DOI: 10.48305/jims.v44.i878.1754

مجله دانشکده پزشکی اصفهان — سال چهل و چهارم/شماره‌ی ۸۷۸/هفته‌ی اول آبان ۱۴۰۵

تاریخ دریافت: ۱۴۰۵/۰۶/۱۶ — تاریخ پذیرش: ۱۴۰۵/۰۷/۲۴ — تاریخ چاپ: ۱۴۰۵/۰۸/۰۷

بررسی و مقایسه‌ی تحلیل کیفی و سطح سؤالات ۴ دوره آزمون جامع علوم پایه پزشکی از سال ۱۳۹۹ تا ۱۴۰۲

شهرام شایان[۱]، محمد حسن عبدی اردستانی[۲]، وجیهه عسگری[۳]

مقاله پژوهشی

چکیده

مقدمه: این مطالعه با هدف ارزیابی کیفیت و سطح سؤالات چندگزینه‌ای آزمون‌های جامع علوم پایه پزشکی در فاصله‌ی سال‌های ۱۳۹۹ تا ۱۴۰۲ انجام گرفته است. آزمون‌ها، ابزارهایی کارآمد برای سنجش پیشرفت تحصیلی دانشجویان، ایجاد انگیزه‌ی یادگیری و اطمینان از تسلط بر دانش و مهارت‌های لازم محسوب می‌شوند. در این بین آزمون جامع علوم پایه پزشکی که در پایان اولین مرحله از آموزش پزشکی عمومی در ایران برگزار می‌شود، یکی از مهم‌ترین آزمون‌های حوزه‌ی علوم پزشکی است که رعایت دقیق اصول و مبانی یک آزمون مطلوب از ضروریات انجام آن به حساب می‌آید. لذا این مطالعه با تمرکز بر دو شاخص کیفیت سؤالات آزمون مشتمل بر ساختار شکلی سؤالات بر اساس چک‌لیست میلمن و سطح شناختی سؤالات بر اساس تاکسونومی بلوم، طراحی و اجرا گردیده است.

روش‌ها: مطالعه از نوع توصیفی- مقطعی بوده و تمامی سؤالات ۸ آزمون سراسری (هر آزمون ۲۰۰ سؤال) و در یک روند ۴ ساله (مجموعاً ۱۶۰۰ سؤال) مورد بررسی و تحلیل قرار گرفته است. سؤالات از نظر اشکالات ساختاری، شامل کامل بودن ساقه، تکیه بر محفوظات، وضوح و هماهنگی با اهداف آموزشی، مورد بررسی قرار گرفتند. سطح شناختی نیز به سه دسته پایین (تاکسونومی I)، متوسط (تاکسونومی II) و بالا (تاکسونومی III) تقسیم‌بندی شد.

یافته‌ها: نتایج نشان داد که ۱۱ درصد از سؤالات حداقل دارای یک اشکال ساختاری بودند. شایع‌ترین مشکلات، تکیه بر محفوظات به جای کاربرد اطلاعات (۵۷ درصد) و ناقص بودن ساقه سؤال (۴۱ درصد) بودند. توزیع سطح شناختی شامل ۵۶ درصد تاکسونومی I، ۳۷ درصد تاکسونومی II و ۸ درصد تاکسونومی III بود. خرده آزمون‌هایی مانند انگل‌شناسی، قارچ‌شناسی، حشره‌شناسی و انقلاب اسلامی عمدتاً سؤالات با تاکسونومی پایین داشتند، در حالی که درس علوم تشریح بهبود سطح شناختی را نشان داد. همچنین تفاوت معنی‌داری در کیفیت ساختاری سؤالات بین دروس مختلف، طی چهار سال مشاهده گردید ($P > ۰/۰۵$). مقایسه‌ی نتایج این پژوهش با مطالعات پیشین، وجود مشکلات ساختاری و تمرکز بر سطوح پایین شناختی را تأیید می‌کند.

نتیجه‌گیری: با وجود اینکه اکثر سؤالات، فاقد اشکال ساختاری بودند، غالب بودن سؤالات با تاکسونومی پایین و عدم بهبود کیفیت به صورت معنی‌دار در طول چهار سال نشان‌دهنده‌ی عدم مداخله کافی و مؤثر در فرایند طراحی آزمون‌هاست. پیشنهاد می‌شود اقداماتی شامل برگزاری کارگاه‌های آموزشی هدفمند برای طراحان سؤال، بهره‌گیری از بانک سؤالات ملی، بکار گیری جدول مشخصات آزمون (بلوپرینت) در روند طراحی و سایر روش‌های کنترل کیفیت برای ارتقاء روایی و کیفیت آزمون‌ها اعمال گردد.

واژگان کلیدی: آزمون‌های چندگزینه‌ای (MCQ)؛ تاکسونومی بلوم؛ تحلیل کیفی سؤالات؛ آزمون جامع علوم پایه پزشکی

ارجاع: شایان شهرام، عبدی اردستانی محمدحسن، عسگری وجیهه. **بررسی و مقایسه‌ی تحلیل کیفی و سطح سؤالات ٤ دوره آزمون جامع علوم پایه پزشکی از سال ۱۳۹۹ تا ۱٤۰۲.** مجله دانشکده پزشکی اصفهان ۱۴۰۵؛ ۴۴ (۸۷۸): ۱۷۵۴- ۱۷۶۲.

مقدمه

آزمون‌ها، مهم‌ترین وسیله‌ی اندازه‌گیری پیشرفت تحصیلی دانشجویان محسوب می‌شود (۱)؛ به طوری که هم باعث ایجاد انگیزه‌ی یادگیری می‌شوند و هم ابزاری هستند برای حصول اطمینان از اینکه دانش‌آموختگان به سطح مورد نظر از دانش و توانمندی رسیده‌اند. بدین جهت برنامه‌های ارتقای کیفیت و استانداردسازی آزمون‌ها از اثربخشی بالایی برخوردار است (۲). آزمون‌ها از نظر محتوای آزمون (دانش، مهارت و ...)، هدف آزمون (پیشرفت تحصیلی، گزینش داوطلبان و ...)، معیار قبولی و جنبه‌های گوناگون دیگر قابل طبقه‌بندی می‌باشند. برای آنکه هر آزمون نتایج قابل استناد داشته باشد، باید از روایی، پایایی و قابلیت اجرا برخوردار باشد که هر یک جنبه‌های متفاوتی از آزمون را در بر می‌گیرند (۳، ٤). آزمون‌های چندگزینه‌ای

۱- مرکز تحقیقات قلب و عروق، انستیتو قلب و عروق شهید رجایی، تهران، ایران

۲- کارشناس ارشد علوم تشریح، دانشجوی دکترای تخصصی علوم تشریح گروه علوم تشریحی و بیولوژی تولید مثل، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

۳- دکترای تخصصی علوم تشریح، استادیار، گروه علوم تشریحی و بیولوژی تولید مثل، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

نویسنده‌ی مسؤول: محمدحسن عبدی اردستانی: کارشناس ارشد علوم تشریح، دانشجوی دکترای تخصصی علوم تشریح گروه علوم تشریحی و بیولوژی تولید مثل، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران

Email: m.h.abdie@gmail.com

MCQ (Multiple choice Question) که اولین بار در سال ۱۹۱٤ در دانشگاه کانزاس توسط فردریک ابداع گردید، از رایج‌ترین نوع آزمون‌ها در آموزش پزشکی می‌باشند که به علت عینیت کامل، از پایایی بالا (حداکثر پایایی ممکن) برخوردار می‌باشند (۵، ٦) و از لحاظ یکنواختی سؤال‌ها، حساسیت کم در مقابل حدس کورکورانه نسبت به سؤالات صحیح و غلط و از نظر سهولت تصحیح پاسخنامه‌ها بهترین نوع آزمون‌های عینی هستند (۷)، اما روایی آنها به ویژه در مورد سطح دانش (تاکسونومی) و رعایت اصول طراحی ساختار MCQ از مسائل مورد توجه در مطالعات و پژوهش‌های آموزشی بوده است (٦).

مطالعه‌ی Tarrant و Ware در دانشکده‌ی پزشکی دانشگاه هنگ کنگ، نشان داد که حدود نیمی از سؤالات چندگزینه‌ای طراحی شده توسط اعضای هیأت علمی دارای اشکال ساختاری بوده و عمدتاً سطوح پایین شناختی را مورد سنجش قرار داده بودند (۸). در این زمینه، شواهد متعددی از جمله مطالعه Brady (۹)، Larsen و همکاران (۱۰) و Reid و همکاران (۱۱) نشان داده‌اند که کیفیت آزمون‌ها، به طور اساسی کیفیت آموزش و روند یاددهی- یادگیری را تحت تأثیر قرار می‌دهند.

همچنین مطالعه‌ی سیاح و همکاران در کاشان نشان داد که بخش عمده سؤالات چهار گزینه‌ای دانشکده پرستاری کاشان دارای اشکال است (۱۲). 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 به عنوان جزء تکمیل کننده و حیاتی چرخه آموزش از دست می رود، بلکه این امر برانگیزه‌های فراگیران اثرات منفی بر جای گذاشته و خطر به هدر رفتن تلاش اساتید و سیستم آموزشی را بدنبال خواهد داشت، لذا بررسی کیفیت سؤالات و میزان استاندارد بودن آزمون ها امری ضروری است (٦).

از طرفی سنجش از طریق آزمون‌های رسمی در دانشگاه‌ها نقش بنیادی در ارزیابی میزان یادگیری دانشجویان دارد و به اساتید و فراگیران کمک می‌کند تا نقاط قوت و ضعف آموزشی را شناسایی کرده و برنامه‌های درسی را اصلاح نمایند (۱۳). در دانشگاه‌های علوم پزشکی این اهمیت دوچندان است، زیرا ارزیابی معتبر و استاندارد تضمین کننده این است که دانشجویان پزشکی علاوه بر تسلط بر مبانی نظری، مهارت‌های بالینی، ارتباطی و تصمیم‌گیری موردنیاز برای ارائه خدمات ایمن و مؤثر را کسب کنند و از آنجا که دانش آموختگان این حوزه مستقیماً بر سلامت جامعه تأثیر می‌گذارند، ارتقای کیفیت آموزش و نظام‌های ارزیابی دقیق، نقشی کلیدی در حفظ و ارتقای سلامت عمومی دارد (١٤).

همچنین ذکر این نکته که یکی از وظایف مهم کارکرد آموزشی در دانشگاه‌ها، تربیت و آموزش دانشجویان دوره پزشکی عمومی می‌باشد لذا می‌بایست فرایند آموزشی این دانشجویان مورد ارزیابی دقیق قرار گیرد (۱۵).

برنامه‌ی تحصیلات پزشکی عمومی در ایران به‌طور معمول از دو مقطع اصلی تشکیل می‌شود، مقطع پایه شامل آموزش‌های علوم پایه و پسا پایه که زیربنای دانش نظری پزشک را می‌سازد و مقطع بالینی شامل کارآموزی بالینی و دوره‌ی کارورزی/اینترن شیپ که مهارت‌های بالینی و مواجهه با بیمار را فراهم می‌کند (١٦). دوره‌ی آموزشی علوم‌پایه پزشکی، پیش‌زمینه پیشرفت تحصیلی و فهم دقیق مطالب دوره‌های بعدی پزشکی عمومی است (۱۷) و در فهم دروس بالینی و ارزیابی وضعیت تحصیلی در مقاطع بعدی تحصیلی از اهمیت ویژه‌ای برخوردار است (۱۸). در این بین امتحان جامع علوم پایه پزشکی در پایان اولین مرحله از آموزش پزشکی عمومی در ایران برگزار می‌شود و یکی از مهم‌ترین آزمون‌های ارزشیابی دانشجویان است. این آزمون از سال ۱۳٦۸ همه ساله در دو نوبت شهریور و اسفند، به صورت کتبی با ۲۰۰ پرسش چهار گزینه‌ای برگزار می‌شود که حد نصاب قبولی در آن ۷۰ درصد میانگین نمره ۵ درصد افرادی است که بالاترین نمره را در کشور کسب کرده‌اند (۱۹).

آزمون جامع علوم پایه پزشکی یکی از مهم‌ترین آزمون‌های سرنوشت‌ساز در نظام آموزش پزشکی ایران است که به‌عنوان دروازه ورود دانشجویان از مرحله علوم پایه به دوره فیزیوپاتولوژی و آموزش‌های بالینی عمل می‌کند. این آزمون علاوه بر سنجش میزان دستیابی دانشجویان به دانش و شایستگی‌های پایه، نقش تعیین‌کننده‌ای در تصمیم‌گیری برای ارتقای تحصیلی، تضمین آمادگی جهت ورود به محیط‌های بالینی و حفظ کیفیت آموزش پزشکی دارد. از آنجا که عملکرد دانشجویان در این مرحله بر روند شکل‌گیری صلاحیت‌های حرفه‌ای و عملکرد بالینی آینده آنان تأثیر مستقیم می‌گذارد، کیفیت آزمون و به‌ویژه اعتبار و استاندارد بودن سؤالات آن از اهمیت ویژه‌ای برخوردار است. شواهد اخیر نیز تأکید می‌کنند که آزمون‌های سرنوشت‌ساز زمانی می‌توانند مبنای قابل اعتمادی برای تصمیمات آموزشی باشند که علاوه بر رعایت اصول استاندارد سؤال‌نویسی، تولنایی سنجش دانش کاربردی، استدلال بالینی و سطوح بالاتر شناختی را داشته باشند (۲۰، ۲۱).

اکثر مطالعات انجام شده در خصوص آزمون‌های جامع علوم پایه به تحلیل نتایج دانشجویان در آزمون‌ها و بررسی برخی عوامل مؤثر بر نتایج آنها پرداخته‌اند، طبق مطالعات لندک انجام شده در خصوص تحلیل کیفی آزمون‌های جامع علوم پایه، خواجه علی و همکاران در سال ۱۳۹۹ با تحلیل آزمون جامع دانشگاه جندی شاپور اهواز در سال

۱۳۹۳ مشخص کردند که در بررسی سؤالات پزشکی و دندانپزشکی، به ترتیب در ٦ و ٥ درصد موارد اشکال ساختاری وجود داشته است (۲۲).

تعیین سطح دانش به کار گرفته و اشکال‌های ساختاری سؤال‌های چندگزینه‌ای و بررسی علت‌ها و فراوانی آن‌ها موضوع با اهمیتی است که می‌تواند به کاهش نقاط ضعف و تقویت نقاط قوت در آزمون‌های چندگزینه‌ای در نهایت به بهبود کیفیت ارزشیابی در دانشگاه منجر شود. بررسی مستندات موجود نشان داد که مطالعات جامعی که سؤالات چندگزینه‌ای را هم از نظر سطح و تاکسونومی و هم از نظر رعایت اصول ساختاری سؤال بررسی کرده باشد، کمتر انجام شده است و بر اساس بررسی های صورت گرفته، تاکنون مطالعه‌ی طولی در سطح ملی که به‌طور همزمان، هم ایرادات ساختاری سؤالات و هم سطوح شناختی آزمون‌های جامع علوم پایه پزشکی ایران را در چندین سال متوالی مورد ارزیابی قرار داده باشد، انجام نشده است. لذا این مطالعه ساختار آزمون‌های چند گزینه‌ای جامع علوم پایه پزشکی را هم از نظر سطح و تاکسونومی و هم از نظر رعایت اصول ساختاری سؤال بین سال‌های ۱۳۹۹ تا ۱٤۰۲ مورد بررسی قرار داده است.

روش‌ها

این مطالعه از نوع توصیفی- مقطعی بود که در آن آزمون‌های جامع علوم پایه پزشکی برگزارشده به‌صورت سراسری (کشوری) در فاصله سال‌های ۱۳۹۹ تا ۱٤۰۲ مورد بررسی قرار گرفت. در این مطالعه کلیه سؤالات چندگزینه‌ای آزمون‌ها مشتمل بر ۱۱ خرده آزمون در هر سال از نظر ساختار و کیفیت طراحی سؤال مورد ارزیابی قرار گرفتند. ساختار سؤالات با استفاده از چک‌لیست تجزیه و تحلیل کیفی آزمون که توسط Millman در سال ۱۹۸۱ ارائه شده است، تحلیل گردید (۲۳).

جمع‌آوری داده‌ها و ابزار پژوهش

ابزار سنجش در این مطالعه یک پرسشنامه سه‌بخشی بود که پس از بومی‌سازی، به‌طور مکرر در مطالعات داخلی مورد استفاده قرار گرفته و استاندارد بودن و روایی آن براساس منابع متعدد توسط محققین داخلی تأیید شده است. بخش اول پرسشنامه شامل اطلاعات شناسنامه‌ای سؤال، از جمله رشته‌ی آموزشی، سال برگزاری آزمون و شماره سؤال بود.

ارزیابی کیفیت سؤالات

بخش دوم پرسشنامه شامل چک‌لیستی مبتنی بر اصول Millman و مشتمل بر ۱۸ آیتم در سه حیطه تنه سؤال (۳ آیتم)، گزینه‌ها (۹ آیتم) و کلیات سؤال (٦ آیتم) بود. برای هر آیتم، بر اساس رعایت یا عدم رعایت معیار مربوطه، به‌ترتیب نمره ۱ یا صفر اختصاص داده شد. بر این اساس، تنها سؤالاتی که مجموع امتیاز ۱۸ را کسب می‌کردند، به‌عنوان سؤالات فاقد اشکال ساختاری در نظر گرفته شدند. همچنین ملاک وجود یا عدم وجود اشکال ساختاری در هر یک از سه بخش تنه سؤال، گزینه‌ها و کلیات سؤال، دریافت یا عدم دریافت نمره کامل در همان بخش بود.

تعیین سطح شناختی سؤالات

بخش سوم پرسشنامه شامل یک سؤال بسته برای تعیین سطح تاکسونومی سؤالات بود که توسط متخصصین هر درس بر این اساس، سؤالات دارای تاکسونومی I به‌عنوان سؤالات با سطح شناختی پایین و سؤالات دارای تاکسونومی II و III به‌عنوان سؤالات با سطح شناختی بالاتر طبقه‌بندی شدند. برای تعیین سطح تاکسونومی و تحلیل کیفی سؤالات، از نظرات اساتید طراح سؤال آزمون جامع در گروه‌های علوم پایه پزشکی دانشگاه علوم پزشکی اصفهان و پاسخ آنان به یک سؤال بسته استفاده شد (۲٤).

برای تحلیل داده‌ها از نرم‌افزار SPSS ۲۱ (version 21, IBM Corporation, Armonk, NY) و آمار توصیفی (جداول توزیع فراوانی، شاخص‌های آماری و نمودار) و آزمون تحلیلی مقایسه‌ی نسبت‌ها (آزمون Z) با خطای آلفا کمتر از ۰/۰۵ و آزمون Chi-square استفاده گردید و فاصله اطمینان ۹۵٪ مدنظر قرار گرفت.

ملاحظات اخلاقی

این مطالعه پس از تصویب در کمیته اخلاق (IR.ARI.MUI.REC.1403.136) در معاونت تحقیقات و فناوری دانشگاه علوم پزشکی اصفهان انجام شد. از آنجا که پژوهش حاضر بر پایه تحلیل سؤالات آزمون‌های جامع علوم پایه پزشکی و بدون استفاده از اطلاعات هویتی شرکت‌کنندگان انجام شده است، تمامی داده‌ها به‌صورت محرمانه نگهداری و تحلیل شدند. همچنین در تمامی مراحل پژوهش، اصول اخلاق در پژوهش مطابق با دستورالعمل‌های دانشگاه و بیانیه هلسینکی رعایت گردید.

یافته‌ها

در مجموع ۱٦۰۰ سؤال مربوط به ۸ آزمون چند گزینه‌ای مورد بررسی قرار گرفت که هر آزمون مربوط به نوبت شهریورماه و یا اسفندماه یک سال از سال‌های ۱۳۹۹ الی ۱٤۰۲ و شامل ۲۰۰ سؤال چهارگزینه‌ای بود. سوالات بر مبنای درس (خرده آزمون) تفکیک شد و سپس سؤالات هر درس از نظر کیفیت و تاکسونومی طبق پرسشنامه‌ی مذکور بررسی گردید. در مجموع ۱۱ درصد از سؤالات دچار حداقل یک اشکال ساختاری بود و ٥٦ درصد سؤالات تاکسونومی I و ۳۷ درصد سؤالات تاکسونومی II و ۸ درصد سوالات تاکسونومی III را واجد بودند (جدول ۱). در خصوص تاکسونومی(سطح) سؤالات این آزمون‌ها به تفکیک هر سال در جدول ۱ مشخص گردیده که در مجموع خرده آزمون‌های قارچ‌شناسی، حشره‌شناسی و انگل‌شناسی، کمترین تاکسونومی را در طی این چهار سال

جدول ۱. مقادیر سطح یا تاکسونومی سؤالات آزمون‌های جامع علوم پایه پزشکی کشوری در سال‌های ۱۳۹۹ تا ۱۴۰۲ (P < ۰/۰۵)

سطح (تاکسونومی) سؤالات	۱۳۹۹ تعداد (درصد)	۱۴۰۰ تعداد (درصد)	۱۴۰۱ تعداد (درصد)	۱۴۰۲ تعداد (درصد)	کل	P
پایین	۲۲۸(۵۷/۲۹)	۲۲۸(۵۷/۲۹)	۲۳۴(۵۸/۷۹)	۲۰۰(۵۰/۲۵)	۸۹۰(۵۵/۶)	
متوسط	۱۴۲(۳۵/۱۸)	۱۳۸(۳۴/۱۷)	۱۴۰(۳۴/۶۷)	۱۶۸(۴۱/۷۱)	۵۸۸ (۳۶/۷)	۰/۲۱
بالا	۳۰(۷/۵۴)	۳۴(۸/۵۴)	۲۶(۶/۵۳)	۳۲(۸/۰۴)	۱۲۲(۷/۶)	

داشته است و تاکسونومی سؤالات در درس‌های اصول خدمات بهداشتی و درمانی (P = ۰/۰۰۹) و درس علوم تشریح (P = ۰/۰۰۱) در این چهار سال تفاوت معنی‌داری نشان داد.

در مورد رعایت اصول طراحی سوالات (اصول Millman)، نتایج نشان داد بیشترین اشکال در طی سال‌های اخیر در طراحی سؤالات آزمون‌های علوم پایه پزشکی مربوط به محور کلیات سؤال و زیر محور تکیه سوال بر محفوظات با ۵۷/۱۲ درصد درصد از کل سؤالات بود (P = ۰/۰۳) و پس از آن جامع بودن ساقه‌ی سؤال به نحوی که بدون مشاهده‌ی گزینه‌ها بتوان پاسخ داد با ۱۴/۱۲ درصد، بیشترین مشکلات در کیفیت سؤالات بود.

همچنین در مطالعه‌ی حاضر، در بخش مقایسه‌ی بین خرده‌آزمون‌های جامع علوم پایه، درس زبان انگلیسی با ۱۳/۵ درصد، درس قارچ‌شناسی با ۱۲/۵ درصد و اپیدمیولوژی با ۱۲/۱ درصد سؤالات مشکل‌دار در مجموع سال‌های ۱۳۹۹تا ۱۴۰۲ از مجموع خرده آزمون‌ها با بیشترین اشکال ساختاری بودند (P = ۰/۰۰۱) و درصد سؤالات مشکل‌دار مابقی خرده آزمون‌ها در محدوده‌ی ۵/۵ درصد تا ۱۰ درصد قرار داشت، همچنین در درس بیوشیمی نمره مربوط به گزینه‌های سؤالات بین سال‌ها تفاوت معنی‌دار داشت که در سال ۱۴۰۲ از نظر کیفیت دچار افت شده است. در درس باکتری‌شناسی، نمره‌ی آیتم اول مربوطه به بخش تنه‌ی سؤالات و کل محور تنه‌ی سؤالات بین سال‌های مورد مطالعه، تفاوت معنی‌دار مشاهده شد که به میزان ۳ درصد تا سال ۱۴۰۲ افت کیفیت داشته است ولی در درس اصول خدمات سلامت، معنی داری به سمت ارتقاء کیفیت از۹۳ به ۹۵ درصد بوده است و در درس ویروس‌شناسی نیز آیتم‌های مربوط به ساقه‌ی

جدول ۲. میزان رعایت اصول Millman در سؤالات آزمون‌های علوم پایه پزشکی سال‌های ۱۳۹۹ تا ۱۴۰۲ (P = ۰/۰۵)

سؤالات چک‌لیست Millman	سال ۱۳۹۹	سال ۱۴۰۰	سال ۱۴۰۱	سال ۱۴۰۲	جمع کل	P
ایده اصلی در ساقه سؤال گنجانده شده است.	۳۴۶ (۸۶/۵)	۳۶۴ (۸۶/۵)	۳۵۴ (۸۸/۵)	۳۶۸ (۹۰)	۱۴۳۲ (۸۹/۵)	۰/۸
بدنه سؤال واضح، روان و از نظر دستوری صحیح است.	۳۸۸ (۹۷)	۳۷۲ (۹۳)	۳۵۶ (۸۹)	۳۴۰ (۸۷)	۱۴۵۶ (۹۱)	۰/۳
از عبارات مبهم و قیود مطلق پرهیز شده است.	۴۰۰ (۱۰۰)	۳۶۸ (۹۲)	۳۹۸ (۹۹/۵)	۳۶۴ (۹۳)	۱۵۳۰ (۹۵/۶۳)	۰/۴
گزینه‌ها مستقل و غیرهمپوشانی هستند.	۴۰۰ (۱۰۰)	۳۷۶ (۹۴)	۴۰۰ (۱۰۰)	۳۶۸ (۹۴)	۱۵۴۴ (۹۶/۵)	۰/۵
گزینه‌ها همگن و مرتبط با یک موضوع واحد هستند.	۳۶۸ (۹۲)	۳۷۲ (۹۳)	۳۹۸ (۹۹/۵)	۳۶۰ (۹۲)	۱۴۹۸ (۹۳/۶۳)	۰/۵
طول گزینه‌ها تقریباً یکسان است.	۳۵۴ (۸۸/۵)	۳۵۴ (۸۸/۵)	۳۸۸ (۹۷)	۳۲۸ (۸۴)	۱۴۲۴ (۸۹)	۰/۱
هیچ سرنخی درباره گزینه صحیح داده نمی‌شود.	۳۹۸ (۹۹/۵)	۳۷۶ (۹۴)	۳۹۸ (۹۹/۵)	۳۷۲ (۹۵)	۱۵۴۴ (۹۶/۵)	۰/۶
از عبارات مبهم (غالباً، معمولاً، گاهی، هرگز) استفاده نشده است.	۳۹۸ (۹۹/۵)	۳۷۲ (۹۳)	۳۹۴ (۹۸/۵)	۳۷۲ (۹۵)	۱۵۳۶ (۹۶)	۰/۶
گزینه‌ها عمودی و مرتب شده‌اند.	۳۹۸ (۹۹/۵)	۳۷۶ (۹۴)	۴۰۰ (۱۰۰)	۳۷۲ (۹۵)	۱۵۴۶ (۹۶/۶۳)	۰/۶
تکرار عبارات در گزینه‌ها به حداقل رسیده است.	۳۴۲ (۸۵/۵)	۳۴۸ (۸۷)	۳۸۸ (۹۷)	۳۵۶ (۹۱)	۱۴۳۴ (۸۹/۶۳)	۰/۳
یک گزینه کاملاً صحیح وجود دارد.	۴۰۰ (۱۰۰)	۳۷۴ (۹۳/۵)	۳۹۴ (۹۸/۵)	۳۵۶ (۹۱)	۱۵۲۴ (۹۵/۲۵)	۰/۳
از گزینه‌های منفی برای ساقه منفی یا متضاد خودداری شده است.	۳۹۸ (۹۹/۵)	۳۷۶ (۹۴)	۴۰۰ (۱۰۰)	۳۶۴ (۹۳)	۱۵۳۸ (۹۶/۱۳)	۰/۴
سؤال یک موضوع مهم و اختصاصی را ارزیابی می کند.	۳۹۸ (۹۹/۵)	۳۳۰ (۸۲/۵)	۳۸۴ (۹۶)	۳۵۲ (۹۰)	۱۴۶۴ (۹۱/۵)	۰/۰۵
محور سؤال مناسب و محدود به یک مسئله یا موقعیت بالینی است.	۴۰۰ (۱۰۰)	۳۳۲ (۸۳)	۳۷۶ (۹۴)	۳۴۸ (۸۹)	۱۴۵۶ (۹۱)	۰/۰۵
سؤال مستقل از سایر سؤالات است.	۴۰۰ (۱۰۰)	۳۷۶ (۹۴)	۳۹۲ (۹۸)	۳۷۲ (۹۵)	۱۵۴۰ (۹۶/۲۵)	۰/۷۱
بدون نگاه کردن به گزینه‌ها می‌توان به سؤال پاسخ داد.	۲۰۲ (۵۰/۵)	۲۱۶ (۵۴)	۲۸۰ (۷۰)	۲۴۴ (۶۲)	۹۴۲ (۵۸/۸۸)	۰/۰۰۱
سؤال به جای محفوظات، توانایی به کارگیری اطلاعات را می‌سنجد.	۱۹۶ (۴۹)	۱۵۲ (۳۸)	۱۵۴ (۳۸/۵)	۱۸۴ (۴۷)	۶۸۶ (۴۲/۸۸)	۰/۰۳
سؤال مثبت طراحی شده است و لغات منفی هدایت کننده مشخص هستند.	۴۰۰ (۱۰۰)	۳۷۶ (۹۴)	۳۹۴ (۹۸/۵)	۳۷۲ (۹۵)	۱۵۴۲ (۹۶/۳۸)	۰/۶
کل	۶۵۸۶ (۹۱/۵)	۶۲۱۰ (۸۶)	۶۶۴۸ (۹۲)	۶۱۹۲ (۸۵/۷)	۲۵۶۲۴ (۸۸/۹۰)	<۰/۰۰۱

سؤال بهبود داشته است. درس علوم تشریح در کیفیت ساقه‌ی سؤال معنی‌داری بین سال‌های مختلف افت ٦ درصدی نشان داده است و در مجموع در سال ١٤٠٢ کیفیت سؤالات کاهش داشته است (P < ٠/٠٠١) (جدول ٢).

مقایسه‌ی اشکالات ساختاری سؤالات در همه خرده آزمون‌های آزمون جامع علوم پایه نشان داد که محدوده‌ی اشکالات ساختاری در آزمون‌ها بر اساس محورهای چک‌لیست Millman از صفر تا ٦٢ درصد و این اختلاف در هر سال معنی‌دار بود (P = ٠/٠٠١). همچنین در توزیع فراوانی کلی میزان رعایت اصول طراحی سؤالات بین سال‌های ١٣٩٩ الی ١٤٠٢ تفاوت معنی‌دار مشاهده شد و به طور کلی ١١/١ درصد از سؤالات دارای حداقل یک اشکال بودند (P = ٠/٠٠١). فراوانی اشکالات در هر آیتم از چک‌لیست در کل آزمون‌ها و به تفکیک سال در جدول ٢ نشان داده شده است.

بحث

در این مطالعه، دو شاخص مهم کیفیت آزمون‌های چند گزینه‌ای شامل تاکسونومی یا سطح سؤال و اصول Millman در مورد سؤالات آزمون‌های جامع علوم پایه کشوری بین سال‌های ١٣٩٩ تا ١٤٠٢ مورد ارزیابی قرار گرفت که در مورد شایع‌ترین اشکالات در ساختار سؤالات، نتایج نشان داد که در کل آزمون‌های مورد بررسی، کامل نبودن ساقه‌ی سؤال و تکمیل شدن منظور طراح سؤال بر مشاهده‌ی گزینه‌ها با ٤١ درصد و نیز ارزیابی محفوظات صرف بجای توان بکارگیری اطلاعات در موقعیت‌های جدید با ٥٧ درصد در کل سؤالات مشهود بود.

در این زمینه، مطالعه‌ی جوادی و همکاران در بم نشان داد، شایع‌ترین اشکالات در سؤالات، «درج مطالب تکراری در گزینه‌ها» و «وجود لغات منفی در ساقه‌ی سؤال» بود (٢٥).

نتایج مطالعه‌ی شکورنیا و همکاران در اهواز در مورد شایع‌ترین اشکالات در ساختار سؤالات نشان داد که در کل آزمون‌های مورد بررسی، عدم رعایت «چیدمان عمودی گزینه‌ها»، «همسنگ بودن گزینه‌ها» و «آوردن بخش اعظم اطلاعات در ساقه‌ی سؤال»، در بیش از ١٠ درصد سؤالات وجود داشت (٢٦).

همچنین در مطالعه‌ی طاهری و همکاران در مورد آزمون‌های دستیاری پزشکی، شایع‌ترین اشکالات در ساختار سؤالات «هم سنگ نبودن گزینه‌ها» و « منفی بودن ساقه‌ی سؤال» بود (٢٧) و در مطالعه‌ی حق شناس و همکاران نیز هم سنگ نبودن گزینه‌ها و نیاوردن بخش اعظم اطلاعات در ساقه‌ی سؤال ذکر شده بود (٦).

همچنین در مطالعات بین‌المللی نیز Hammond و همکاران از دانشگاه هامپتون نشان دادند که بخش اعظم سؤالات دارای حداقل یک مشکل ساختاریست که مجموع این مطالعات، نتایج مطالعه‌ی حاضر را تأیید می‌کند (٢٨).

درمقایسه با مطالعه‌ی خواجه علی و همکاران که آزمون‌های علوم پایه پزشکی و دندانپزشکی سال ١٣٩٣ که به صورت منطقه‌ای و در دانشگاه جندی شاپور اهواز برگزار شده بود را بررسی نمودند. سؤالات علوم پایه پزشکی ٦ درصد در مجموع دارای اشکال ساختاری بوده است به طوری که بیشترین فراوانی اشکالات ساختاری در سؤالات، وجود اشکال در ساقه‌ی سؤال بوده است که همین مورد نیز در سؤالات دندانپزشکی نیز بیشترین فراوانی را داشته و مطالعه‌ی حاضر نشان داد در سال‌های اخیر این عدد تا ١١/١ درصد افزایش داشته است (٢٢).

مطالعه‌ی Masters و همکاران نیز نشان داد که در آزمون پرستاری با اینکه بخش اعظم سؤالات از بانک سؤالات انتخاب شده بود، دارای اشکالات جزئی و برخی از سوالات دارای اشکالات اساسی بود (٢٩).

Jozefowicz و همکاران، وجود اشکالات زیاد در سؤالات امتحانی را عدم صرف وقت کافی در این ارتباط می‌دانند. آنها معتقدند، اساتید وقت زیادی را برای آماده کردن مطالب درسی و رفتن به کلاس اختصاص می‌دهند اما برای طراحی سؤالات امتحانی خود وقت زیادی صرف نمی‌کنند و اطلاعات آنها از اصول ارزشیابی و فنون طرح سؤالات امتحانی ناکافی است (٣٠).

مطالعه‌ی طاهری و همکاران در این زمینه نشان داد، مداخله با برگزاری ١٢ دوره کارگاه آموزشی یک روزه جهت طراحان آزمون (به تفکیک گروه‌های آموزشی) در طی یک سال، باعث رشد استانداردهای کیفیت آزمون گردید و نسبت سؤالات بدون اشکال از ٦٦ درصد قبل از مداخله به ٧٥ درصد بعد از انجام مداخله ارتقا یافت همچنین در خصوص سطح یا تاکسونومی نیز سؤالات با تاکسونومی بالا از ١ درصد به ١٠ درصد ارتقا پیدا کرد (٢٧)؛ بنابراین، برای جلوگیری از تداوم این روند، پیشنهاد می‌شود طراحی آزمون‌ها بر اساس Blueprint و اهداف یادگیری انجام شود، هر سؤال پیش از ورود به آزمون توسط حداقل دو ارزیاب مستقل بر اساس چک‌لیست Millman و سطح شناختی Bloom بررسی شود و پس از اجرای آزمون نیز شاخص‌های دشواری، تمایز و عملکرد گزینه‌های انحرافی به‌صورت منظم تحلیل شوند. ایجاد بانک سؤال پویا، حذف یا اصلاح سؤالات دارای عملکرد روان‌سنجی نامطلوب، ارائه بازخورد به طراحان و برگزاری دوره‌های مستمر و هدفمند توانمندسازی اعضای هیئت علمی می‌تواند به‌عنوان راهبردی عملی برای ارتقای اعتبار، پایایی و کیفیت آزمون‌های علوم پایه مورد استفاده قرار گیرد.

بررسی تاکسونومی سؤالات در مطالعه‌ی حاضر بین سال‌های

۱۳۹۹ تا ۱۴۰۲ تغییر معنی‌داری نداشته و در مجموع ۵۶ درصد سؤالات تاکسونومی I و ۳۷ درصد تاکسونومی II و ۸ درصد تاکسونومی III بوده است. همچنین مشخص شد در خرده آزمون‌های انگل‌شناسی، انقلاب اسلامی، حشره‌شناسی و قارچ‌شناسی اکثر سؤالات با تاکسونومی پایین بوده و در درس علوم تشریح بین سال‌های مورد مطالعه از نظر تاکسونومی ارتقا داشته است.

همچنین مشخص گردید تاکسونومی سؤالات بین سال‌های اخیر و حتی نسبت به سال‌های گذشته تغییر معنی‌داری نداشته است. در مورد تاکسونومی سؤالات نتایج این پژوهش با پژوهش‌های قبلی همخوانی داشت، بدین‌گونه که در مطالعه‌ی طاهری و همکاران، اکثر سؤالات حتی پس از مداخله هنوز سطوح یادگیری پایین را مورد هدف قرار می‌داد (۲۷).

همچنین در مطالعه‌ی شکوری‌نیا نیز مشخص شد، دو سوم سؤالات در تاکسونومی یک طبقه‌بندی گردیده است (۲۶). در مطالعات بین‌المللی نیز اهمیت کیفیت سؤالات چندگزینه‌ای و ارتباط آن با روایی و اثربخشی ارزشیابی، مورد توجه قرار گرفته است. به‌طور مثال، بررسی گسترده اصول و فنون نگارش سؤالات چند گزینه‌ای نشان داده است که استفاده از دستورالعمل‌های استاندارد می‌تواند به بهبود سنجش سطوح بالای شناختی کمک کند و از مشکلات رایج در ساختار سؤالات جلوگیری نماید (۳۱).

همچنین تحقیقات در بستر ابزارهای خودکار و هوش مصنوعی نشان داده‌اند که رویکردهای نوین می‌توانند خطاهای نوشتاری در سؤالات را به‌طور مؤثر شناسایی کنند و به ارتقای کیفیت آزمون‌ها کمک کنند (۳۲).

مطالعات دیگر نیز رابطه‌ی معنی‌داری بین خطاهای نگارشی سؤالات و شاخص‌های دشواری و تمایز در آزمون‌ها پیدا کرده‌اند، که بر لزوم توجه به اصول طراحی سؤالات تأکید می‌کند (۳۳). لازم به ذکر است با توجه به اهمیت تعیین حدنصاب قبولی در آزمون‌های جامع علوم پایه، می بایست تعیین حداقل سطح قبولی (MPL) با استفاده از روش‌های استانداردگذاری معتبر و مبتنی بر حداقل شایستگی مورد انتظار نیز مورد توجه قرار گیرد. تعیین علمی حدنصاب می‌تواند به افزایش عدالت و اعتبار تصمیمات قبولی و مردودی کمک کند و از وابستگی نتایج به یک حدنصاب ثابت جلوگیری نماید (۳۴).

از سوی دیگر، تحقیقات نشان می‌دهند که تحلیل پس از آزمون (Item analysis) ابزاری ضروری برای بازنگری سؤالات و توسعه‌ی بانک سؤالات معتبر است، به‌ویژه زمانی که هدف، ارزیابی دقیق سطوح مختلف شناختی باشد (۳۵). بررسی‌های متعدد نیز بیانگر آن بود که حتی در سیستم‌های آموزشی پیشرفته، بخش زیادی از سؤالات در سطوح پایین شناختی قرار دارند و نیازمند بازنگری بیشتر هستند (۳۶).

علاوه بر این، تحلیل کیفیت سؤالات با استفاده از شاخص‌های روان‌سنجی مانند ضریب تمایز و دشواری، به بهبود اعتبار و قابلیت اعتماد آزمون‌ها کمک می‌کند و برای توسعه ارزشیابی‌های معتبر در آموزش پزشکی ضروری است (۳۷). گرچه مطالعات جدید تر نشان داده است که رویکرد طراحان به سمت طرح سؤالات با تاکسونومی III بیشتر شده اما سایر مطالعات و همچنین موارد مشابه داخلی نشان می‌دهد که کماکان تمرکز عمده طراحان سؤال بر محفوظات دانشجویان و نه فهم عمیق و تحلیل و کاربرد مفاهیم، بوده است و این امر 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 دانشجو برخوردار نیست، بلکه از کارکرد لازم در زمینه‌ی پایه‌ریزی بستر دانش آینده و گسترش توانایی‌های ذهنی دانشجو کاملاً دور است. همچنین باید به این حقیقت اذعان کرد که یکی از خصوصیات آزمون‌های چند گزینه‌ای ماهیتاً ارزیابی دانش در حیطه‌ی حافظه است و ارزیابی تاکسونومی‌های بالا با این مدل آزمون دور از انتظار می‌باشد.

نتیجه‌گیری

اگرچه در تمامی آزمون‌های بررسی شده در این مطالعه بیش از نیمی از سؤالات فاقد اشکال ساختاری بودند اما نقص در رعایت اصول ساختاری به هر میزان، روایی سازه را در آزمون مختل نموده و در این صورت کارکردهای آموزشی را تحت تأثیر قرار می‌دهد. اساساً گرایش به طراحی سؤالات آسان و بسیار آسان باعث اخلال در مشخصات روان‌سنجی آزمون اعم از ضریب دشواری و بدنبال آن قدرت تمیز سؤال خواهد شد. همچنین نبود معنی‌داری تاکسونومی و کیفیت سؤالات آزمون‌های جامع علوم پایه کشوری در بین چهار سال متوالی در این مطالعه و استفاده بیشتر از سؤالات با تاکسونومی‌های پایین نشان‌دهنده‌ی نقص در مداخله‌ی مؤثر توسط مدیران و مسوولان فنی آزمون‌های کشوری در این زمینه و ضعف در بهبود کیفیت سؤالات است که ضرورت اقدامات مداخله‌ای در این زمینه شامل برگزاری کارگاه‌ها آموزشی هدفمند و آموزش طراحان سؤال، استفاده از بانک سؤالات ملی و کاربست جدول مشخصات آزمون (بلوپرینت) در فرایند طراحی سؤالات و سایر تکنیک‌های ارزیابی سؤالات قبل از برگزاری آزمون‌های جامع علوم پایه پزشکی، جهت ارتقاء روایی و کیفیت آن‌ها را ضروری می‌نماید.

همچنین پیشنهاد می‌گردد سایر ویژگی‌های روان‌سنجی این آزمون‌ها مورد بررسی قرار گرفته و نشان داده شود که آیا صرفاً برگزاری یک آزمون چند گزینه‌ای می‌تواند ملاک ارزشیابی همه‌ی پارامترهای یادگیری علوم پایه پزشکی باشد یا بایستی از سایر روش‌های ارزیابی که قابلیت سنجش فرایندهای عمیق ذهنی را دارد، بهره جست. همچنین با توجه به اهمیت تصمیم‌گیری عادلانه درباره‌ی قبولی در آزمون‌های جامع علوم پایه، پیشنهاد می‌شود حداقل سطح قبولی (MPL) با استفاده از روش‌های استانداردگذاری معتبر، مانند Angoff، تعیین و با حدنصاب ثابت موجود مقایسه شود. بررسی تأثیر MPL بر نرخ قبولی، پایایی و اعتبار تصمیمات آزمون می‌تولند به تعیین حدنصابی مبتنی بر حداقل شایستگی مورد انتظار و ارتقای کیفیت و عدالت آموزشی کمک کند.

تشکر و قدردانی

در پایان از اساتید گروه‌های علوم پایه دانشکده‌ی پزشکی و نیز سرکار خانم نازنین زارع و سرکار خانم مریم برون دانشجویان پزشکی دانشکده‌ی پزشکی اصفهان که در طی انجام مراحل کار ما را یاری نمودند تشکر و قدردانی می‌نماییم.

References

1. Kaveh Tabatabaee MS, Bahreyni Toosi MH, Derakhshan A, Khajeh Dalloee M, Gholami H. An Analytical Review of Multiple-Choice Examinations at the Faculty of Medicine, Mashhad University of Medical Sciences, in the Second Semester of the 1979–1980 Academic Year [in Persian]. Med J Mashhad Univ Med Sci 45(2): 89-95.
2. Schuwirth LW, van der Vleuten CP. A history of assessment in medical education. Advances in Health Sciences Education 2020; 25(5): 1045-56.
3. McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. Medical Teacher 2004; 26(8): 709-12.
4. Seif A. Educational measurement, assessment and evaluation. Tehran: AGAH. 2007:60-132.
5. Azizi F, Rezaei Ghale N, Riahi Asl N. Medical education: challenges and prospects [in Persian]. Tehran, Iran: Vice Chancellor and Student Affairs, Ministry of Health and Medical Education; 2003.
6. Haghshenas M, Vahidshahi K, Mahmudi M, Shahbaznejad L, Parvinnejad N, Emadi A. Evaluation of multiple choice questions in the school of Medicine Mazandaran University of Medical Sciences the first Semester of 2007 [in Persian]. Strides in development of Medical Education. 2009; 5(2): 120-7.
7. Chernoff EJ, Mamolo A, Zazkis R. An investigation of the representativeness heuristic: The case of a multiple choice exam. Eurasia J Math Sci Tech Ed 2016; 12(4): 1009-31.
8. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. Med Educ 2008; 42(2): 198-206.
9. Brady A-M. Assessment of learning with multiple-choice questions. Nurse Educ Pract 2005; 5(4): 238-42.
10. Larsen DP, Butler AC, Roediger III HL. Test-enhanced learning in medical education. Med Educ 2008; 42(10): 959-66.
11. Reid WA, Duvall E, Evans P. Relationship between assessment results and approaches to learning and studying in year two medical students. Med Educ 2007; 41(8): 754-62.
12. Sayyah M, Vakili Z, Masoudi Alavi N, Bigdeli M, Soleymani A, Assarian M, et al. An item analysis of written multiple-choice questions: Kashan University of Medical Sciences [in Persian]. Nurs Midwifery Stud 2012; 1(2): 83-7.
13. Lakhtakia R, Otaki F, Alsuwaidi L, Zary N. Assessment as learning in medical education: feasibility and perceived impact of student-generated formative assessments. JMIR Med Educ 2022; 8(3): e35820.
14. Maeshiro R, Jarris YS, Prunuske J, Carney JK, Strelnick AH, Townsend JM, et al. Public and population health in US medical education: a review of guidance in extraordinary times. Acad Med 2023; 98(12): 1366-80.
15. Nasri K, Kahbazi M, Nasri S. Medical Students' Viewpoints toward Basic Sciences and Preinternship Comprehensive Exams in Arak University of Medical Sciences [in Persian]. Iranian Journal of Medical Education 2010; 10(1): 82-91.
16. Ahmady S, Khani H. The situational analysis of teaching-learning in clinical education in Iran: a postmodern grounded theory study. BMC Med Educ 2022; 22(1): 520.
17. Namdari P, Ebrahimzadeh F, Mardani M. Study of effective factors on comprehensive test of basic medical sciences of the medical students of Lorestan University of Medical Sciences [in Persian]. Yafte 2010; 12(1): 5-12.
18. Roudbari M, Dadgar F. Effective factors on the results of the basic sciences examinations at Zahedan University of Medical Sciences [in Persian]. J Inflamm Dis 2004; 8(8): 32-9.
19. Roudbari M, Sheibak A. The survey of the result and the associated factors in the preinternship examination at Zahedan University of Medical Sciences in September & Mars 2001 [in Persian]. Zahedan J Res Med Sci 2006; 8(1): e94916.
20. Bhanji F, Naik V, Skoll A, Pittini R, Daniels VJ, Bacchus CM, et al. Competence by design: the role of high-stakes examinations in a competence based medical education system. Perspect Med Educ 2024; 13(1): 68-74.

21. Staudenmann D, Waldner N, Lörwald A, Huwendiek S. Medical specialty certification exams studied according to the Ottawa Quality Criteria: a systematic review. BMC Med Educ 2023; 23(1): 619.
22. Khajeali N, Aslami M, Araban M. Analysis of medical and dentistry basic sciences examinations: A case study [in Persian]. Payesh 2020; 19(4): 383-9.
23. Millman J. A checklist procedure. New techniques for evaluation. Beverly Hills, CA: Sage Publications; 1981. p. 309-20.
24. Vyas R, Supe A. Multiple choice questions: a literature review on the optimal number of options. Natl Med J India 2008; 21(3): 130-3.
25. Javadi M, Abbas Zadeh A, Borhani F, Abdoli R. Analytical survey of summative MCQs of Bam Nursing faculty. Proccedings of the 8th National Congress of Medical Education Kerman: Kerman Univerity of Medical Sciences; 2007. [in Persian]
26. Shakurnia A, Ghafourian Boroujerdnia M, Khodadadi A, Ghadiri A, Amari A. Analytical Study of Quantitative Indices of Multiple-choice Questions of Immunology Department in Ahvaz Jundishapur University of Medical Sciences [in Persian]. Educational Development of Judishapur 2018; 9(2): 72-83.
27. Taheri M, Khoshrang H, Asadi Louyeh A, Hidarzadeh A. Quality of residents' promotion exams before & after educational intervention in 2010-2011 [in Persian]. Iranian Journal of Medical Education 2013; 13(7): 551-60.
28. Hammond E, McIndoe A, Sansome A, Spargo P. Multiple-choice examinations: adopting an evidence-based approach to exam technique. Anaesthesia 1998; 53(11): 1105-8.
29. Masters JC, Hulsmeyer BS, Pike ME, Leichty K, Miller MT, Verst AL. Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. J Nurs Educ 2001; 40(1): 25-32.
30. Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH. The quality of in-house medical school examinations. Acad Med 2002; 77(2): 156-61.
31. Schmauder R, Daugherty K, Rudolph M, Hutson B, McCarthy K, McCauley S, et al. The Art and Science of Item Writing: A Review of Established Guidelines for Multiple-Choice Questions. Intersection: A Journal at the Intersection of Assessment and Learning 2025; 6(3): 222–39.
32. Moore S, Nguyen HA, Chen T, Stamper J. Assessing the quality of multiple-choice questions using gpt-4 and rule-based methods. European conference on technology enhanced learning; 2023: Springer.
33. Schmucker R, Moore S. The impact of item-writing flaws on difficulty and discrimination in item response theory. Computers and Education: Artificial Intelligence 2026; 11: 1-15.
34. Heidari A, Shayan S. Determining the Passing Score as A Strategy to Achieve Justice in Assessment: A Review Study [in Persian]. J Isfahan Medi Sch 2025; 43(802): 53-62.
35. Rashwan NI, Aref SR, Nayel OA, Rizk MH. Postexamination item analysis of undergraduate pediatric multiple-choice questions exam: implications for developing a validated question Bank. BMC Med Educ 2024; 24(1): 168.
36. Iñarrairaegui M, Fernandez-Ros N, Lucena F, Landecho MF, García N, Quiroga J, et al. Evaluation of the quality of multiple-choice questions according to the students' academic level. BMC Med Educ 2022; 22(1): 779.
37. Hye T, Hossain MF, Roni MA. Interventions to Improve the Quality of Multiple-Choice Questions: A Systematic Review and Meta-Analysis. Am J Pharm Educ 2025; 89(11): 101872.

DOI: *10.48305/jims.v44.i878.1754*

Journal of Isfahan Medical School **Vol. 44, No. 878, 1st Week, November 2026**

Received: 07.09.2026 Accepted: 16.10.2026 Published: 29.10.2026

A Comparative Analysis of Item Quality and Cognitive Levels in Four National Comprehensive Basic Medical Sciences Examinations (2020–2023)

Shahram Shayan[1], Mohammad Hassan Abdi Ardestani[2], Vajiheh Asgari[3]

Original Article

Abstract

Background: This study aimed to evaluate the quality of multiple-choice questions (MCQs) used in the Iranian National Comprehensive Basic Medical Sciences Examination between 2020 and 2023. Examinations are essential educational tools for assessing students' academic achievement, promoting learning motivation, and ensuring the acquisition of the knowledge and competencies required for professional practice. Among these, the National Comprehensive Basic Medical Sciences Examination, administered at the completion of the preclinical phase of undergraduate medical education in Iran, represents one of the most important high-stakes assessments. Therefore, strict adherence to established principles of high-quality test construction is essential. The present study focused on two key indicators of question quality: structural item-writing quality based on the Millman checklist and cognitive level according to Bloom's taxonomy.

Methods: This descriptive cross-sectional study analyzed all questions from eight consecutive nationwide Comprehensive Basic Medical Sciences Examinations conducted over a four-year period (2020–2023). Each examination consisted of 200 MCQs, yielding a total of 1,600 questions. Structural quality was evaluated using the Millman item-writing checklist, including stem completeness, reliance on factual recall, clarity, and alignment with intended educational objectives. Cognitive level was classified into three categories according to Bloom's taxonomy: lower-order (Taxonomy I), intermediate (Taxonomy II), and higher-order (Taxonomy III).

Findings: Overall, 11% of the questions contained at least one structural item-writing flaw. The most frequent deficiencies were excessive reliance on factual recall rather than application of knowledge (57%) and incomplete question stems (41%). Regarding cognitive level, 56% of the questions were classified as Taxonomy I, 37% as Taxonomy II, and only 8% as Taxonomy III. Subtests such as Parasitology, Mycology, Medical Entomology, and Islamic Revolution Studies predominantly consisted of lower-order cognitive questions, whereas the Anatomy subtest demonstrated improvement in cognitive complexity over time. A statistically significant difference in structural quality was observed among different subject areas across the four-year study period ($P < 0.05$). Comparison with previous studies confirmed the persistence of structural item-writing flaws and the predominance of lower-order cognitive questions.

Conclusion: Although the majority of questions were free of structural item-writing flaws, the predominance of lower-order cognitive questions and the absence of meaningful improvement in overall examination quality over the four-year period suggest insufficient and ineffective interventions in the test development process. Implementing targeted faculty development workshops for item writers, establishing a national item bank, incorporating examination blueprinting into the test development process, and adopting additional quality assurance measures are recommended to enhance the validity and overall quality of future examinations.

Keywords: Multiple-choice questions (MCQs); Bloom's taxonomy; Item quality analysis; National Comprehensive Basic Medical Sciences Examination

Citation: Shayan Sh, Abdi Ardestani MH, Asgari V. **A Comparative Analysis of Item Quality and Cognitive Levels in Four National Comprehensive Basic Medical Sciences Examinations (2020–2023).** J Isfahan Med Sch 2026; 44(878): 1754- 62.

1- Rajaie Cardiovascular Medical and Research Center, Rajaie Cardiovascular Institute, Tehran, Iran
2- MSc in Anatomical Sciences, PhD Candidate in Anatomical Sciences, Department of Anatomical Sciences and Reproductive Biology, Isfahan University of Medical Sciences, Isfahan, Iran
3- PhD in Anatomical Sciences, Department of Anatomical Sciences and Reproductive Biology, Isfahan University of Medical Sciences, Isfahan, Iran
Corresponding Author: Mohammad Hassan Abdi Ardestani, MSc in Anatomical Sciences, PhD Candidate in Anatomical Sciences, Department of Anatomical Sciences and Reproductive Biology, Isfahan University of Medical Sciences, Isfahan, Iran; Email: m.h.abdie@gmail.com