز بالایی بودند و بنابراین این آزمون دارای ضریب تمیز بالا است؛ زیرا سؤال‌های خوب ضریب تمیز بالا دارند. این نتیجه با یافته‌های ربیپهاوی (۱۳۸۹)، میر آقایی و همکاران (۱۳۹۳) و قسمتی از تحقیق گواهی و همکاران (۱۳۹۱) هم‌سو بوده و با یافته‌های شونند غربی (۱۳۸۹)، امیری مجد (۱۳۹۴) و بخشی از سؤال‌های گواهی و همکاران (۱۳۹۱) متفاوت است. در مورد پرسش سوم یعنی میزان روایی سؤال‌های آزمون طبق نتایجی که به دست آمد؛ تعدادی از سؤال‌ها دارای روایی محتوایی بودند که می‌توانند معرف خوبی برای محتوا و اهداف کتاب علوم تجربی پایه نهم باشند؛ اما تعدادی دیگر روایی لازم را نداشتند. از آنجاکه در بررسی روایی این آزمون در بعضی سؤالات ۸ متخصص و در برخی دیگر ۹ متخصص دیدگاه خود را ارائه داده‌اند، طبق جدول ۱ (جدول روایی محتوایی لاشه) روایی لازم برای این آزمون ۰/۸۲ است. در مجموع روایی به‌دست‌آمده برای کل سؤال‌ها ۰/۵۶ بود که این مقدار کم‌تر از مقدار روایی لازم یعنی ۰/۸۲ بود؛ لذا می‌توان اظهار داشت آزمون از روایی محتوایی لازم برخوردار نیست. در پاسخ به چرایی این امر می‌توان بیان کرد که برای بررسی روایی محتوایی یک آزمون پیشرفت تحصیلی، لازم است هدف‌های آموزشی آن درس که آزمون برای اندازه‌گیری آن ساخته شده است مشخص و به‌صورت رفتاری با دقت لازم تعریف شود. تا بر اساس آن‌ها داوری روی محتوای آزمون صورت گیرد. برای این کار مهم‌ترین و بهترین روش در تهیه آزمون‌های پیشرفت تحصیلی طراحی جدول مشخصات^۱ برای موضوع یا درسی است که می‌خواهیم از آن آزمون تهیه کنیم؛ و این کار روایی سؤال‌ها را بالا می‌برد، می‌توان گفت احتمالاً طراح سؤال این مهم را مدنظر قرار نداده است. ضمناً کتاب علوم تجربی پایه‌ی نهم فاقد اهداف آموزشی و درسی در ابتدای فصل‌هایش است. نتایج به‌دست‌آمده از پرسش چهارم در مورد پایایی سؤال‌های آزمون نشان داد پایایی به‌دست‌آمده برای سؤال‌های عینی (چهار جوابی، صحیح-غلط، جور کردنی و جای خالی) بسیار پایین و به میزان ۰/۱۲۶ بود. این یافته‌ها با یافته‌های میر آقایی و همکاران (۱۳۹۳) و امیری مجد (۱۳۹۴) هم‌سو نبود؛ اما پایایی سؤال‌های غیر عینی (تشریحی، تعریف مفاهیم و مسئله‌ها) ۰/۸۲۹ بود که نشان‌دهنده ثبات

^۱. Table of specifications

و دقت خوب این سؤال‌ها است؛ نتایج این بخش با یافته‌های میر آقایی و همکاران (۱۳۹۳) و امیری مجد (۱۳۹۴) هم‌سو است. متخصصان برای آزمون‌های معلم ساخته پایایی بالاتر از ۰/۸۰ را مناسب می‌دانند. بر این اساس سؤال‌های عینی از پایایی برخوردار نبوده و می‌توان گفت طراح سؤال در انتخاب آن‌ها دقت کافی را نداشته است. البته ضریب تمیز پایین نصف سؤالات عینی می‌تواند دلیلی علمی برای پایایی پایین آن‌ها باشد. ولی سؤال‌های غیر عینی دارای ثبات و پایایی لازم و قابل قبول بودند. در تبیین این نتیجه می‌توان اظهار نمود که ضریب تمیز بالا و سطح دشواری متوسط این سؤالات می‌تواند دلیلی محکم برای پایایی بالای آن‌ها باشد. در خصوص پرسش پنجم نتایج نشان داد بیش‌ترین تعداد سؤالات این آزمون بر سطح «به یاد آوردن» تمرکز داشتند. پس از آن بیش‌ترین سؤالات در سطح «فهمیدن» طراحی شده بودند؛ و تعداد بسیار کمی نیز در سطوح «به کار بستن» و «تحلیل کردن» انتخاب شده بودند؛ اما هیچ سؤالی در سطوح «ارزشیابی کردن» و «آفریدن» طرح نشده بود. با توجه به این که طبقه‌های بعد فرایند شناختی به صورت سلسله‌مراتبی و از عینی به انتزاعی و از ساده به پیچیده تنظیم شده‌اند، نتیجه می‌گیریم طراح آزمون سعی بر آن داشته است از سؤال‌های ساده‌تر و عینی‌تر استفاده نماید. این نتایج با یافته‌های ربی‌هاوی (۱۳۸۹)، میر آقایی و همکاران (۱۳۹۳)، درس‌های شیمی و علوم اجتماعی در تحقیق امیری مجد (۱۳۹۴) هم‌سو بود ولی با نتایج پژوهش جعفری (۱۳۸۸)، زمانی (۱۳۹۱) و ادواردز (۲۰۱۰) ناهم‌سو بود. برای چرایی این یافته می‌توان به دلایل ذیل اشاره نمود: تدریس سنتی، غیرفعال و معلم محور بیشتر دبیران، تراکم زیاد کلاس‌ها، نبود امکانات و فضای آزمایشگاهی، ساعات کم اختصاص یافته به درس علوم تجربی (۳ ساعت در هفته)، حساسیت زیاد ادارات آموزش و پرورش، مدیران، والدین و دانش‌آموزان به بالا بودن نمرات و افزایش درصد قبولی. مجموع عوامل بالا معلمین و طراحان سؤالات و آزمون‌های پیشرفت تحصیلی را به سمت طراحی پرسش‌های سهل و ساده و بیشتر در سطوح پایین حیطه‌ی شناختی سوق می‌دهد. بودن نمرات دانش‌آموزان نتایج حاصل از پرسش ششم نشان داد از مجموع ۳۴ سؤال این آزمون ۲۳ سؤال آن در بعد دانش امور واقعی، هم‌سو با یافته‌های هانگ و همکاران (۲۰۰۶)، امیری مجد (۱۳۹۴) ۴ سؤال در بعد دانش مفهومی، ناهم‌سو با یافته‌های هانگ و همکاران (۲۰۰۶) و ۷ سؤال در بعد دانش روندی ناهم‌سو با یافته‌های هانگ و همکاران (۲۰۰۶) و هم‌سو با نتایج امیری مجد (۱۳۹۴) طراحی شده‌اند؛ اما هیچ سؤالی در بعد دانش فراشناختی انتخاب نشده است. این قسمت با یافته‌های هانگ و همکاران (۲۰۰۶) ناهم‌سو و با یافته‌های امیری مجد (۱۳۹۴) هم‌سو بود.

امانی تهرانی و همکاران (۱۳۹۵) بر اساس تجارب اولیه پژوهش‌گر در طول بیست سال برنامه‌ریزی و تألیف کتاب‌های درسی و آموزش ضمن خدمت معلمان و نیز مطالعه پیشینه ادبیات و پژوهش‌های بین‌المللی، در پژوهش خود با عنوان "طراحی و تدوین مدلی کارآمد برای آموزش علوم تجربی در دوره‌ی اول متوسطه" طرحی شش مرحله‌ای که تمام عناصر برنامه درس علوم تجربی را پوشش می‌دهد، ارائه دادند. این مراحل شامل: مرتبط ساختن موضوعات یادگیری با زندگی، بهره‌گیری از رویکرد ساختن‌گرایی، استفاده از راهبرد یادگیری مشارکتی و همکاری گروهی، تأکید بر فعالیت‌های دست‌ورزی و ذهن‌ورزی (فکری)، تأکید بر ارزشیابی تکوینی و استفاده از شیوه‌های جدید در ارزشیابی پایانی. از جمله راهکارهای آن‌ها در رابطه با به‌کارگیری روش‌های جایگزین و متفاوت مانند افزودن ارزشیابی عملکردی در ارزشیابی پایانی عبارت‌اند از: استفاده از ارزشیابی کتبی که ویژگی‌های زیر در آن‌ها رعایت شده باشد: بها دادن به سؤالات کاربردی؛ استفاده از سؤالات درک مطلب؛ استفاده از سؤالات تفکر برانگیز؛ استفاده از سؤالات انتخابی و امتیازی و استفاده از سؤالاتی که دانش‌آموزان آن‌ها را طرح کنند و پاسخ آن عیناً در کتاب نباشد. همچنین به‌تمامی معلمان و طراحان سؤالات آزمون‌های پایانی پیشنهاد می‌شود از جدول مشخصات آزمون استفاده نمایند.

کتابنامه

- امیری‌مجد، حشمت اله. (۱۳۹۴). تحلیل مواد آزمون پایان سال دروس ریاضی، فیزیک، شیمی و علوم اجتماعی سال اول دبیرستان شهرستان سرابله با توجه به طبقه‌بندی اهداف شناختی اندرسون و کراتول. پایان‌نامه کارشناسی ارشد. گروه علوم تربیتی. دانشگاه ملایر.
- بست، جان. (۱۳۹۰). روش‌های تحقیق در علوم تربیتی و رفتاری. (ترجمه حسن پاشا شریفی و نرگس طالقانی). تهران: رشد. (تاریخ انتشار به زبان اصلی ۱۹۸۳).
- بقائی، رحیم؛ رسولی، داود؛ نادری، ژاله؛ شمس، شمس‌الدین و فیضی، آرام. (۱۳۹۵). ارزیابی سؤالات چندگزینه‌ای پایان‌ترم دانشجویان پرستاری دانشگاه علوم پزشکی ارومیه. *مجله دانشکده پرستاری و مامایی*، ۴، ۲۹۹-۲۹۱.
- پاشا شریفی، حسن، شریفی، نسترن. (۱۳۹۵). *اصول روان‌سنجی و روان‌آزمایی (ویرایش پنجم)*. تهران: انتشارات رشد.
- جعفری، طیبه. (۱۳۸۸). تحلیل محتوای سؤال‌های کتاب‌های ریاضی دوم و سوم دبیرستان در رشته ادبیات و علوم انسانی، بر اساس طبقه‌بندی تجدیدنظر شده بلوم در حیطه شناختی. پایان‌نامه کارشناسی ارشد. گروه علوم تربیتی. دانشگاه علامه طباطبائی.
- حسینی، محمدحسن. (۱۳۸۶). *سنجش و اندازه‌گیری*. تهران: فرهنگ نشر.
- حسینی، محمد و کیامنش، علیرضا. (۱۳۸۸). نقد و بررسی نظام ارزشیابی تحصیلی ایران از منظر ارزشیابی نظام آموزشی. فصلنامه نوآوری‌های آموزشی، سال ۸، شماره ۳۰، ۱۰۳-۷۶.
- خاکی، غلامرضا. (۱۳۹۱). *روش تحقیق با رویکرد پایان‌نامه نویسی*. تهران: نشر فوژان.
- خورشیدی، عباس و ملکشاهی راد، محمدرضا. (۱۳۸۵). *ارزشیابی آموزشی (چاپ دوم)*. تهران: سیسترون.
- ریبهاوی، فیصل. (۱۳۸۹). بررسی میزان انطباق سؤالات امتحانی دروس مراکز تربیت‌معلم استان خوزستان با اصول روان‌سنجی. پایان‌نامه کارشناسی ارشد. دانشگاه شهید چمران اهواز.
- رضوانی، رضا و زمانی، گران ناز. (۱۳۹۱). ارزیابی هم ترازوی آزمون‌های سراسری، محتوای کتاب‌ها و اهداف آموزشی رشته آموزش زبان انگلیسی بر اساس اهداف آموزشی اندرسون و کراتول. *فصلنامه رهیافتی نو در مدیریت آموزشی*، دوره ۳، شماره ۳، ۱۲۰-۱۰۵.
- رضوانی، رضا و حق‌شناس، بهناز. (۱۳۹۳). ارزیابی هم ترازوی کتاب‌های کارشناسی زبان انگلیسی تخصصی رشته‌های علوم انسانی انتشارات سمت و استانداردهای آموزشی مربوطه بر اساس اهداف آموزشی اندرسون و کراتول. *فصلنامه رهیافتی نو در مدیریت آموزشی*، دوره ۵، شماره ۴، ۱۱۰-۹۵.
- زمانی، گران ناز. (۱۳۹۱). بررسی هم ترازوی آزمون کارشناسی ارشد رشته‌های ترجمه و آموزش زبان انگلیسی در ایران و سرفصل‌های آموزشی و کتاب‌های رسمی کشور با استفاده از طبقه‌بندی اهداف آموزشی اندرسون و کراتول. پایان‌نامه کارشناسی ارشد، دانشکده ادبیات و علوم انسانی، دانشگاه یاسوج.
- سرمد، زهره، بازرگان، عباس و حجازی، الهه. (۱۳۹۵). *روش‌های تحقیق در علوم رفتاری*. تهران: آگاه.
- سیف، علی‌اکبر. (۱۳۹۴). *اندازه‌گیری، سنجش و ارزشیابی آموزشی*. تهران: نشر دوران.

کاظمی، اشرف و احسان پور، سهیلا. (۱۳۸۹). تحلیل سؤالات امتحانات دروس تخصصی نظری دانشجویان مامایی دانشگاه علوم پزشکی اصفهان. *مجله ایرانی آموزش در علوم پزشکی (ویژه‌نامه توسعه آموزش)*، دوره ۵، شماره ۱۰، ۶۵۰-۶۴۳.

کیامنش، علیرضا. (۱۳۹۳). روش‌های ارزشیابی آموزشی. تهران: پیام نور.
گواهی، محمد؛ علیزاده، حجت اله و اورانوس بخت، میترا. (۱۳۹۱). تحلیل آزمون هماهنگ کشوری درس فیزیک ۱ و آزمایشگاه. *مجله آموزش فیزیک*، دوره ۳، شماره ۲۸.
میرآقایی، علی عباس؛ سپاسی، حسین؛ مهاجران، بهناز و قلعه‌ای، علیرضا. (۱۳۹۴). بررسی و تحلیل شاخص‌های روان‌سنجی و سطوح حیطة شناختی سؤالات امتحانات نهایی دروس ریاضیات و علوم پایه سوم راهنمایی شهرستان خرم‌آباد. *مجله‌ی روان‌شناسی مدرسه*، دوره ۴، شماره ۳، ۱۱۸-۱۰۲.
نوروزی، ابوالقاسم. (۱۳۸۶). مقدمه‌ای بر سنجش و اندازه‌گیری در علوم تربیتی. (چاپ دوم). تهران: آوای نور.

ولف، ریچارد. (۱۳۹۵). ارزشیابی آموزشی (مبانی سنجش توانایی و بررسی برنامه‌ها). (ترجمه علیرضا کیامنش). تهران: مرکز نشر دانشگاهی. (تاریخ انتشار اثر به زبان اصلی، ۱۹۸۴).
هومن، حیدر علی. (۱۳۸۹). *پایه‌های پژوهش در علوم رفتاری (شناخت روش علمی)*. تهران: پارسا.

Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Longman.

Edwards, N. (2010). An analysis of the alignment of the Grade 12 Physical Sciences examination and the core curriculum in South Africa. *South African Journal of Education*, 30(4), 571-590.

Gado, I. (2004). Psychometric properties of teacher-made science tests used in national examinations for middle-grade students in Benin (West Africa): A longitudinal study.

Hung, S. I., Chung, W. H., Jee, S. H., Chen, W. C., Chang, Y. T., Lee, W. R., & Wong, T. W. (2006). Genetic susceptibility to carbamazepine-induced cutaneous adverse drug reactions. *Pharmacogenetics and genomics*, 16(4), 297-306.

Kara, Y., & Cepni, S. (2011). Investigation the alignment between school learning and entrance examinations through item analysis. *Journal of Baltic Science Education*, 10 (2), 73-86.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41(4), 212-218.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel psychology*, 28(4), 563-575.

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79(4), 1332-1361.

Nitko, A.J. (1983). *Educational tests and measurement*. New York: Harcourt Brace Jovanovich.

Payne, D.A. (2003). *Applied educational assessment (2d ad.)*. Wandsworth.

Shultz, k. s., & Whitney, D. J. (2005). *Measurement theory in action*. London: sage.

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7-25.

Whitney, D. R., & Sabers, D. L. (1970). *Improving essay examinations: Use of item analysis*. University of Iowa.

