

بررسی و تحلیل شاخص‌های روان‌سنجی و سطوح حیطة شناختی سوالات امتحانات نهایی دروس ریاضیات و علوم پایه سوم راهنمایی شهرستان خرم‌آباد

علی عباس میرآقایی^۱، حسین سپاسی^۲، بهناز مهاجران^۳ و علیرضا قلعه‌ای^۴

چکیده

هدف اصلی این تحقیق بررسی و تحلیل شاخص‌های روان‌سنجی و سطوح حیطة شناختی سوالات امتحانات نهایی دروس ریاضیات و علوم پایه سوم راهنمایی شهرستان خرم‌آباد بود. از هر کدام از جامعه‌های آماري برگه‌های نمره‌گذاری شده دروس ریاضیات و علوم ۱۰۰۰ برگه به روش تصادفی ساده انتخاب شدند و برای تعیین سطوح حیطة ای شناختی نمونه ای به حجم ۲۵ نفر از دبیران ریاضیات و علوم با روش تصادفی ساده انتخاب و از روش توافقی دلفی دو مرحله‌ای استفاده شد. نتایج محاسبات نشان داد که ضریب دشواری سوالات امتحانی مناسب نبودند ولی ضریب تمیز سوالات امتحانی به استثنای تعداد کمی از سوالات مناسب بودند. بررسی پایایی و روایی محتوایی نشان داد که سوالات دارای پایایی و روایی محتوایی بوده‌اند. نتایج همچنین نشان داد که سوالات امتحان درس ریاضیات به اندازه‌گیری سطوح بالای حیطة شناختی (ترکیب و ارزشیابی) اختصاص نداشتند.

واژه‌های کلیدی: ضریب دشواری، ضریب تمیز، ضریب پایایی، روایی محتوایی، حیطة شناختی

۱. نویسنده‌ی رابط: کارشناسی ارشد تحقیقات آموزشی، دانشگاه شهید چمران اهواز (a.a.miraghaie@gmail.com)

۲. استاد گروه علوم تربیتی، دانشگاه شهید چمران اهواز

۳. استادیار علوم تربیتی، دانشگاه ارومیه

۴. استادیار علوم تربیتی، دانشگاه ارومیه

تاریخ دریافت مقاله: ۹۳/۳/۱۱

تاریخ پذیرش مقاله: ۹۳/۸/۱۸

مقدمه

امتحان یکی از وسایل و تسهیلاتی است که به کمک آن معلم می‌تواند، نتایج تغییراتی را که در فرد ایجاد شده است، ارزشیابی کند و منظور از آن در واقع این است که معلوم شود آیا تغییرات معینی در فرد به وجود آمده است یا نه؟ آیا این تغییرات، مطلوب و رضایت بخش است؟ و میزان این تغییرات نزد افراد مختلف تا چه اندازه متفاوت است؟ (باقری، ۱۳۸۵).

در بررسی منابع ایجاد انگیزه در دانش‌آموزان و تشخیص نقاط قوت و ضعف آنان، آگاه کردن دانش‌آموزان از میزان عملکرد تحصیلی و اعلام نتایج امتحانات به اولیاء، پرورش استعدادها، پرورش روحیه تلاش، ابتکار و رقابت‌های سالم بین دانش‌آموزان، تشخیص استعدادها و علائق دانش‌آموزان به منظور راهنمایی آنان در فعالیت‌های آینده و تحصیلات پیش‌دانشگاهی و دانشگاهی و نهایتاً، حصول اطمینان از تحقق شرایط لازم به منظور ارتقای دانش‌آموزان به پایه تحصیلی بالاتر از جمله اهداف امتحانات ذکر شده است (سیف، ۱۳۷۲؛ شریفی، ۱۳۷۴؛ مهرنر و لهمان^۱، ۱۹۸۴؛ بوام‌گارتنر و جکسون^۲، ۱۹۹۱؛ گیج^۳، ۱۹۹۲؛ پیچ و پیترسون^۴، ۲۰۰۳).

متأسفانه با اینکه شاهد اجرای آزمون‌های بسیاری در نظام‌های آموزشی هستیم ولی تاکنون از آزمون‌های مناسب، کم و از آزمون‌های بسیار مناسب خیلی کمتر استفاده کرده‌ایم. بسیاری از سازندگان آزمون بدون اینکه از آموزش کافی در سنجش و اندازه‌گیری و شیوه ساخت آزمون‌ها برخوردار باشند به ساخت آزمون مبادرت می‌کنند، به ویژه این امر در مورد معلمان که اطلاعات آنها درباره ساخت و ویژگی‌های امتحان به زمان دانشجویی خلاصه می‌شود شایع‌تر است. علاوه بر این، وقتی از دانش‌آموزان درباره سؤالات امتحانی معلمان‌شان سؤال می‌شود پاسخ می‌دهند که سؤالات نامفهوم بودند، سؤالات بسیار دشوار بودند، سؤالات به پرسش مفاهیم ساده و پیش‌پا افتاده که فقط

- 1 . Mehrens & Lehman
- 2 . Baumgartner & Jackson
- 3 . Gage
- 4 . Page & Petron

به حفظ طوطی وار مطالب نیاز داشتند اختصاص داشت (مهرنزو و لهما، ۱۹۸۴؛ گادو، ۲۰۰۴). سازندگان آزمون باید توجه کنند که اگر سوالات آزمون‌ها نامفهوم و یا بسیار دشوار یا بسیار آسان باشند، نمرات حاصل از آنها منعکس کننده توانایی واقعی دانش‌آموزان در موضوع مورد مطالعه نخواهند بود و سوالات آزمون قادر نیستند بین دانش‌آموزان قوی و ضعیف تفاوت بگذارند و چنانچه دانش‌آموزان تلاش زیادی برای یادگیری بکنند ولی به نتایج مورد انتظارشان دست نیابند، پس از یک دوره زمانی انگیزه آنان برای یادگیری پایین خواهد آمد و اگر این عدم موفقیت تکرار شود دانش‌آموزان به نوعی درماندگی مبتلا خواهند شد و امکان دارد از تلاش برای جبران نقاط ضعف خود باز بمانند (گادو^۱، ۲۰۰۴). علاوه بر این که می‌توان از نتایج آزمون‌ها برای تعیین پیشرفت تحصیلی دانش‌آموزان و فراهم کردن انگیزه برای یادگیری آنها استفاده کرد، برای ارزشیابی عملکرد معلمان در کلاس‌های درس نیز کاربرد دارد. بنابراین، ضروری است که شیوه و کیفیت وسایل سنجش مورد ارزشیابی قرار گیرد. از طریق این ارزشیابی می‌توان به نقاط قوت و ضعف ساخت سوالات و خطاهای ناشی از قضاوت طراح سؤال پی‌برد. متخصصان آزمون‌سازی برای ارزشیابی سوالات 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 سوالات امتحانی افزایش دهند و در کیفیت سوالات آزمون‌ها که در آینده مورد استفاده قرار خواهند داد، تجدیدنظر کنند.

جندقی و شاطریان (۲۰۰۸) در تحقیقی به بررسی شاخص‌های روایی، پایایی و دشواری آزمون‌های معلم ساخته در درس شیمی سال اول دانشگاه قم پرداختند. نتایج حاکی از آن بودند که ضریب پایایی و روایی سوالات و ضریب دشواری سوالات بالا بود. نتایج جانبی تحقیق نشان داد که ۳۹ درصد از سوالات در سطح دانش، ۵۷ درصد در سطح فهم و ۴ درصد در سطح کاربرد بودند. شرقی (۱۳۸۳) در تحقیقی به بررسی شاخص‌های روان‌سنجی سوالات امتحانات نهایی

دروس عربی، حسابان و زیست‌شناسی دانش‌آموزان پسر پایه سوم متوسطه سه منطقه متفاوت اجتماعی-اقتصادی استان خوزستان پرداخت. نتایج یافته‌هایی وی نشان دادند که ضرایب دشواری، تمیز و همسانی درونی سؤالات امتحانی نهایی دروس عربی، حسابان و زیست‌شناسی در مقایسه با ملاک‌ها از وضعیت مناسبی برخوردار بودند. سپاسی (۱۳۸۳) در تحقیقی به بررسی و تحلیل سطوح حیطه‌شناختی و شاخص‌های روان‌سنجی سؤالات امتحانات نهایی دروس نهایی عربی، حسابان، و زیست‌شناسی دانش‌آموزان دختر پایه‌ای سوم متوسطه در مناطق برخوردار، نیمه برخوردار و غیر برخوردار استان خوزستان پرداخت. نتایج همچنین نشان دادند که در درس زیست‌شناسی ۵۶/۲۵ از آزمون به اندازه‌گیری سطح دانش و ۱۵ درصد از آزمون به اندازه‌گیری سطح فهم، ۵ درصد به اندازه‌گیری سطح کاربرد و ۲۳/۷۵ به اندازه‌گیری سطح تحلیل پرداخته بودند. در درس حسابان نیز بیشترین بارم، مربوط به سطح فهم با ۴۰ درصد بود. در درس عربی به ترتیب ۷/۵ درصد، ۴۵ درصد، ۲۵ درصد، ۲۲/۵ درصد از بارم آزمون به اندازه‌گیری سطوح دانش، فهم، کاربرد و تحلیل پرداخته بودند. تاگاموری و بی‌شاب^۱ (۱۹۹۴) و زانک و بری^۲ (۱۹۹۵) مطالعاتی در زمینه شیوه‌های ساخت سؤالات امتحانات معلمان دوره‌های مختلف آموزشی در ارتباط با مدارک تحصیلی، تجربه و سابقه تدریس آنان انجام دادند. نتایج یافته‌های آنان نشان داد که دبیران با مدارک کارشناسی و کارشناسی ارشد در ساخت سؤالات آزمون‌های معلم ساخته از معلمان با مدارک دیپلم در دوره راهنمایی و ابتدایی موفق‌تر عمل می‌کنند. مک‌موریس و بوترید^۳ (۱۹۹۳) در تحقیقی به بررسی آزمون‌های کلاسی در دروس ریاضی و علوم سال اول و دوم راهنمایی پرداختند. آنها یک نمونه ۴۱ نفری را انتخاب کردند. نتایج نشان دادند که معلمان از همه شکل‌های سؤال در آزمون‌های کلاسی‌شان استفاده می‌کنند، با این وجود، معلمان علوم

-
1. Tagomori & Bishop
 2. Zhang & Burry
 - 3 . McMorris & Boothroyd

سؤالات چند گزینه‌ای و معلمان ریاضی سؤالات محاسباتی را ترجیح داده بودند. مرتلر^۱ (۱۹۹۹) در تحقیقی به بررسی تصورات معلمان از روایی و پایایی آزمون‌های کلاسی پرداخت. نتایج نشان دادند که معلمان زمان بسیار برای تحلیل‌های آماری داده‌های بدست‌آمده از آزمون‌هایشان اختصاص نمی‌دهند. همچنین نتایج حاکی از آن بودند که علیرغم این که بسیاری از معلمان بیان کرده بودند که سنجش‌های آنها دارای روایی و پایایی است، و زمان زیادی را برای تعیین روایی و پایایی اختصاص می‌دهند، اما تحلیل روش‌های که آنها برای تعیین روایی و پایایی به کار می‌برند، نشان داد که آنها برای تعیین روایی و پایایی آزمون‌هایشان از روش‌های معتبر استفاده نمی‌کنند. فلمینگ و چامبرز^۲ (۱۹۸۳)، به نقل از اوشر و کربی^۳، (۱۹۹۰) در تحقیقی به بررسی میزان رعایت سطوح حیطة شناختی و انواع فرم‌های آزمون‌ها معلم ساخته پرداختند. نتایج تحقیق حاکی از آن بود که کمتر از ۲۰ درصد سؤالات آزمون‌های معلم ساخته بالاتر از سطح دانش بودند، و اکثریت سؤالات در فرم کوتاه پاسخ بودند، از فرم‌های جور کردنی، چندگزینه‌ای و صحیح - غلط به مراتب کمتر استفاده شده بود، و از فرم سؤال انشایی استفاده نشده بود. گادو (۲۰۰۴)، پایان نامه دکتری، به بررسی شاخص‌های روان‌سنجی آزمون‌های ملی معلم ساخته (آزمون‌های نهایی) فیزیک و شیمی سال‌های اول، دوم و سوم دبیرستان در بنین 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 برگه‌ای 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 توجه با مطالب فوق می‌توان گفت که همه سؤالاتی که توسط معلمان برای اندازه‌گیری پیشرفت دانش‌آموزان ساخته می‌شود، بویژه سؤالات امتحان نهایی که پیشرفت تحصیلی تعداد

1. Mertler
2. Fleming & Chambers
3. Oescher & Kirby

زیادی از دانش‌آموزان را اندازه‌گیری می‌کنند و نتایج آنها دارای پیامدهای بسیاری است، لازم است که شاخص‌های روان‌سنجی (دشواری، تمیز، پایایی و روایی محتوایی) و اهداف آموزشی که سؤالات این امتحانات می‌سنجند، مورد بررسی قرار گیرند. تحقیق حاضر به بررسی و تحلیل شاخص‌های روان‌سنجی و سطوح حیطه‌شناختی سؤالات امتحانات دروس نهایی ریاضیات و علوم پایه سوم راهنمایی شهرستان خرم‌آباد پرداخت. برای انجام این تحقیق می‌توان چهار دلیل زیر را ذکر کرد: اول، مدل بازخورد نشان می‌دهد که باید کیفیت وسایل سنجش مورد ارزشیابی قرار گیرد. چنانچه عملکرد دانش‌آموزان پایین‌تر از حد انتظار ارزیابی شود، شاید به جای این که به یادگیری غیر کافی از طرف یادگیرندگان مربوط گردد به دلیل ضعف روش‌های ارزشیابی باشد. پس برای اینکه نمره‌ها برای هدفی که در اصل مورد نظر بوده ارزشمند باشند، آزمون‌ها باید مورد ارزشیابی قرار گیرند. دوم، اطلاع و آگاهی معلمان از عملکرد ضعیف برخی سؤالات امتحان توانایی بیشتری را برای کنترل بحث‌های کلاسی در اختیار معلمان قرار می‌دهد. سوم، آزمون و نمره سؤالات می‌توانند ضعف‌های ساخت سؤالات و خطاهای ناشی از قضاوت نویسنده سؤال را روشن سازند. معلمان می‌توانند با برطرف کردن این نقائص و قابل استفاده مجدد نمره‌ها، مهارت‌های نوشتن سؤال خود را پیشرفت دهند و در کیفیت آزمونی که در آینده مورد استفاده قرار خواهد گرفت تجدید نظر کنند (ایبل^۱، ۱۹۹۱، ترجمه سپاسی، ۱۳۸۱). چهارم، با تعیین سطح حیطه شناختی که هر کدام از سؤالات امتحان اندازه‌گیری می‌کنند، می‌توان در مورد اهدافی که دانش‌آموزان به آن دست یافته‌اند، اطلاع پیدا کرد (سعیدی آرانی، ۱۳۷۷).

روش

روش این تحقیق توصیفی و از نوع ارزشیابی بود. هدف روش تحقیق ارزشیابی، جمع‌آوری اطلاعات تفصیلی به عنوان پایه‌ای برای قضاوت درباره‌ی موثر بودن و ثمربخشی محصول، فرایند یا

1. Ebell

برنامه برحسب هدف‌ها یا ارزش‌های پذیرفته شده می‌باشد (بست^۱، ۱۹۸۳، مترجمان شریفی و طالقانی، ۱۳۸۱). برای تعیین سطوح حیطة شناختی سوالات از تکنیک دلفی دو مرحله‌ای استفاده شد. **جامعه، نمونه و روش نمونه‌گیری:** در این تحقیق جامعه‌ی آماری عبارتند از: کلیهٔ اوراق امتحانی نهایی نمره‌گذاری شده دروس ریاضیات و علوم پایه سوم راهنمایی شهرستان خرم‌آباد در خردادماه ۱۳۸۶ و دبیران دروس ریاضیات و علوم شهرستان خرم‌آباد، دارای سابقه تدریس بالای ۵ سال. از طریق روش نمونه‌گیری تصادفی ساده از هر کدام از جامعه‌های آماری برگه‌های نمره‌گذاری شده ۱۰۰۰ برگه انتخاب شد. از هر کدام از جامعه‌های آماری دبیران ریاضیات و علوم با سابقه تدریس بالای ۵ سال ۲۵ نفر به صورت تصادفی ساده انتخاب شدند.

روش اجرا: روش اجرای تحقیق بدین صورت بود که ابتدا برگه‌های انتخاب شده به ترتیب نمره از کوچکترین به بزرگترین نمره رتبه بندی شدند و سپس برای انتخاب دو گروه ملاک بالا و پایین ۲۷ درصد برگه‌های بالا و ۲۷ درصد از برگه‌های پایین انتخاب شدند این بود که ۲۷ درصد بهترین توافق را بین دو هدف مطلوب ولی متناقض به وجود می‌آورد که این دو هدف عبارتند از: (۱) بزرگ کردن هرچه بیشتر گروه‌های انتهایی (۲) متفاوت کردن گروه‌های انتهایی از یکدیگر تا حد ممکن (ایبل، ۱۹۹۱، ترجمه سپاسی، ۱۳۸۱). ضرایب دشواری و تمیز از طریق فرمول‌های مربوطه محاسبه شدند. برای محاسبه ضریب پایایی نمرات از فرمول آلفای کرانباخ و برای تعیین معنی دار بودن ضریب پایایی بدست آمده از Z فیشر استفاده شد. روایی محتوایی آزمون‌ها در هر درس از طریق ۲۵ دبیر دروس مربوطه با سابقه تدریس بالای ۵ سال بررسی شد. برای تعیین میزان سوالات مربوط به هر یک از سطوح حیطة شناختی از روش توافقی دلفی دو مرحله‌ای استفاده شد. در مرحله‌ای اول سوالات ریاضیات و علوم در بین دبیران انتخاب شده توزیع شد، و از آنها خواسته شد که سطح حیطة شناختی را که هر کدام از سوالات مورد اندازه‌گیری قرار می‌دهند، مشخص کنند، پس از گردآوری داده‌ها محقق

به محاسبه سطح حیطة شناختی برای هر کدام از سوالات پرداخت. به دلیل اینکه در برخی موارد بین دبیران در تعیین سطح حیطة شناختی توافق وجود نداشت. سیاهه‌ای با استفاده از نظرات مرحله-ای اول تهیه شد که در آن سطح حیطة شناختی سوالات ذکر شده بود و این سیاهه بین دبیرانی که نظراتی مغایر با نظرات اکثر دبیران داشتند توزیع شد و از آنها خواسته شد که چنانچه نظراتی مخالف با سیاهه‌های مذکور دارند با ذکر دلیل آن را ذکر نمایند. پس از گردآوری داده‌ها در مرحله‌ای دوم مشخص شد که آنچه که در سیاهه ذکر شده بود مورد توافق دبیران بوده است. سپس محقق تعداد سوالات مربوط به هر کدام از سطوح حیطة شناختی را مشخص کرد و براساس آن مشخص کرد که چند درصد از سوالات به اندازه‌گیری هر سطح پرداخته بودند.

نتایج

در این بخش یافته‌های مربوط به ضریب دشواری و تمیزسوالات، ضریب پایایی، روایی محتوایی و سپس یافته‌های مربوط به سطوح اندازه‌گیری سوالات به ترتیب ذکر شده‌اند. یافته‌های مربوط به ضریب دشواری و ضریب تمیز سوالات. جدول ۱ حداقل (L)، حداکثر (H) ضرایب دشواری و تمیز سوالات امتحان دروس ریاضیات، و علوم را نشان می‌دهند. جدول ۲ ضرایب دشواری و تمیز سوالات امتحان درس ریاضیات و جدول ۳ ضرایب دشواری و تمیز سوالات امتحان درس علوم را نشان می‌دهند. در این جداول PU، ضریب دشواری سؤال برای گروه بالا و PL، ضریب دشواری سؤال برای گروه پایین و P ضریب دشواری سؤال برای هر دو گروه می‌باشد، D، ضریب تمیز سؤال، و معذور کای اسکوتر سطح معنی‌داری بودن ضریب تمیز را نشان می‌دهند. همانطور که جدول ۲ نشان می‌دهد از ۱۶ سوال امتحان درس ریاضیات از لحاظ ضریب دشواری ۸ سوال در چهارچوب ملاک تعیین شده ($p < 0.030$) قرار نداشتند. و از لحاظ ضریب تمیز تمام سوالات به استثنای سوال (۱۶) در چهارچوب ملاک قابل قبول پیشنهاد ($D \leq 0.20$) قرار داشتند. از ۲۴ سوال امتحان درس علوم اکثر سوالات به استثنای ۵ سوال (۲، ۳، ۵، ۶، ۹) از ضریب دشواری مناسب برخوردار نبودند. با وجود این، اکثر سوالات از ضریب

بررسی و تحلیل شاخص‌های روان‌سنجی و سطوح حیطة شناختی سوالات امتحانات نهایی دروس ...

تمیز مطلوب برخوردار بودند و توانسته بودند بین توانایی‌های پاسخ دانش‌آموزان قوی و ضعیف تمیز بگذارند.

جدول ۱. حداقل و حداکثر دشواری و تمیز سوالات به تفکیک درس

درس ضریب	ریاضیات		علوم تجربی	
	H	L	H	L
دشواری	۰/۷۱	۰/۲۰	۰/۷۷	۰/۱۲
تمیز	۰/۴۴	۰/۱۸	۰/۵۴	۰/۱۰

جدول ۲. ضریب دشواری و تمیز سؤالات امتحان درس ریاضیات

شماره سوال	PU	PL	P	D	X ²	α
۱	۰/۸۸	۰/۵۴	۰/۷۱	۰/۳۴	۵۸/۷۵	۰/۰۰۰
۲	۰/۴۴	۰/۱۶	۰/۳۰	۰/۲۸	۵۰/۴۰	۰/۰۰۰
۳	۰/۴۷	۰/۲۱	۰/۳۴	۰/۲۶	۴۰/۶۷	۰/۰۰۰
۴	۰/۴۴	۰/۱۴	۰/۲۹	۰/۳۰	۵۹	۰/۰۰۰
۵	۰/۴۸	۰/۱۲	۰/۳۰	۰/۳۶	۸۳/۳۱	۰/۰۰۰
۶	۰/۴۲	۰/۱۲	۰/۲۷	۰/۳۰	۶۱/۶۴	۰/۰۰۰
۷	۰/۴۶	۰/۱۴	۰/۳۰	۰/۳۲	۶۵/۸۳	۰/۰۰۰
۸	۰/۴۶	۰/۰۸	۰/۲۷	۰/۳۸	۹۸/۹۰	۰/۰۰۰
۹	۰/۶۰	۰/۲۸	۰/۴۴	۰/۳۲	۵۶/۱۰	۰/۰۰۰
۱۰	۰/۴۶	۰/۱۲	۰/۲۹	۰/۳۴	۷۵/۷۹	۰/۰۰۰
۱۱	۰/۵۵	۰/۱۱	۰/۳۳	۰/۴۴	۱۱۸/۲۱	۰/۰۰۰
۱۲	۰/۳۶	۰/۰۴	۰/۲۰	۰/۳۲	۸۶/۴۰	۰/۰۰۰
۱۳	۰/۴۶	۰/۱۴	۰/۳۰	۰/۳۲	۶۵/۸۳	۰/۰۰۰
۱۴	۰/۴۴	۰/۱۲	۰/۲۸	۰/۳۲	۶۸/۵۷	۰/۰۰۰
۱۵	۰/۳۸	۰/۰۴	۰/۲۱	۰/۳۴	۹۴/۰۷	۰/۰۰۰
۱۶	۰/۵۰	۰/۳۲	۰/۴۱	۰/۱۸	۱۸/۰۸	۰/۰۰۰

جدول ۳. ضریب دشواری و تمیز سوالات امتحان درس علوم

شماره سوال	PU	PL	P	D	X^2	α
۱	۰/۹۱	۰/۶۳	۰/۷۷	۰/۲۸	۵۹/۷۶	۰/۰۰۰
۲	۰/۸۴	۰/۵۰	۰/۶۷	۰/۳۴	۷۰/۵۸	۰/۰۰۰
۳	۰/۸۱	۰/۴۱	۰/۶۱	۰/۴۰	۹۰/۷۹	۰/۰۰۰
۴	۰/۴۶	۰/۱۲	۰/۲۸	۰/۳۴	۷۷/۴۱	۰/۰۰۰
۵	۰/۵۸	۰/۲۶	۰/۴۲	۰/۳۲	۵۶/۷۵	۰/۰۰۰
۶	۰/۴۳	۰/۲۵	۰/۳۴	۰/۱۸	۱۹/۴۹	۰/۰۰۰
۷	۰/۴۲	۰/۱۴	۰/۲۸	۰/۲۸	۵۲/۵۰	۰/۰۰۰
۸	۰/۴۰	۰/۰۲	۰/۲۱	۰/۴۲	۱۴۳/۵۴	۰/۰۰۰
۹	۰/۹۴	۰/۴۰	۰/۶۷	۰/۵۴	۱۷۸/۰۴	۰/۰۰۰
۱۰	۰/۴۶	۰/۱۰	۰/۲۸	۰/۳۶	۸۶/۷۹	۰/۰۰۰
۱۱	۰/۴۴	۰/۱۲	۰/۲۸	۰/۳۲	۶۸/۵۷	۰/۰۰۰
۱۲	۰/۳۴	۰/۲۴	۰/۲۹	۰/۱۰	۶/۳۱	۰/۰۰۰
۱۳	۰/۲۲	۰/۰۲	۰/۱۲	۰/۲۰	۵۱/۱۴	۰/۰۰۰
۱۴	۰/۳۴	۰/۰۰	۰/۱۷	۰/۳۴	۱۱۰/۶۰	۰/۰۰۰
۱۵	۰/۳۴	۰/۱۲	۰/۲۳	۰/۲۲	۳۶/۸۹	۰/۰۰۰
۱۶	۰/۴۶	۰/۰۸	۰/۲۷	۰/۳۸	۹۸/۹۰	۰/۰۰۰
۱۷	۰/۳۰	۰/۰۰	۰/۱۵	۰/۳۰	۹۵/۲۹	۰/۰۰۰
۱۸	۰/۴۰	۰/۰۶	۰/۲۳	۰/۳۴	۸۸/۱۲	۰/۰۰۰
۱۹	۰/۴۴	۰/۱۴	۰/۲۹	۰/۳۰	۵۹/۰۰	۰/۰۰۰
۲۰	۰/۴۴	۰/۱۲	۰/۲۸	۰/۳۲	۶۸/۵۷	۰/۰۰۰
۲۱	۰/۴۸	۰/۰۲	۰/۲۵	۰/۴۶	۱۵۲/۳۵	۰/۰۰۰
۲۲	۰/۴۸	۰/۰۸	۰/۲۸	۰/۴۰	۱۹۷/۱۴	۰/۰۰۰
۲۳	۰/۳۸	۰/۰۴	۰/۲۱	۰/۳۴	۹۴/۰۷	۰/۰۰۰
۲۴	۰/۳۶	۰/۰۰	۰/۱۸	۰/۳۶	۱۱۸/۵۴	۰/۰۰۰

یافته‌های ضریب پایایی نمرات. برای برآورد کردن ضریب پایایی نمرات از فرمول آلفای کرانباخ استفاده شد. در آزمون‌های پیشرفت تحصیلی ضریب پایایی بالای ۸۰ صدمه عموماً به عنوان ملاک مطلوب پایایی در نظر گرفته می‌شود (گادو، ۲۰۰۴)، همانطور که جدول ۴ نشان می‌دهد، ضریب پایایی هر کدام از نمرات دروس ریاضیات و علوم برابر با ۰/۹۰ بود. برای آزمون این که آیا ضریب پایایی جامعه نمرات بیشتر از ۰/۸۰ در سطح ۰/۰۱ است، از تبدیل Z فیشر استفاده شد. بعلاوه، فاصله اطمینان ۹۵ درصدی برای ضریب پایایی براساس تبدیل Z فیشر محاسبه شد. همانطور که جدول ۴ نشان می‌دهد، ضریب پایایی نمرات به طور معنی داری در سطح ۰/۰۱ از ارزش بحرانی ۸۰ صدمه بالاتر بود.

جدول ۴. ضرایب پایایی نمرات امتحانات دروس ریاضیات و علوم

شاخص	ضریب پایایی	فاصله اطمینان	Z
ریاضیات	۰/۹۰۴	۰/۸۹ - ۰/۹۱	۸/۸۲
علوم	۰/۹۰۳	۰/۸۹ - ۰/۹۱	۸/۸۲

یافته‌های روایی محتوایی سؤالات امتحانات دروس ریاضیات و علوم. با توجه به اینکه سؤالات امتحانات دروس ریاضیات و علوم براساس جدول مشخصات و با توجه به اهداف و محتوایی این دروس بودجه بندی شده بود، به نظر می‌رسید که این امتحانات از اعتبار محتوایی قابل قبولی برخوردار باشند. علاوه بر این، برای حصول اطمینان از صحت و درستی سؤالات امتحانات نظرات ۲۵ دبیر با سابقه تدریس بالای ۵ سال اخذ گردید. تمام ۲۵ نفر از این که سؤالات امتحانات ریاضیات و علوم روایی محتوایی دارد متفق القول بودند.

یافته‌های سطوح حیطة شناختی اندازه‌گیری شده از طریق سؤالات امتحان دروس ریاضیات و علوم: همانطور که جدول ۵ نشان می‌دهد. ۴ درصد از بارم امتحان دروس ریاضیات به اندازه گیری سطح دانش، ۶ درصد به اندازه گیری سطح فهم، ۸۶ درصد از بارم امتحان سطح کاربرد، و ۴ درصد از بارم آزمون سطح تحلیل را اندازه گیری کرده بودند.

جدول ۵. سطوح حیطة شناختی اندازه‌گیری شده توسط سوالات امتحانات

سطوح درس	دانش	فهم	کاربرد	تحلیل	ترکیب	ارزشیابی
ریاضیات	۴ درصد	۶ درصد	۸۶ درصد	۴ درصد	۰	۰
علوم	۵۳ درصد	۳۲ درصد	۱۵ درصد	۰	۰	۰

با توجه به جدول ۵ درصد از بارم امتحان درس علوم به اندازه‌گیری سطح دانش، ۳۲ درصد از بارم امتحان به اندازه‌گیری سطح فهم، و ۱۵ درصد از بارم امتحان سطح کاربرد را اندازه‌گیری کرده بود.

بحث و نتیجه‌گیری

معلمان در تمام سطوح تحصیلی همیشه خود را در حال ساخت سؤالاتی برای امتحانی که پیش رو دارند می‌بینند و دانش‌آموزان نیز دائماً خود را برای شرکت در امتحانات آماده می‌کنند. یکی از نگرانی‌های موسسه‌های آموزشی از جمله سازمان‌های آموزش و پرورش این است که آیا سؤالات امتحانی ساخته شده توسط معلمان با آئین نامه‌های مربوط که اهداف برگزاری چنین آزمون‌های را تعیین کرده است توافق دارد یا خیر. به بیان دیگر، آیا معلمان در تهیه سؤالات امتحانی با ویژگی‌های آزمون‌سازی آشنایی دارند؟ بنابراین، همانطور ملاحظه شد ویژگی‌های روان‌سنجی سؤالات امتحانات نهایی دانش‌آموزان پایه سوم راهنمایی در دو درس ریاضیات و علوم بررسی و تحلیل شد. در این بخش ابتدا به بحث درباره سطح دشواری و قدرت تمیز سؤالات، پایایی و اعتبار آزمون‌ها خواهیم پرداخت. سپس، سطوح حیطة شناختی که سؤالات این امتحانات اندازه‌گیری کرده بودند را بررسی خواهیم کرد.

بررسی دشواری و قدرت تمیز: متخصصان آزمون‌سازی (مهرنر و لهما، ۱۹۸۴،

ثوراندیک و هیگن^۱، ۱۹۸۶) ضرایب دشواری بین ۰/۷۰ تا ۰/۳۰ را برای آزمون‌های معلم ساخته پیشنهاد می‌کنند. آنها به تجربه دریافتند سوالاتی که دشواری آنها بالاتر از ۰/۷۰ بسیار آسان و سوالاتی که ضریب دشواری آنها پایین تر از ۰/۳۰ است برای دانش آموزان بسیار دشوارند. بررسی ضریب دشواری سوالات درس ریاضیات نشان می‌دهد سوالاتی که ۱۰/۵۰ نمره از بارم امتحان را به خود اختصاص داده بودند دارای ضریب دشواری مناسب نبودند. سهم این سوالات نسبت به کل بارم آزمون ۵۳ درصد بوده است. از این بارم ۱/۵ نمره مربوط به سوالاتی بود که ضریب دشواری آنها بالای ۰/۷۰ بود، که برای آزمون‌شوندگان بسیار آسان بوده است. ۹ نمره دیگر مربوط به سوالاتی بوده است که برای آزمون‌شوندگان بسیار مشکل بوده‌اند. سایر سوالات که ۸/۵۰ نمره از بارم امتحان را به خود اختصاص داده بودند، دارای ضریب دشواری مناسب بودند. با توجه به اینکه بیش از نیمی از بارم امتحان ریاضیات از لحاظ ضریب دشواری نامناسب هستند، سوالات امتحان درس ریاضیات از لحاظ ضریب دشواری مناسب نبوده‌اند. بررسی ضریب دشواری سوالات امتحان درس علوم نشان می‌دهد سوالاتی که ۳ نمره از بارم امتحان به آنها اختصاص داشت و سهم آنها نسبت به کل بارم آزمون ۲۰ درصد بوده است. از لحاظ دشواری مطلوب بوده‌اند ولی سایر سوالات از ضریب دشواری مناسب نداشتند. در نتیجه گیری کلی ضریب دشواری سوالات امتحان درس علوم مناسب نبوده‌اند. با توجه به اینکه اکثر سوالات امتحانات دروس ریاضیات و علوم دارای ضریب دشواری کمتر از ۰/۳۰ بودند، طراحان نتوانسته‌اند به ساخت سوالاتی پردازند که با توانایی دانش آموزان منطبق باشد. متخصصان آزمون سازی معتقدند هر چه قدرت تمیز سوالات آزمون بیشتر باشد، قدرت افتراق آنها بین دانش آموزان قوی و ضعیف بیشتر است. این متخصصان ضریب دشواری ۰/۲۰ را برای سوالات امتحانات معلم ساخته پیشنهاد می‌کنند. در مقایسه با این ملاک، تحلیل و بررسی ضرایب تمیز سوالات امتحانات ریاضیات و علوم نشان داد که اکثر سوالات با این ملاک هم خوانی دارند. فقط تعداد محدودی از سوالات ضرایب تمیزی داشتند که

1 . Thorandike & Hagen

با این ملاک تفاوت داشت. و تمامی ضرایب تمیز در سطح ۰/۰۰۱ معنی دار بودند. به عبارتی دیگر، براساس تحلیل ضرایب تمیز سوالات، نتایج تحقیق نشان داد که طراحان در ساخت سوالاتی که بتوانند بین دانش‌آموزان قوی و ضعیف تفاوت بگذارند، آگاهی کافی داشتند.

بررسی پایایی و روایی محتوایی سوالات. در این تحقیق ضرایب پایایی نمرات آزمون‌های ریاضیات و علوم هر کدام ۰/۹۰ بود. این ضرایب در سطح ۰/۰۱ از ملاک تعیین شده برای پایایی نمرات آزمون‌های معلم ساخته (ضریب پایایی بالای ۰/۸۰) بالاتر بود. بنابراین، نمرات آزمون‌ها برای اندازه‌گیری پیشرفت دانش‌آموزان از ثبات و دقت کافی برخوردار بودند. بررسی اعتبار محتوایی سوالات نشان داد که سوالات براساس جدول مشخصات و با رعایت اهداف رفتاری، محتوایی دروس، سطوح حیطه شناختی و با در نظر گرفتن نسبی اهداف هر یک از آزمون‌های دروس ریاضیات و علوم ساخته شده بودند. بنابراین، معلمان توانسته بودند با تهیه جدول مشخصات آزمون‌ها هماهنگی لازم بین مطالب را فراهم آورند و بر اعتبار محتوایی امتحانات خود بیفزایند.

بررسی سطوح حیطه شناختی سوالات امتحانات. طراحان سوالات درس ریاضیات بیشتر به سطح کاربرد توجه کرده بودند. و برای اندازه‌گیری دو سطح بالای حیطه شناختی (ترکیب، و ارزشیابی) سوالی را طراحی نکرده بودند. برای سوالات امتحان درس علوم طراحان بیشتر به اندازه‌گیری سطح دانش توجه کرده بودند و برای اندازه‌گیری سه سطح بالای حیطه شناختی (تحلیل، ترکیب و ارزشیابی) سوالی را طراحی نکرده بودند. شاید به نظر سازندگان سوالات امتحانی چنین رسیده است که افزودن سوالاتی برای اندازه‌گیری سطوح بالاتر ذهنی (تحلیل، ترکیب و ارزشیابی) بر دشواری امتحان می‌افزاید و پیدا کردن پاسخ چنین سوالاتی برای دانش‌آموزان امکان‌پذیر نباشد. نتایج این تحقیق می‌تواند اطلاعات با ارزشی در ارتباط با ساخت سوالات آزمون‌های امتحانات نهایی در اختیار مسئولین آموزش و پرورش قرار دهد. علاوه بر این، پیشنهاد می‌شود بررسی ساخت سوالات آزمون‌های سایر دروس در دستور کار سازمان آموزش و

پرورش قرار گیرد.

References

- ایبل، رابرت ال. (۱۹۹۱). اصول اندازه‌گیری در علوم تربیتی. ترجمه حسین سپاسی (۱۳۸۱). اهواز: انتشارات دانشگاه شهید چمران اهواز.
- باقری، غلامعلی (۱۳۸۵). بررسی و ارزیابی چگونگی استفاده از سطوح حیطة شناختی و قواعد و اصول طراحی سوالات امتحانی کلاس سوم راهنمایی شهرستان زنجان. پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبایی.
- بست، جان (۱۹۸۳). روش‌های تحقیق در علوم تربیتی و رفتاری. مترجمان: حسن پاشا شریفی و نرگس طالقانی (۱۳۸۱). تهران: رشد.
- سپاسی، حسین (۱۳۸۳). بررسی و تحلیل سطوح حیطة شناختی و شاخص‌های روان‌سنجی سوالات امتحان نهایی دروس نهایی عربی، حسابان و زیست‌شناسی دانش‌آموزان دختر پایه سوم متوسطه در مناطق برخوردار، نیمه برخوردار و غیربرخوردار استان خوزستان. شورای تحقیقات سازمان آموزش و پرورش استان خوزستان.
- سیف، علی اکبر (۱۳۷۲). روش‌های اندازه‌گیری و ارزشیابی آموزشی. تهران: نشر آگاه.
- سعیدی آرانی، احمدرضا (۱۳۷۷). تجزیه و تحلیل محتوای کتب جغرافیای دوره راهنمایی تحصیلی (سال اول، دوم و سوم) و انطباق آن با حیطة شناختی بلوم. پایان‌نامه کارشناسی ارشد دانشگاه علامه طباطبایی.
- شریفی، حسن پاشا (۱۳۷۴). آزمون‌های باز پاسخ و روش‌های عینی نمره‌گذاری آنها. سازمان سنجش آموزش کشور، ۱۰۷-۹۷.
- شرقی، حمید (۱۳۸۳). بررسی و تحلیل سطوح حیطة شناختی و شاخص‌های روان‌سنجی سوالات امتحان نهایی دروس نهایی عربی، حسابان و زیست‌شناسی دانش‌آموزان پسر پایه ای سوم متوسطه در مناطق برخوردار، نیمه برخوردار و غیربرخوردار استان خوزستان. انتشارات شورای تحقیقات سازمان آموزش و پرورش استان خوزستان.

- Baumgartner, T.A.J. & Jackson, A.S. (1991). *Measurement for evaluation in physical education and exercise science*. Wis: Brown Publisher.
- Gado, I. (2004). *Psychometric properties of teacher- made science tests used innational examination for middle grade students in benin*. A dissertation submitted to the Kent State University for the degree of Doctor of Philosophy in Evaluation and Measurement.
- Gage, L. R. (1992). *Educational evaluation and measurement*. Maxwell, Macmillan, International.
- Jandaghi, G., & Shaterian, F. (2008). Validity, Reliability and difficulty indices for instructor- built questions. *Evaluation of Academic Activities in Universities Journal*. 3 (2), 151-155. (Persian)
- McMorris, R., Boothroyd, Roger. (1993). Tests that teachers build: An Analysis of classroom tests in science and mathematics. *Applied-Measurement-in-Education*. 6 (4), 321-341 .
- Mehrens, W. A. J & Lehman, I. J. (1984). *Measurement and evaluation in education and psychology* (3rd Ed), New York: Holt, Rinehart and Winston.
- Mertler, C. (1999). Teachers (mis)conception of classroom test validity and reliability. Paper presented at the annual meeting of the Mid- Western Educational Research Association. Chicago, Illinois.
- Oescher, J., & Kirby, P. (1990). Assessing teacher – made tests in secondary math and science classrooms. Paper presented at the annual meeting of the National Council on Measurement in Education. Boston.
- Page, B. E. & Peterson, N .S. (1995). The computer moves into essay grading, *Delta Kappa*, 79, 7-15.
- Tagomori, T., & Bishop, L. (1994). Content analysis of evaluation instruments used for student evaluation of classroom teaching of performance in higher education. Paper presented at the annual meeting of the American Educational Research Association. New Orleans.
- Thorandike, H. T. & Hagen, E. (1986). *Examiners manual- cognitive ability tests*. From 4. Chicago: Riverside Publishing Company.
- Zhang, L., & Burry, S. J. (1995). A multivariate analysis teachers perceived assessment competency as a function of measurement training and years of teaching. Paper presented at the annual meeting of the Med- South Educational Research Association.
- Ebel, Robert L. (1991). *Measurement principles in educational science*. Translated by Sepasi Hossein (2002). Ahwaz. Published by Ahwaz CHamran Shahid.
- Bagheri, Gholamali (2006). *Survey and evaluation of use of cognitive domain and principles of test questions design of guidance three degree in Zanjan city*. M. A dissertation of Allamehtabbatabaee. (Persian)
- Best, John (1983). *Resarch methods in Educational sciences*. Translated by Sharifi Hasanpasha and Thalaghani Nargess (2002). Tehran: Roshd.
- Seif, Aliakbar, (1993). *Measurement and assessment methods*. Tehran. Agah.

- Saiediarani, A (1998). Survey of gographi book material in guidance level and accomadation with Bloom cognitive domain. M. A dissertation of Allamehtabbatabae.
- Sepasi H. (2002). Survey and analysis cognitive domain and psychometric propertice of final test questionnaires in arabic, accounting and biology in secondry school in Khosestan Province. Published By research concual of Khosestan educational organizational. (Persian)
- Shareefi , Hassanpasha (1995). Essey tests and objective methods of those scored. Sanjesh organization. 97-107. (Persian)
- Sharghi, H. (2002). Survey and analysis cognitive domain and psychometric propertice of final test questionnaires in arabic, accounting and biology in secondry school in Khosestan Province. Published By research concual of Khosestan educational organizational. (Persian)

A study of the psychometric properties and levels of cognitive domain of final exams questions of mathematics and sciences of the third grade junior high school in Khorramabad

A.A. Miraghaie¹, H. Sepasi², B. Mohajeran³ & A.R.Ghalaie⁴

Abstract

The basic aim of this research was the analysis of psychometric properties and the The main aim of this research was to analyze the psychometric properties and the levels of cognitive domain of final exam questions of mathematics and sciences of the third grade junior high school in Khorramabad. From each pool, 1000 graded sheets of math and sciences were selected through simple random sampling. To determine the levels of cognitive domain, a sample size of 25 teachers were randomly selected among mathematics and science teachers using Delphi technique. The results showed that the difficulty coefficient of exam questions was not appropriate but the discrimination index was acceptable except for a few cases. Analysis of the reliability and content validity showed that the questions enjoyed reliability and content validity. The results showed that the questions of math exam were not allocated to assess the high level of cognitive field (combination and evaluation).

Keywords: Difficulty coefficient, discrimination coefficient, reliability, content validity, cognitive domain.

1 . Corresponding Author: M.A educational research, Shahid Chamran University of Ahvaz (a.a.miraghaie@gmail.com).

2 . Professore of educational sciences, Shahid Chamran University of Ahvaz.

3. Assistant Professore of educational sciences, University of Urmia

4. Assistant Professore of educational sciences, University of Urmia