

بررسی آزمون‌های چند گزینه‌ای دانشکده پزشکی دانشگاه علوم پزشکی مازندران در نیم‌سال اول سال تحصیلی ۸۶-۸۵

محمد رضا حق‌شناس^۱، کوروش وحیدشاهی^{۲*}، میترا محمودی^۳، لیلا شهبازنژاد^۴، نیکو پروین‌نژاد^۴، اعظم عمادی^۴

۱. دکترای ویروس‌شناسی، استادیار گروه میکروب‌شناسی و انگل‌شناسی دانشکده پزشکی، دانشگاه علوم پزشکی مازندران، ایران
 ۲. متخصص اطفال، استادیار مرکز توسعه آموزش دانشگاه علوم پزشکی مازندران، ایران
 ۳. دکترای فارماکولوژی، دانشیار گروه فیزیولوژی و فارماکولوژی دانشکده پزشکی، دانشگاه علوم پزشکی مازندران، ایران
 ۴. پزشک عمومی، مرکز مطالعات و توسعه آموزش علوم پزشکی، دانشگاه علوم پزشکی مازندران، ایران
- دریافت مقاله: ۸۷/۵/۱ • پذیرش مقاله: ۸۷/۹/۱۰

زمینه و هدف: آزمون‌های چند گزینه‌ای رایج‌ترین انواع آزمون‌ها در آموزش پزشکی می‌باشند که از درجه پایانی بالایی برخوردارند، اما مسئله روایی محتوا، روایی ساختار، توزیع سطح دانش و رعایت اصول ساختاری آنها همواره مورد توجه محققین قرار داشته است. در این مطالعه کیفیت آزمون‌های چند گزینه‌ای دانشکده پزشکی دانشگاه علوم پزشکی مازندران از نظر توزیع سطح دانش و رعایت اصول ساختاری مورد بررسی قرار گرفت.

روش کار: در این مطالعه توصیفی کلیه سؤالات آزمون‌های کتبی چند گزینه‌ای یک نیم‌سال تحصیلی دانشکده پزشکی دانشگاه علوم پزشکی مازندران مورد بررسی قرار گرفت. سؤالات از نظر سطح دانش (در سه سطح یاد آوری، فهم، کاربرد) و ساختار به کمک چک لیستی که مبتنی بر اصول میلمن برای طراحی ساقه و گزینه‌های سؤال بود، ارزیابی گردید.

یافته‌ها: از مجموع ۱۴۷۸ سؤال مربوط به ۲۵ آزمون، سؤالات مربوط به دروس علوم پایه، فیزیوپاتولوژی و کارآموزی به ترتیب ۲۸/۷، ۱۸/۷ و ۵۲/۶ درصد بود. ۷۷/۴ درصد سؤالات در تاکسونومی I و بقیه در تاکسونومی II و III طراحی شده بود. از نظر اشکالات ساختاری ۴۶ درصد سؤالات بدون اشکال و بقیه یک یا چند اشکال ساختاری داشتند. تاکسونومی و اشکالات ساختاری آزمون‌های دروس کارآموزی و فیزیوپاتولوژی به‌طور معنی‌داری از دروس علوم پایه کمتر بود (به ترتیب $P=0/023$ و $P=0/001$). سؤالات با تاکسونومی بالاتر اشکالات ساختاری کمتری داشتند.

نتیجه‌گیری: مطالعه حاضر نشان داد، کیفیت آزمون‌های چند گزینه‌ای مورد مطالعه از نظر توزیع تاکسونومی و رعایت اصول ساختار MCQ بین آزمون‌های مقاطع مختلف متفاوت بوده و در برخی موارد نیازمند اتخاذ تدابیر اصلاحی می‌باشد.

کلید واژه‌ها: پیشرفت تحصیلی، آزمون چند گزینه‌ای، تاکسونومی، اصول میلمن

* نویسنده مسؤل: مرکز مطالعات و توسعه آموزش پزشکی دانشگاه علوم پزشکی مازندران، سه راه جویبار، ساری، ایران

مقدمه

ارزیابی یا ارزشیابی عبارت از فرآیند نظام‌مند گردآوری، تحلیل و تفسیر داده‌ها به منظور ارزش‌گذاری و قضاوت می‌باشد (۱). اگر این فرآیند در مورد افراد، گروه‌ها و سازمان‌ها انجام گردد، معمولاً به عنوان ارزیابی و اگر در مورد برنامه انجام شود، به آن ارزشیابی اطلاق می‌گردد (۱،۲).

آزمون‌ها از انواع رایج ارزیابی هستند که در سیستم آموزشی صورت می‌گیرند. آزمون‌ها از نظر محتوای آزمون (دانش، مهارت و ...)، هدف آزمون (پیشرفت تحصیلی، گزینش داوطلبان و ...)، معیار قبولی و جنبه‌های گوناگون دیگر قابل طبقه بندی می‌باشند. برای آن‌که هر آزمون نتایج قابل داشته باشد، باید از روایی، پایایی و قابلیت اجرا برخوردار باشد که هر یک جنبه‌های متفاوتی از آزمون را در بر می‌گیرند (۱،۲).

در میان آزمون‌هایی سنجش دانش، سؤالات چند گزینه‌ای (Multiple choice Question: MCQ) که اولین بار در سال ۱۹۱۴ در دانشگاه کانزاس توسط فردریک (Fredrick) ابداع گردید، اکنون رایج‌ترین نوع آزمون‌ها در آموزش پزشکی می‌باشند. آزمون‌های چند گزینه‌ای به علت عینیت کامل، از پایایی بالا (حداکثر پایایی ممکن) برخوردار می‌باشند، اما روایی آنها به‌ویژه در مورد سطح دانش (تاکسونومی) و رعایت اصول طراحی ساختار MCQ از مسایل مورد توجه در مطالعات و پژوهش‌های آموزشی بوده است. بدیهی است اگر یک آزمون از طراحی مطلوب در زمینه انتخاب تاکسونومی سؤالات و رعایت قواعد ساختاری برخوردار نباشد واجد روایی محتوا و روایی ساختار مطلوب نبوده و در این صورت نه تنها نقش اصلی آزمون به‌عنوان جزء تکمیل‌کننده و حیاتی چرخه آموزش از دست می‌رود، بلکه این امر برانگیزه‌های فراگیران اثرات منفی بر جای گذاشته و خطر به هدر رفتن تلاش اساتید و سیستم آموزشی را خواهد داشت، لذا بررسی کیفیت سؤالات و میزان استاندارد بودن آزمون‌ها ضروری است. مطالعات موجود در دانشگاه‌های مختلف در این زمینه

نتایج متفاوتی را نشان داده است. در مطالعه کاوه و همکاران (۲۰۰۲)، مختل بودن ملاک‌های آنالیز پرسش‌ها در بخش قابل توجهی از سؤالات آزمون‌های چند گزینه‌ای نشان داده شد (۳) ولی شکورنیا و همکاران (۲۰۰۷) بیش از نیمی از سؤالات را فاقد اشکالات ساختاری گزارش کردند (۴).

مطالعات جامعی که سؤالات چند گزینه‌ای را هم از نظر تاکسونومی و هم از نظر رعایت اصول ساختاری سؤال بررسی کرده باشد، کمتر انجام شده است. لذا در این مطالعه ساختار آزمون‌های چند گزینه‌ای دانشکده پزشکی دانشگاه علوم پزشکی مازندران مورد بررسی قرار گرفت. از آنجا که این مطالعه یک طرح تحقیقاتی به منظور انتشار مقاله و مستند سازی شواهد حاصل بوده است، لذا داده‌ها به شیوه‌ای دقیق‌تر نسبت به رویه رایج کمیته آزمون‌های مرکز توسعه آموزش که ارزشیابی آزمون‌ها را به طور متداول انجام می‌دهند، گردآوری گردید. برای بررسی هر آزمون از تعداد بیشتری از صاحب‌نظران استفاده شد. تا با شناخت دقیق‌تر ساختار آزمون‌ها و اشکالات موجود، راهکارهای لازم جهت ارتقای کیفیت برنامه‌های موجود در زمینه استاندارد سازی آزمون‌ها که در مرکز توسعه آموزش در دست اجرا می‌باشد شناسایی شده و اجرا گردد.

روش کار

در این مطالعه توصیفی کلیه سؤالات آزمون‌های چند گزینه‌ای پایان نیم‌سال اول سال تحصیلی ۸۶-۸۵ دانشکده پزشکی دانشگاه علوم پزشکی مازندران شامل ۱۴۷۸ سؤال از ۲۵ آزمون مربوط به دروس: آناتومی تنه، آناتومی اندام، بیوفیزیک، بهداشت یک، قارچ‌شناسی، بهداشت سه، فیزیولوژی، آناتومی اندام، پاتولوژی عمومی، پاتولوژی اختصاصی، فیزیوپاتولوژی ریه، روماتولوژی، گوارش، کلیه، فارماکولوژی، بیماری‌های اعصاب، زنان، اورولوژی، عفونی، جراحی عمومی، جراحی اختصاصی، اپیدمیولوژی، روان‌پزشکی، اورتوپدی و اطفال مورد بررسی قرار گرفت.

کلیه سؤالات آزمون‌ها از نظر سطح دانش (تاکسونومی) و همچنین رعایت اصول ساختاری طراحی سؤالات چند گزینه‌ای بررسی گردید. سطح دانش سؤالات بر اساس سطوح شناختی بلوم در سه سطح شامل تاکسونومی I (سؤال مبتنی بر حافظه و یادآوری)، تاکسونومی II (سؤال مبتنی بر فهم) و تاکسونومی III (سؤال مبتنی بر کاربرد مفاهیم)، قرار گرفت. ساختار طراحی سؤالات بر اساس اصول استاندارد طراحی ساقه و گزینه‌های میلمن (Millman) با چک لیستی شامل ۱۲ مورد بررسی گردید و روایی آن با توجه به استاندارد بودن پرسش‌نامه (۵) و پایایی آن در ۱۰ درصد از واحدهای مورد مطالعه در سطح بین-ناظر (Interrater) یعنی میزان همخوانی نتایج ۲ ارزیابی کننده از آزمون‌ها ۰/۸ و در سطح خود-ناظر (intrarater) از طریق آزمون مجدد به فاصله یک هفته در مورد هر فرد نمره دهنده ۰/۹ و در مجموع با ضریب همخوانی $r = 0/85$ تأیید گردید.

بررسی سؤالات توسط اعضای کمیته آزمون‌های مرکز توسعه آموزش که در زمینه آزمون‌های چند گزینه‌ای کاملاً مسلط بودند، انجام گردید. مسئولیت بررسی هر آزمون برعهده سه نفر صاحب‌نظر از اعضای کمیته آزمون‌ها بود که از میان اعضای منتخب گروه‌های آموزشی دانشکده پزشکی (با تخصص‌های زنان و زایمان، داخلی، جراحی، بیهوشی، رادیولوژی، فیزیولوژی، اطفال، آناتومی، عفونی) برگزیده شده و با گذراندن کارگاه‌ها و دوره‌های فشرده ارزیابی فراگیران (با تأکید بر سؤالات چند گزینه‌ای) در هر دو زمینه محتوای علمی و دانش آموزش پزشکی صاحب‌نظر شده بودند. بررسی سؤالات توسط افراد به صورت مستقل انجام گردید و نظر غالب در مورد تاکسونومی و آیت‌های چک لیست میلمن به عنوان نتیجه نهایی تلقی گردید. تجزیه و تحلیل داده‌ها، به وسیله نرم افزار آماری SPSS.V.13 و با استفاده از شاخص‌های مرکزی و پراکندگی و آزمون مربع کای صورت گرفت و $P < 0/05$ به عنوان سطح معنی‌داری در نظر گرفته شد.

یافته‌ها

در مجموع ۱۴۷۸ سؤال مربوط به ۲۵ آزمون کتبی مورد بررسی قرار گرفت. ۲۸/۷ درصد سؤالات مربوط به مقطع علوم پایه، ۱۸/۷ درصد مربوط به دوره فیزیوپاتولوژی و ۵۲/۶ درصد مربوط به دروس تئوری مقطع بالینی بود. از نظر توزیع سطح دانش ۷۷/۴ درصد سؤالات در تاکسونومی I، ۱۶/۸ درصد در تاکسونومی II و بقیه در تاکسونومی III طراحی شده بودند. درصد سؤالات در تاکسونومی I در مقطع علوم پایه به‌طور معنی‌داری بیشتر از سایر مقاطع بود (جدول ۱). مقایسه تاکسونومی سؤالات در هر مقطع نشان داد که طیف درصد سؤالات تاکسونومی I در آزمون‌های دروس مقطع علوم پایه از ۸۶ تا ۱۰۰ درصد متغیر بود و تفاوت دروس در این زمینه معنی‌دار بود ($P = 0/048$). در دوره فیزیوپاتولوژی این محدوده از ۶۳ تا ۱۰۰ درصد ($P = 0/021$) و در مقطع بالینی از ۳۵ تا ۹۲ درصد بود که اختلاف دروس در این زمینه معنی‌دار بود ($P = 0/013$).

جدول ۱: توزیع فراوانی تاکسونومی سؤالات چند گزینه‌ای آزمون‌های دانشکده پزشکی دانشگاه علوم پزشکی مازندران در یک نیم‌سال

تاکسونومی مقطع	I	II	III	P
علوم پایه	۴۵۱(۹۴)	۲۴(۶)	۰	۰/۰۲۳
فیزیوپاتولوژی	۲۱۲(۷۶)	۴۱(۱۵)	۲۲(۹)	
بالینی	۵۳۶(۷۰)	۱۸۳(۲۳)	۵۸(۷)	
کل	۱۱۵۰(۷۷۱۴)	۲۴۸(۱۶۲۸)	۸۰(۵۱۸)	

در مورد رعایت اصول طراحی سؤالات (اصول میلمن)، نتایج نشان داد که در کل آزمون‌ها ۴۶ درصد سؤالات بدون اشکال بودند و ۳۲ درصد یک اشکال، ۱۵ درصد دو اشکال و ۶/۸ درصد بیش از دو اشکال ساختاری داشتند. میزان اشکالات ساختاری سؤالات آزمون‌ها در دروس بالینی به طور معنی‌داری کمتر از آزمون‌های فیزیوپاتولوژی و علوم پایه بود (جدول ۲).

($P=0/001$). همچنین در دوره فیزیوپاتولوژی طیف اشکالات آزمون‌ها از $31/8$ تا 70 و در دوره بالینی از 16 تا $59/6$ درصد و در هر دو مورد تفاوت معنی‌داری وجود داشت ($P=0/0001$).

فراوانی اشکالات در هر آیتیم از چک لیست در کل آزمون‌ها و به تفکیک مقطع در جدول ۳ نشان داده شده است. در هر سطر درصد سؤالات آزمون که آیتیم مربوطه در آنها رعایت نشده و در آن آیتیم دارای اشکال بوده، آورده شده است. بررسی رابطه تاکسونومی سؤالات با میزان اشکالات ساختاری نشان داد که در سؤالات با تاکسونومی I، میزان سؤالات بدون اشکال ساختاری ۴۴ درصد، در سؤالات تاکسونومی II، ۵۰ درصد و در سؤالات با تاکسونومی III، ۶۳ درصد بوده است ($P=0/0001$).

جدول ۲: توزیع فراوانی مطلق (درصد) اشکالات ساختاری سؤالات چند گزینه‌ای بر حسب مقطع دروس در آزمون‌های دانشکده پزشکی دانشگاه علوم پزشکی مازندران

مقطع	تعداد اشکال	ندارد	یک	دو	بیش از دو	P
علوم پایه	۱۶۷	۱۳۳	۷۶	۴۸	(۱۱/۵)	۰/۰۰۰۱
فیزیوپاتولوژی	۱۱۴	۹۶	۴۵	۱۹	(۷)	
بالینی	۴۰۶	۲۴۷	۹۳	۲۹	(۴/۳)	
کل	۶۸۷	۴۷۶	۲۱۹	۹۶	(۶/۸)	

مقایسه اشکالات ساختاری سؤالات در دروس علوم پایه نشان داد که محدوده اشکالات ساختاری در آزمون‌ها از صفر تا ۹۵ درصد و این اختلاف در دروس مختلف معنی‌دار بود

جدول ۳: درصد اشکال ساختاری در آزمون‌های نیم‌سال اول سال تحصیلی ۸۶-۸۵ دانشکده پزشکی دانشگاه علوم پزشکی مازندران

نتیجه آزمون (P)	بالینی	فیزیوپاتولوژی	علوم پایه	کل	موضوع
۰/۰۱۲	۷	۴	۱۱	۱۰/۴	آیا بخش اعظم اطلاعات در ساقه سؤال گنجانده شده است؟
۰/۰۰۸	۶	۴	۱۶	۸/۲	آیا سؤال یک هدف اختصاصی یادگیری را مورد ارزیابی قرار می‌دهد؟
۰/۳۱	۴	۹	۵	۵/۵	آیا لغات استفاده شده در ساقه یا گزینه‌ها، شفاف و مستقیم بیان شده‌اند؟
۰/۴	۱	۰	۰/۵	۰/۵	آیا از کاربرد گزینه منفی برای ساقه منفی خودداری شده است؟
۰/۲۴	۳	۲	۵	۵/۳	آیا از کاربرد گزینه‌های نظیر همه موارد هیچ کدام و گزینه‌های ترکیبی خودداری شده است؟
۰/۸	۱	۰	۱	۰/۹	آیا از کاربرد گزینه‌های متضاد یکدیگر خودداری شده است؟
۰/۰۲۱	۸	۲۱	۳	۶/۲	آیا از لغات مثبت در ساقه سؤال استفاده شده است یا در صورت منفی بودن ساقه سؤال، لغات منفی مشخص شده‌اند؟
۰/۶۵	۲	۰	۲	۱/۲	آیا هر سؤال مستقل از سؤالات دیگر می‌باشد؟
۰/۹۲	۱۳	۱۸	۷	۱۲/۲	آیا گزینه‌ها از نظر طول، ساختار لغوی و سبک نگارش هم‌سنگ هستند؟
۰/۸	۱	۰	۱	۰/۹	آیا تا حد امکان از کاربرد عبارت‌های تکراری در گزینه‌ها خودداری شده است؟
۰/۰۴۱	۲	۹	۲	۳/۷	آیا کلمات به کار رفته در ساقه و یا گزینه‌ها از نظر املائی صحیح هستند؟
۰/۰۴۵	۱۲	۱۱	۳۲	۱۸/۷	آیا گزینه‌ها به‌طور عمودی لیست شده‌اند؟

بحث و نتیجه‌گیری

در این مطالعه دو جنبه مهم از عواملی که در کیفیت سؤالات چهار گزینه‌ای نقش تعیین کننده‌ای دارند شامل سطح دانش بکار رفته، تاکسونومی برای روایی محتوا و رعایت اصول ساختاری میلمن برای روایی ساختار مورد بررسی قرار گرفت. ارزش و اهمیت این‌گونه مطالعات و بازخورد آنها در زمینه ارتقاء کیفیت آزمون‌های چند گزینه‌ای در مطالعات مختلفی نشان داده شده است: مانند مطالعه اپشتین (Epstein) از دانشگاه روچستر و فریدمن (Friedman) از دانشگاه داندی (۷،۶) و از مطالعات داخلی مطالعه رضانی‌بدر و همکاران در تهران (۸) و غلامی و همکاران در اصفهان (۹).

با توجه به آن‌که آزمون‌های چند گزینه‌ای از نظر پایایی و عینی بودن، که از چالش‌انگیزترین جنبه‌های آزمون می‌باشد، از کفایت بالایی برخوردارند، (۱۰-۱۲)، با تلاش در زمینه ارتقاء جنبه‌های اصلی روایی محتوا و ساختار آزمون‌های چند گزینه‌ای، این نوع از آزمون‌ها به یکی از مطلوب‌ترین شیوه‌های ارزیابی دانش فراگیران تبدیل می‌شوند و همان‌طور که ماویس (Mavis) و همکاران از دانشگاه میشیگان (۱۳) نشان دادند تنها در این صورت استفاده از سؤالات چند گزینه‌ای به‌عنوان یک روش رایج، اعتبار قابل توجهی برای آزمون‌ها ایجاد می‌کند.

از نظر توزیع سطح دانش در سؤالات مورد بررسی، نتایج نشان داد که بخش اعظم سؤالات مورد بررسی (بیش از سه چهارم) در تاکسونومی I (یادآوری) طرح شده بودند که این یافته با مطالعه رسولی‌نژاد و همکاران در کاشان (۱۴) که نشان دادند اکثریت سؤالات آزمون ارتقاء دستیاران آن دانشکده در تاکسونومی I طرح شده بود، هم‌خوانی داشت. در این زمینه مطالعات خارجی، در حالی که شواهد قدیمی‌تر مانند مطالعه هاکسهام (Huxham) و همکاران در آمریکا (۱۵) یافته‌های مشابه، مبنی بر غلبه کاربرد تاکسونومی I در آزمون‌های پیشرفت تحصیلی، را نشان دادند، اما مطالعات جدیدتر مانند مطالعه پلمر (Plemer) نشان داد که در سال‌های اخیر سهم

بیشتری از سؤالات به سطوح عمیق‌تر دانش (تاکسونومی II و III) اختصاص یافته‌اند. به‌هر حال این واقعیت که در آزمون‌های مورد بررسی در این مطالعه، تمرکز عمده طراحان سؤال بر محفوظات دانشجویان و نه فهم عمیق و تحلیل و کاربرد مفاهیم، بوده است، جای تأمل دارد، چرا که بدیهی است چنین رویکردی موجب سوق دادن دانشجویان به محفوظات و برداشت سطحی مطالب می‌گردد که در مورد مطالب پزشکی این شیوه یادگیری نه تنها از ماندگاری لازم در ذهن دانشجویان برخوردار نیست، بلکه از کارکرد لازم در زمینه پایه‌ریزی بستر دانش آینده و گسترش توانایی‌های ذهنی دانشجویان کاملاً دور است. گرچه میزان سؤالات طرح شده در تاکسونومی‌های بالا در دروس بالینی و برخی دروس علوم پایه به حد مطلوب نزدیک بوده است، اما با توجه به آن‌که اساتید طراح سؤال از آموزش‌های لازم توسط مرکز توسعه در زمینه توزیع مطلوب تاکسونومی‌ها، تهیه blueprint در مورد مفاهیم و سطح دانش سؤالات برخوردار شده‌اند، به‌نظر می‌رسد تلاشی افزون‌تر و مستمر در زمینه کاربرد روش‌های آموزش داده شده و توجه ویژه به طرح سؤال در سطوح عمیق‌تر دانش ضروری می‌باشد.

در زمینه رعایت اصول ساختار طراحی سؤالات چند گزینه‌ای (اصول میلمن) مطالعه حاضر نشان داد که نزدیک به نیمی از سؤالات فاقد اشکال و درست طراحی شده بودند که در این زمینه شکورنیا و همکاران در اهواز نشان دادند بیش از نیمی از سؤالات فاقد اشکالات ساختاری است (۴). اما برخی مطالعات دیگر داخلی نتایج متفاوتی را نشان دادند. در این زمینه مک کوبری (McCaubri) و همکاران در دانشگاه بریستول و هاموند (Hammond) و همکاران از دانشگاه هامپتون (South Hampton)، وجود اشکالات در درصد قابل توجهی از سؤالات را نشان دادند (۱۶،۱۷). گرچه در اغلب آزمون‌های بالینی و برخی آزمون‌های فیزیوپاتولوژی، بخش اعظم سؤالات، درست و فاقد اشکالات بودند، اما عدم رعایت اصول ساختاری به هر میزان، روایی سازه را در آزمون

گروه‌های آموزشی دانشکده پزشکی در مورد نگرش و انگیزش اساتید و نیز برگزاری نشست‌های مشترک کمیته آموزشی با اساتید گروه‌ها و به‌ویژه استقرار مکانیسمی برای اعمال نتایج بررسی آزمون‌ها در ارزشیابی اساتید مانند اعطای امتیازات و همچنین محدودیت‌های آموزشی.

البته در این مطالعه محدودیت‌هایی از قبیل بررسی سؤالات یک نیم‌سال تحصیلی و انجام مطالعه در یک دانشکده وجود داشت. همچنین اگر نتایج این مطالعه با یک بررسی پس آزمونی همراه می‌گردید، تصویر جامع‌تری از آزمون‌ها حاصل می‌شد.

البته توجه به این نکته نیز ضروری است که با توجه به محدودیت‌های آزمون‌های چند گزینه‌ای، همان طور که در بسیاری از دانشگاه‌های معتبر جهان از جمله توسط سیمکین (Simkin) و همکاران در انگلستان (۲۲) و بوش (Bush) و همکاران (۲۳) نشان داده شده است، باید حرکت اساتید به استفاده از سایر روش‌های استاندارد سنجش دانش در کنار سؤالات چند گزینه‌ای سوق یابد تا ارزیابی جامع‌تر و دقیق‌تری انجام گردد.

در مجموع این مطالعه نشان داد کیفیت آزمون‌های چند گزینه‌ای از نظر سطح دانش و رعایت اصول ساختار سؤالات چند گزینه‌ای در بین آزمون‌های مختلف از سطح مطلوب تا متوسط متفاوت بوده و نیاز به پی‌گیری و استمرار برنامه‌های موجود در زمینه ارتقاء و کیفیت آزمون‌ها وجود دارد.

سپاسگزاری

محققین لازم می‌دانند از زحمات اعضای کمیته آزمون‌های مرکز توسعه آموزش پزشکی و همچنین جناب آقای دکتر سید حمزه حسینی، دکتر علیرضا رفیعی و همچنین همکاران محترم امور آموزشی دانشکده پزشکی خانم‌ها مهشید صابری و فاطمه مستشرق، قدردانی نمایند.

مختل نموده و در این صورت کارکردهای آموزشی آزمون را تحت تأثیر قرار داده و این مسأله با ایجاد زمینه برای کسب نمره با روش‌هایی غیر از تسلط علمی (حدس زدن، گمراه شدن دانشجویان و...) نگرش دانشجویان را نسبت به آزمون با اشکال مواجه می‌سازد. در این زمینه یراقی و همکاران (۱۸) در اصفهان، مختل بودن ملاک‌های آنالیز آیتم را در بخش قابل توجهی از سؤالات آزمون‌های چند گزینه‌ای و تأثیر نامطلوب آن را بر کیفیت آزمون نشان دادند. مطالعات وکیلی و همکاران (۱۹) در کاشان و کاوه در مشهد (۳) نتایج مشابهی را نشان دادند.

در مورد شایع‌ترین اشکالات در ساختار سؤالات نتایج نشان داد که در کل آزمون‌های مورد بررسی، عدم رعایت "چیدمان عمودی گزینه‌ها"، "هم‌سنگ بودن گزینه‌ها" و "آوردن بخش اعظم اطلاعات در ساقه سؤال"، در بیش از ۱۰ درصد سؤالات وجود داشت. در این زمینه مطالعه جوادی و همکاران در بم نشان داد شایع‌ترین اشکالات در آیتم‌ها، "درج مطالب تکراری در گزینه‌ها" و "وجود لغات منفی در ساقه سؤال" بود (۲۰). مطالعه شکورنیا و همکاران (۴) در اهواز و دری‌پور (۲۱) در تهران موارد متفاوت دیگری را نشان داد. با توجه به مجموعه یافته‌های این مطالعه و شواهد فوق به‌نظر می‌رسد باید علاوه بر بازنگری، تقویت و به‌روزرسانی برنامه‌های موجود، استاندارد سازی و ارتقای کیفیت ارزیابی فراگیران و به‌ویژه آزمون‌های چند گزینه‌ای، جهت ارزیابی عملکرد بهینه اساتید در طراحی و اجرای آزمون چند گزینه‌ای راهکارهای لازم جهت بهینه نمودن دانش و نگرش اساتید در مورد ارزیابی فراگیران اتخاذ گردد که شامل مواردی چون: تکرار کارگاه‌های ارزیابی فراگیران در زمینه سؤالات چند گزینه‌ای برای گروه‌های آموزشی که ارزیابی مطلوبی نداشته‌اند، تهیه پمفلت آموزشی در مورد سؤالات چند گزینه‌ای و فرم‌های blueprint و ارسال آن در زمان طراحی سؤال برای کلیه

References

1. Bazargan A. Educational assessment. Tehran: SAMT; 2001: 9-17. [In Persian]
2. Saif A. Educational measurement, assessment and evaluation. 4th ed. Tehran: Agah; 2007: 70-80. [In Persian]
3. Kaveh M. Analytical assessment of multiple – choice test at medical school. Medical Journal of Mashhad University of Medical Sciences 2002; 76 (45): 89-95. [In Persian]
4. Shakoornia A, Khosravi A, Shariati A, Zarei A. Survey on multiple choice questions of faculty members of Jondi Shapor Medical University of Ahwaz. The 8th National Congress of Medical Education. kerman: kerman University of Medical Sciences; 2007: 44. [In Persian]
5. Seif A. Educational measurement, assessment and evaluation. 4th ed. Tehran: Doran; 2004: 426-8 [In Persian]
6. Epstein RM. Assessment in medical education. New England J Med 2007; 356 (4): 387-96.
7. Friedman Ben DM. Standard setting in student assessment. Med Teach 2000; 22 (2): 120-30.
8. Ramezani Badr F, Shaban M. Survey on effect of evaluation and Analysis of summative examination on test formation of faculties. The 8th National Congress of Medical Education. kerman: kerman University of Medical Sciences; 2007: 124. [In Persian]
9. Gholami V, Dadgostarina M. Survey on effect of evaluation of examination on reliability of MCQs of ICM in Isfahan during 2001-2005. The 8th National Congress of Medical Education. kerman: kerman University of Medical Sciences; 2007: 213. [In Persian]
10. Downing SM. Reliability on the reproduce ability of assessment data. Med Educ 2004; 38(9): 1006 -12.
11. Collins J. Education techniques for lifelong learning: Writing multiple-choice question for continuing medical education activities and self assessment modules. Radio Graphics 2006; 26(2): 543-51.
12. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice question for use in nursing research and education. Collegian 2005; 12(1): 14-24.
13. Mavis DE, Cole BL, Hoppe RB. A survey of student assessment in US medical schools. The balance of breadth and fidelity. Teach learn Med 2001; 13(2): 74-9.
14. Rasolinejad SA, Vakihi Z, Fakhariion E, Mosayebi Z, Moniri R. Comparative survey of taxonomies of residents. Promotion examination, kashan Medical University 2006. The 8th National Congress of Medical Education. kerman: kerman University of Medical Sciences; 2007: 68. [In Persian]
15. Huxham GJ, Naeraa N. Is Bloom’s taxonomy reflected in the response pattern to MCQ items? Med Educ 1980; 14(1): 23-6.
16. McCourbie P. Improving the fairness of multiple choice questions: A literature review. Med Teach 2004; 26(8): 709 –12
17. Hammond EJ, Mcindo AK, Sansome AJ, Spargo PM. Multiple choice examination: Adapting an evidence – based approach to exam technique. Anesthesia 1998; 53(11): 1105-8.
18. Toghyanifar N, Gorgyaraghi M, Esmaili A, Avizhgan M, Mazaheri M. Survey on quality of Multiple Choice Questions of clinical clerkship Isfahan Medical Faculty 2005–2006. The 8th National Congress of Medical Education. kerman: kerman University of Medical Sciences; 2007: 73. [In Persian]
19. Vakili Z, Sayah M, Fakhrian E, Rasouli Nejad SA, Mosayebi Z, Moniri R. Survey on psychometric characteristics of resident examination Kashan Medical University. The 8th National Congress of Medical Education. kerman: kerman University of Medical Sciences; 2007: 174. [In Persian]
20. Javadi M, Abbas zadeh A, Borhani F, Abdoli R. Analytical survey of summative MCQs of Bam Nursing Faculty. The 8th National Congress of Medical Education. kerman:

- kerman University of Medical Sciences; 2007: 217. [In Persian]
21. Daripor F, karimi SH, Pakaeen J. Evaluation of MCQs of nursing student 2006. The 8th National Congress of Medical Education. kerman: kerman University of Medical Sciences; 2007: 67. [In Persian]
 22. Sim kin MG, Kuechler WL. Multiple choice test and student understanding: What is the connection? *Decision Scie J Innovat Educ* 2005; 3(1): 73-98.
 23. Bush ME. Quality assurance of multiple choice tests. *Qual Assur Educ* 2006; 14(4): 398-404.

Evaluation of Multiple Choice Questions in the School of Medicine, Mazandaran University of Medical Sciences, the First Semester of 2007

Mohammadreza Haghshenas^{1*}, *Kourosh Vahidshahi*², *Mitra Mahmudi*³, *Leyla Shahbaznejad*⁴, *Nikoo Parvinnejad*⁴, *Aazam Emadi*⁴

1. Ph.D. in Virology, Assistant Professor of Microbiology and Parasitology Dept., School of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

2. Pediatrician, Assistant Professor of Education Development Center, School of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

3. Ph.D. in Pharmacology, Associate Professor of Physiology and Pharmacology Dept., School of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

4. General Practitioner, Education Development Center, Mazandaran University of Medical Sciences, Sari, Iran

• Received: 22 Jul, 2008 • Accepted: 30 Nov, 2008

Background & Objective: Multiple Choice Questions (MCQs) are the commonest form of exams in medical education which are highly reliable; however their structural and content validity, taxonomy and structural principles have always been considered by researchers. This study was designed to evaluate quality of MCQs in the school medicine, Mazandaran University of Medical Sciences, regarding taxonomy and structural principles.

Methods: In this descriptive study, all MCQs of written exams held during a whole semester in the school of medical in Mazandaran University of Medical Sciences were evaluated. All questions were assessed according to taxonomy (in three levels of recall, comprehension, and application) and structure using a checklist which was based on Millman's principles for designing stem and choices of questions.

Results: Among 1478 questions related to 25 exams, questions related to lessons of basic sciences, physiopathology, and clerkship were 28.7%, 18.7%, 52.6% respectively. 77.4 percent of questions were in Taxonomy I and the rest were designed in Taxonomy II and III. 46 percent of all questions had no structural problems while the rest had one or more structural problems. Taxonomy and structural problems in exams related to clerkship and physiopathology were significantly less than basic science lessons ($p=0.023$ and $p=0.001$ respectively). Questions of higher taxonomy had less structural problems.

Conclusion: Our present study showed that the quality of MCQs regarding structural principles and taxonomy were different in exams of different levels which necessitate further revisions.

Key words: Educational improvement, Multiple Choice Question, Taxonomy, Millman's principles