بررسی نتایج آنالیز کمی و کیفی سؤالات چهارگزینه‌ای دوره مقدمات پزشکی بالینی

وجیهه وفامهر، محمد دادگسترنیا*

چکیده

مقدمه: در اغلب دانشکده‌های پزشکی، حتی قبل از اجرای آزمون‌های بالینی، یک آزمون چندگزینه‌ای برای اطمینان از دانش پایه‌ای فراگیران برگزار می‌شود.

روش‌ها: در دوره مقدمات پزشکی بالینی، ۳۹۷۳سؤال ۴ گزینه‌ای مربوط به آزمون‌های یک دوره چهار ساله از ۱۳۸۴ تا ۱۳۸۸ جمع‌آوری شد. سپس این سؤالات از نظر ۱۰ آیتم کیفی مربوط به وجود خطا در ساختار سؤال، سطح شناختی سؤال و ضریب دشواری و تمییز مورد بررسی قرار گرفتند.

نتایج: نتایج گویای آن بود که ۶۶/۹ درصد از سؤالات دارای حداقل یک خطا در ساختار خود بودند. ۷۹/۱ درصد سؤالات در سطوح پایین شناختی طراحی شده بودند. سؤالاتی که دارای سطح پایین شناختی بودند به طور معناداری دارای خطاهای ساختاری بیشتری بودند. ۴۷/۳ درصد سؤالات دارای ضریب تمییز مناسب و ۲۸/۸ درصد سؤالات دارای ضریب دشواری مناسب بودند. در حالی که ارتباطی بین ضریب دشواری و سطح شناختی سؤالات به دست نیامد، سؤالات با سطح شناختی بالاتر با ضریب تمییز مناسب رابطه معناداری داشتند.

نتیجه‌گیری: برای بهبود کیفیت آزمون‌ها و استفاده از نتایج تجزیه و تحلیل کمی و کیفی سؤالات می‌بایست راه‌کارهای عملی اندیشیده شود. اساتید می‌بایست در زمینه طراحی و استفاده از سؤالات مناسب و تهیه بانک سؤالات چهارگزینه‌ای آموزش داده شوند.

واژه‌های کلیدی: آنالیز سؤالات، بانک سؤال، سؤالات چهار گزینه‌ای

مجله ایرانی آموزش در علوم پزشکی (ویژه‌نامه توسعه آموزش) / زمستان ۱۳۸۹؛ ۱۰(۵): ۱۱٤٦ تا ۱۱۵۲

مقدمه

آزمون‌های کتبی هسته مرکزی ارزشیابی دانشجویان پزشکی را تشکیل می‌دهد. از این میان به خصوص آزمون‌های چندگزینه‌ای (MCQs) به خاطر سهولت و عینیت در نمره‌دهی، بیشتر مورد استفاده قرار می‌گیرند(۱تا٤). اغلب اوقات حتی قبل از اجرای آزمون‌های بالینی، یک آزمون چندگزینه‌ای به منظور اطمینان از دانش پایه‌ای دانشجویان برگزار می‌شود(۱). از طرف دیگر پایاترین روش مورد استفاده برای اندازه‌گیری دانش فراگیران، آزمون‌های چند گزینه‌ای است. سؤالات چندگزینه‌ای خوب طراحی شده، فراتر از یک یادآوری ساده اطلاعات هستند. چنین سؤالاتی آزمون‌شونده را مجبور می‌کنند که برای یک مشکل کلینیکی، به تحلیل و نتیجه‌گیری از مطالب تئوری ذهن خود بپردازد(۵). این سؤالات می‌تواند برای سنجش اغلب سطوح یادگیری به کار برده شود. همچنین توانایی اندازه‌گیری بیشتر اهداف آموزشی را دارا

نویسنده مسؤول: دکتر محمد دادگسترنیا، کارشناس ارشد آموزش پزشکی و مسؤول دوره مقدمات پزشکی بالینی دانشکده پزشکی، دانشگاه علوم پزشکی، اصفهان، ایران.
(dadgostar@med.mui.ac.ir)

دکتر وجیهه وفامهر، کارشناس ارشد آموزش پزشکی و کارشناس ارزشیابی دوره مقدمات پزشکی بالینی دانشکده پزشکی، دانشگاه علوم پزشکی، اصفهان، ایران. vgrenani@yahoo.com

این مقاله در تاریخ ۸۹/۱۰/۱۱ به دفتر مجله رسیده، در تاریخ ۸۹/۱۱/۲۱ اصلاح شده و در تاریخ ۸۹/۱۱/۲۵ پذیرش گردیده است.

[Downloaded from ijme.mui.ac.ir on 2026-11-09]

می‌باشد(٦). از طرف دیگر آزمون‌های چهارگزینه‌ای توانایی ارزیابی تعداد زیادی دانشجو با بکارگیری حداقل نیروی انسانی را دارند (٤). استفاده روزافزون از آزمون‌های چندگزینه‌ای در ارزشیابی دانشجویان پزشکی باعث ایجاد بعد جدیدی از مدیریت آزمون‌ها شده است. ذخیره کردن و آزمون مجدد تعداد زیادی سؤال، به همراه حجم بالای داده‌های عددی منتج از item- analysis سؤالات، باعث ایجاد یک مشغله پاره وقت برای اساتید پزشکی شده است(٧).

در اغلب دانشکده‌های پزشکی ایران هنوز ارزشیابی دانشجویان حتی در اغلب بخش‌های بالینی هم توسط آزمون‌های کتبی و به خصوص MCQ انجام می‌شود. طی سال‌های متوالی کیفیت پایین این آزمون‌ها در دانشکده‌های پزشکی ما همواره به چشم می‌خورد. آگاهی کم اساتید از اصول طراحی سؤال خوب، عدم ارائه باز خورد مناسب به ایشان و فرصت کم اساتید برای بررسی سؤالات، از علل این چالش به نظر می‌رسد. این مسأله به خصوص در مورد سؤالات طراحی شده توسط اساتید بالینی بیشتر است. چرا که ایشان با توجه به مشغله کاری فراوان‌تری که دارند، یا فرصت کمتری برای شرکت در کارگاه‌های آموزشی و مطالعه در این زمینه دارند و یا این که اگر اطلاعات کافی در این زمینه داشته باشند، وقت مناسبی که با آرامش ذهنی به طراحی سؤال بپردازند، در اختیار ندارند. دیده می‌شود برخی از اساتید گاهی در راهروهای بیمارستان ودر زمان درگیری با مسائل بیماران‌شان، اقدام به طرح و نوشتن سؤالات می‌نمایند که این چنین طراحی سؤال، تنها برای رفع مسؤولیت صورت می‌گیرد. از طرف دیگر تایپ و آماده‌سازی سؤالات اغلب زیر نظر افراد غیر پزشک انجام می‌شود که حتی توانایی تشخیص اشکالات واضح در سؤالات را هم ندارند. پس از اجرای آزمون‌ها نیز اغلب نه استاد و نه دانشجویان، بازخوردی دربارهٔ سؤالات مورد اشکال و اعتراض، دریافت نمی‌کنند و به این ترتیب گاهی این گونه سؤالات حتی در چندین امتحان تکرار می‌شود. واحد امتحانات نیز پس از هر امتحان حجم بالایی از اعداد مربوط به ارزشیابی سؤالات را در اختیار واحدهای برگزارکننده، قرار می‌دهد. اما عملاً تنها فایده این اعداد، پرکردن بایگانی‌ها است و هیچ استفاده‌ای در جهت ارتقای کیفیت و پایایی آزمون‌ها از آن نمی‌شود. هدف از انجام این مطالعه بررسی کیفیت آزمون‌های چهارگزینه‌ای برگزار شده در دوره مقدمات پزشکی بالینی (بررسی بانک سؤال این دوره) می‌باشد.

روش‌ها

در یک مطالعه توصیفی آزمون‌های پاتوفیزیولوژی بیماری‌ها که از سال ١٣٨٤ تا ١٣٨٨ در دوره مقدمات پزشکی بالینی(ICM) برگزار شده بود، مورد بررسی قرار گرفت. تعداد ٨ آزمون از هریک از مباحث هفتگانه پاتوفیزیولوژی (قلب، کلیه، تنفس، خون، غدد، گوارش، روماتولوژی) و یا به عبارت دیگر در کل ٥٦ آزمون مورد بررسی قرار گرفت. تمامی این آزمون‌ها، آزمون‌های پایانی و به صورت چند گزینه ای MCQ بوده‌اند. در این ارزیابی کلیه سؤالات چهارگزینه‌ای از نظر ضریب دشواری، ضریب تمیز، ١٠ مورد مربوط به آنالیز کیفی و سطح شناختی مورد بررسی قرار گرفتند. برای بررسی کیفیت، شاخص‌های زیر اندازه‌گیری شده است: (١) فراوانی و نوع نقص‌های موجود در طراحی سؤال (٢) فراوانی سؤالات با ضریب تمیز مناسب (٣) فراوانی سؤالات با ضریب دشواری مناسب (٤) سطح شناختی سؤالات (٥) ارتباط بین سطح شناختی سؤالات و وجود نقص در سؤال (٤) ارتباط بین سطح شناختی سؤالات و ضریب تمییز و دشواری سؤالات. از میان ٣٩٧٣ سؤال، ٦٤ سؤال تکراری (سؤالاتی که دقیقا ساقه و گزینه‌ها به جز تغییرات کوچک تایپی، شبیه به هم بودند) بود. از آنجا که اولاً نتایج آنالیز کمی یک سؤال در آزمونهای مختلف می‌تواند متفاوت باشد و دوماً

[Downloaded from ijme.mui.ac.ir on 2026-11-09]

تغییرات کوچک در بازنویسی سؤال می‌تواند در نتایج آنالیز کیفی مؤثر باشد، و سوماً هدف این مطالعه بررسی کیفیت سؤالات مورد استفاده در آزمون‌ها در یک دوره زمانی بوده است، لذا سؤالات صرف نظر از تکراری بودن یا نبودن مورد بررسی قرار گرفت.

در آنالیز کمی سؤالات، ضریب دشواری و تمییز هر سؤال بررسی گردید. این ضرایب توسط واحد امتحانات دانشکده پزشکی همزمان با تصحیح پاسخ نامه‌ها، توسط نرم‌افزار محاسبه و به دفتر ICM ارسال می‌گردد. ضریب دشواری (P) بین ۰/۳ تا ۰/۷ مناسب، بیشتر از ۰/۷آسان و کمتر از ۰/۳ دشوار در نظر گرفته شده است. ضریب تمیز بالاتر از ۰/۳ مناسب، صفر تا ۰/۳ کم و زیر صفر نامناسب در نظر گرفته شده است. از آنجا که در آزمونهای پیشرفت تحصیلی سوالات بر اساس جدول دوبعدی آزمون طراحی می شوند، لذا میزان مطلوب از پیش تعیین شده ای نمی توان برای تعداد سوالات آسان و مناسب و دشوار در نظر گرفت. در این آزمونها میزان پایایی آزمون اهمیت دارد(۸)،

برای بررسی کیفی سؤالات از یک جدول با ۱۰ معیار مربوط به آنالیز کیفی سؤالات استفاده شد. طی چند سال آنالیز کمی و کیفی سؤالات در دفتر ICM، مشخص گردیده که از میان آیتم‌های موجود در راهنماهای طراحی سؤال(۹) ، ۱۰ مورد بیش از همه مورد نقض قرار می‌گیرد. بنابراین این ۱۰ مورد(جدول ۲) برای آنالیز کیفی سؤالات در این مطالعه مورد استفاده قرار گرفته است.

از طرف دیگر سؤالات براساس دو سطح شناختی طبقه بندی شدند. از آنجا که تاکسونومی بلوم برای طبقه‌بندی سؤالات در ارزیابی‌های پایانی کتبی به آسانی قابل استفاده است(۹)، از این تاکسنومی استفاده گردید. اگر چه تاکسونومی بلوم دارای ٦ سطح است (آگاهی، دریافت، کاربرد، تجزیه و تحلیل، ترکیب و ارزشیابی)، اما فقط ٤ سطح اول قابل ارزیابی در سؤالات ٤ گزینه‌ای می‌باشد. آگاهی و دریافت از سطوح پایین شناختی و کاربرد و تجزیه و تحلیل از سطوح بالای شناختی در سؤالات ٤ گزینه‌ای می‌باشد(۹). بنابراین در این مطالعه برای اینکه پایایی بین نمره دهندگان افزایش پیدا کند، طبقه بندی بلوم ساده شده و سؤالات چهارگزینه‌ای به دو گروه T_1 و T_2 تقسیم شدند(۱۰). یک سؤال با تاکسونومی T_1 فقط یادآوری واقعیت‌ها یا دریافت مسائل پایه‌ای را سنجیده است. یک سؤال با تاکسونومی T_2 کاربرد دانش و تجزیه و تحلیل را ارزیابی نموده است. از آنجا که این آزمونها آزمونهای پیشرفت تحصیلی هستند، انتظار میرود که حداقل ۱۵-۱۰ درصد سوالات در گروه T_2 قرار داشته باشند(۲).

فرآیند ارزیابی سؤالات به این‌گونه بوده است که با بررسی بانک سؤال دفتر دوره مقدمات پزشکی بالینی توسط دو پژوهشگر (دو پزشک کارشناس ارشد آموزش پزشکی) داده‌های مورد نظر استخراج گردید. از آنجا که این دو نفر تجربه زیادی در ارتباط با آزمونهای دوره مقدمات پزشکی بالینی داشته و در ارتباط مستقیم با اساتید طراح سوال بوده و واکاوی محتوا و ساختار سوالات این دوره برعهده ایشان میباشد، میتوان از تبحر در زمینه محتوای سوالات وتبحر در ساختن سوالات چهارگزینه ای اطمینان داشت. لازم به ذکر است که بانک سؤال دوره مقدمات پزشکی بالینی توسط دو پزشک کارشناس ارشد آموزش پزشکی مدیریت می‌شود. در این بانک سؤال، ضریب تمییز و دشواری سؤال، درصد انتخاب گزینه‌ها و رعایت یا عدم رعایت ۱۰ مورد مشخصات کیفی سؤال که در این مطالعه به آن پرداخته شده، وجود دارد. از آنجا که سطح شناختی سؤال در بانک سؤال مشخص نشده بود، در یک بررسی اولیه، هر پژوهشگر درباره سطح شناختی ۲۰۰ سؤال قضاوت نمود. سپس در یک جلسه، با بررسی قضاوت‌های انجام شده، دو پژوهشگر در قضاوت درباره سطح شناختی سؤال، به توافق رسیدند.

[Downloaded from ijme.mui.ac.ir on 2026-11-09]

برای آنالیز داده‌ها از آمار توصیفی استفاده شد. همچنین به منظور تعیین ارتباط میان متغیرهایی مانند تعداد نقص، سطح شناختی سؤال و قدرت تمییز از chi-square استفاده شد. داده‌ها با نرم‌افزار SPSS-11.5 آنالیز شد.

نتایج

با بررسی ۳۹۷۳ سؤال چهار گزینه‌ای مشخص گردید که ۲۶۵۸ (۶۶/۹ درصد) سؤال حداقل دارای یک نقص و ۴۲/۳ درصد سؤالات بیش از یک نقص در ساختار خود دارند (جدول۱). در کل ۴۸۹۳ نقص در طراحی ۳۹۷۳ سؤال یافت شد(برخی از سوالات هیچ نقصی نداشته و برخی چند نقص داشتند). بیشترین نقص‌های موجود، نبودن بخش اعظم اطلاعات در ساقه سؤال (۴۳/۲ درصد) و کاربرد لغات منفی در ساقه سؤال (۳۴/۸ درصد) بودند(جدول ۲).

جدول ۱: فراوانی وجود نقص در ساختار سؤالات

تعداد نقص	n (٪) N=۳۹۷۳
بدون نقص	۱۳۱۵ (۳۳/۱)
۱ نقص	۹۷۸ (۲۴/۶)
۲ نقص	۱۲۰۴ (۳۰/۳)
۳ نقص	۳۹۷ (۱۰)
۴ نقص	۷۹ (۲)

جدول ۲: فراوانی هر یک از ۱۰ نقص موجود در سؤالات چهارگزینه‌ای

نوع نقص	n (٪) N=۴۸۹۳
عدم وجود بخش اعظم اطلاعات در ساقه سؤال	۲۱۱۴(۴۳/۲)
وجود گزینه منفی برای ساقه منفی	۳۰۸(۶/۳)
وجود لغات منفی در ساقه سؤال	۱۷۰۳(۳۴/۸)
وجود لغات منفی در گزینه‌ها	۱۳۲(۲/۷)
وجود گزینه‌های همه موارد، هیچ کدام، گزینه‌های ترکیبی	۱۸۶(۳/۸)
وجود گزینه‌های متضاد	۱۰ (۰/۲)
عدم همبستگی گزینه‌ها از نظر طول و ساختاربندی	۲۴(۰/۵)
کاربرد عبارات تکراری در گزینه‌ها	۱۵۷(۳/۲)
عدم مطابقت سؤال با اهداف دوره	۸۳(۱/۷)
عدم وجود فقط یک گزینه صحیح	۱۷۶(۳/۶)

از میان ۳۹۷۳ سؤال مورد بررسی سؤالات آسان (۴۲/۲درصد)، ۲۸/۸٪ متوسط و ۲۹٪ دشوار بودند. سؤالات با ضریب تمیز کم (۴۷/۹ درصد) بیشترین درصد را به خود اختصاص داده بودند. ۴۷/۳٪ و ۴/۸٪ سوالات بترتیب ضریب تمیز مناسب و نامناسب بودند.

در بررسی تاکسونومی سؤالات بیشتر سؤالات (۷۹/۱ درصد n= ۳۱۴۲) در سطح T_1 (یادآوری و دریافت مسائل پایه‌ای) و بقیه (۲۰/۹ درصد n=۸۳۱) در سطح T2 ارزیابی شدند. ۷۷/۹ درصد (n=۲۴۴۷) از سؤالات با سطح T1 در مقایسه با ۲۵/۴ درصد (n= ۲۱۱) از سؤالات با سطح T2، دارای نقص بودند (P<۰/۰۰۱) (جدول ۳).

جدول ۳: ارتباط سطح شناختی سؤال با وجود نقص در ساختار سؤال

سطح شناختی	وجود نقص در ساختار سؤال		Sig chi-square
	بلی	خیر	
T_1	۲۴۴۷(٪۷۷/۹)	۶۹۵(٪۲۲/۱)	P<۰/۰۰۱
T_2	۲۱۱(٪۲۵/۴)	۶۲۰(٪۷۴/۶)	

اگرچه ارتباط معناداری بین دشواری مناسب سؤال و سطح شناختی وجود نداشت، اما سؤالات با سطح شناختی T2 به طور معناداری دارای درصد بالاتری از سؤالات با ضریب تمیز مناسب بودند.

[Downloaded from ijme.mui.ac.ir on 2026-11-09]

جدول ۴: ارتباط بین سطح شناختی و ضریب دشواری و قدرت تمیز سؤالات

سطح شناختی	ضریب دشواری سؤالات			ضریب تمییز سؤالات		
	آسان	دشوار	مناسب	نامناسب	کم	مناسب
T_1 (n= 3142)	۱٤٤۷	۸۸٤	۸۱۱	۱۲۳	۱٦۷٤	۱۳٤٥
T_2 (n= 831)	۲۳۱	۲٦۷	۳۳۳	٦۸	۲۲۹	٥۳٤
Sig chi-square	P=۰/۲٤۱			P<۰/۰۰۱		

بحث

این مطالعه با هدف بررسی کیفیت آزمون‌های چهارگزینه‌ای برگزار شده در دوره مقدمات پزشکی بالینی (بررسی بانک سؤال این دوره) صورت پذیرفته است. فراوانی وجود نقص در طراحی سؤالات در این مطالعه (٦٦/۹ درصد) بیشتر از مطالعه مشابهی است که تارانت (Tarrant) در سال ۲۰۰٦ (٤٦/۲ درصد) بر روی سؤالات پرستاری انجام داده است(۱۰). دونینگ (Downing) نیز در ۲۰۰٦ نتایج مشابهی با مطالعه ترنت در مورد سؤالات دانشکده پزشکی به دست آورده است(۱۱). اما نتایج این مطالعه با میزان نقص‌های موجود در بانک سؤال روانشناسی (٦۰ درصد) که در مطالعه الورث (Ellsworth) به آن پرداخته شده است نزدیک می‌باشد(۱۲). در مطالعه حاضر، از ۱۰ آیتم بررسی شده، ۲ آیتم با فراوانی زیادی در سؤالات نقض شده بود. از این دو آیتم، وجود لغات منفی در ساقه سؤال دومین فراوانی را دارد. در مطالعه تارانت نیز این آیتم دومین فراوانی را دارد. به نظر تارانت سؤالاتی که به صورت منفی نوشته می‌شوند، راحت‌تر و سریع‌تر ساخته می‌شود و بنابراین استفاده از آنها رایج است(۱۰). این سؤالات باعث گیج شدن دانشجو می‌شوند، به خصوص زمانی که به صورت منفی در منفی ساخته می‌شوند(۱۱). عدم وجود بخش اعظم اطلاعات در ساقه سؤال، در مورد سؤالاتی است که به صورت «کدام مورد صحیح است» و یا «کدام مورد غلط است» طراحی می‌شوند. این سؤالات نیز خیلی ساده و راحت طراحی می‌شود و به همین دلیل در مطالعه ما، بیشترین فراوانی را در میان نقص‌ها دارا می‌باشد.

در این مطالعه اگرچه سؤالات با ضریب دشواری مناسب زیاد نبودند (۲۸/۸ درصد) اما به نظر می‌رسد که فراوانی سؤالات با ضریب تمییز مناسب در حد خوبی باشد(٤۷/۳ درصد). فراوانی سؤالات با سطح شناختی T2 در این مطالعه در سطح مطلوبی بوده است. این فراوانی (۲٥/٤ درصد) نسبت به مطالعه تارانت (۱۰ درصد) بالاتر و تقریباً مشابه با مطاله ماستر (۲۸/۳ درصد) است(۱۰و۱۳). جالب است که در مطاله ما، سؤالات با سطح شناختی T2 نسبت به T1 به طور معناداری دارای قدرت تمییز بالاتری بوده‌اند. همچنین سولات با سطح شناختی T2 به طور معناداری نقص‌های ساختاری کمتری داشته‌اند. تارانت نیز در مطالعه خود چنین ارتباطی را بین سطح شناختی و نقص‌های ساختاری سؤالات اذعان نموده است(۱۰). شاید بتوان نتیجه گرفت که طراحی سؤال در سطوح شناختی بالاتر باعث کاهش نقص‌های ساختاری آن می‌گردد.

استفاده روز افزون از سؤالات چند گزینه ای در ارزشیابی‌ها ضرورت آنالیز دقیق سؤالات توسط اساتید قبل و بعد از آزمون را بیشتر می‌کند(٤). برای این کار لازم است بر تهیه بانک سؤال و آنالیز سؤالات موجود در بانک سؤالات دانشکده‌های پزشکی تأکید گردد. ماستر (Master) و همکاران در مطالعه خود نشان دادند که حتی بیشتر سؤالات انتخاب شده از بانک سؤال دانشکده پرستاری لانزینگ (Lansing)، دارای اشکالات جزیی و برخی از آنها دارای اشکالات اساسی بوده است. آنها اساتید را تشویق می‌کنند که قبل از استفاده از سؤالات بانک سؤال، بازهم سؤال مورد نظر را به دقت مورد بررسی قرار

[Downloaded from ijme.mui.ac.ir on 2026-11-09]

دهند(۱۳). چرا که item-analysis اطلاعات مفیدی را برای بهبود کیفیت و صحت سؤالات چندگزینه‌ای در اختیار آنان قرار می‌دهد(۱٤تا ۱٦).

نتیجه‌گیری

بنابراین با توجه به این مطالعه و سایر مطالعات مشابه می‌توان نتیجه گرفت که طراحی سؤال در سطوح شناختی بالاتر باعث کاهش نقص‌های ساختاری آن می‌گردد. با توجه به اهمیت ارزشیابی در تعیین میزان دستیابی به اهداف و مهارت‌های مورد انتظار و نقش نمره به عنوان پاداش عملکرد دانشجو در ایجاد انگیزه برای پرداختن به فعالیت‌های یادگیری، بر ضرورت اصلاح و بازنگری در روش‌های ارزشیابی دانشجویان تأکید می‌گردد(۲و۳). در نهایت پیشنهاد می‌شود که مطالعات مشابهی بر روی سایر آزمون‌های دانشکده پزشکی انجام گرفته و اساتید طراح سؤال از نتایج آن آگاه گردند.

منابع

1. McManus IC, Mooney-Somers J, Dacre JE, Vale JA. Reliability of the MRCP(UK) Part I Examination, 1984-2001. Med Educ 2003; 37(7): 609-11.
2. Srinivasa DK, Adkoli BV. Multiple choice questions: how to construct and how to evaluate. Indian J Pediatr 1989; 56(1): 69-74.
3. Colbert MA. Statistical Analysis of Multiple Choice Testing. Available at: http://www.stormingmedia.us/64/6447/A644704.html. Accessed May 15, 2006.
4. Vahalia KV, Subramaniam K, Marks SC Jr, De Souza EJ. The use of multiple-choice tests in anatomy: common pitfalls and how to avoid them. Clin Anat 1995; 8(1): 61-5.
5. Owen A, Byrne G, O'Neill P. FDTL4: A core question bank linked to learning outcomes for undergraduate medicine. Available at: http://www.ltsn-01.ac.uk/newsletter/01.3.html#corequestionbank. Accessed May 10, 2006.
6. Terry MA. Writing a multiple-choice test question. J Am Osteopath Assoc 1992; 92(1): 112-4, 123.
7. Fleming PR. The administration of a multiple choice question bank. Med Educ 1984; 18(5): 372-6.
8. zolfaghari B, Adibi N, Derakhshanfar S, Tansaz M, Karbasi A, niroomand P. Academic achivement tests in medical sciences.Isfahan.Isfahan University of medical sciences- Modiriate motaleat va toseae amoozeshe pezeshki 2000.
9. Haladyna T M, Downing S M, Rodrigues M C. A review of multiple choise item writing guidelines for classroom assessment. Applied Measurement in Education 2002; 15(3):309-334.
10. Tarrant M, Knierim A, Hayes SK, Ware J. The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. Nurse Educ Pract. 2006 Dec;6(6):354-63.
11. Downing SM. The effects of violating standard item writing principles on tests and students: the consequences of using flawed test items on achievement examinations in medical education. Adv Health Sci Educ Theory Pract. 2005;10(2):133-43.
12.Ellsworth R A, Dunnell P, Duell O K. Multiple chiose test items: what are textbook authors telling teachers? Journal of educational research 1990; 83(5): 289-293.
13. 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 Jan;40(1):25-32.
14. Davis P J. An Essay on the Statistical Analysis of Exam Questions. Available at: http://www.utexas.edu/pharmacy/resources/guide/stats.pdf. Accessed April 11, 2011.Statistical Analysis of Exam Questions
15. Moussa MA, Ouda BA, Nemeth A. Analysis of multiple-choice items. Comput Methods Programs Biomed. 1991; 34(4): 283-9.
16. Bandaranayake RC, Buzzard AJ, Marshall VC. Effects of publishing multiple-choice questions on their

[Downloaded from ijme.mui.ac.ir on 2026-11-09]

subsequent item analysis. Aust N Z J Surg 1990; 60(12): 937-41.

[Downloaded from ijme.mui.ac.ir on 2026-11-09]

Reviewing the results of qualitative and quantitative analysis of MCQs in Introduction to clinical medicine course

Vajiheh Vafamehr[1], Mohammad Dadgostarnia[2] *

Abstract

Background: *At most of the medical universities, MCQ-based examinations are often use as the first component, particularly to ensure the candidates have an adequate knowledge base, prior to entering subsequent clinical examinations.*
Method: *In ICM , 3973 MCQs were collected from tests over a four-year period from 2005 to2009. Questions were evaluated for 10 frequently occurring item writing flows, for cognitive level and for difficulty and discrimination index.*
Results*: Result show that 66.9% of the questions contained violations of item writing guidelines and79.1% were written at low cognitive levels. MCQs written at a low cognitive level were significantly more likely to contain item writing flows.47.3% of the questions had suitable discrimination index and 28.8% of the questions had suitable difficulty index. While there was no relationship between difficulty index and cognitive level, higher cognitive level and suitable discrimination index had significant relationship.*
Discussion: *We should think about practically ways for using results of item analysis. Faculty members should be taught designing and using suitable questions and developing MCQs bank.*

Key Words: *Item Analysis, Question Bank, MCQ*

Address
[1]Master of Medical Education, evaluation and tests expert, school of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. E mail: vgrenani@yahoo.com
[2](✉)Master of Medical Education, in charge of ICM, school of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. E mail: dadgostar@med.mui.ac.ir

[Downloaded from ijme.mui.ac.ir on 2026-11-09]

This document was created with Win2PDF available at http://www.daneprairie.com.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.

[Downloaded from ijme.mui.ac.ir on 2026-11-09]