

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

وجیهه و فامهر، محمد دادگسترنسی\*

### چکیده

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

**روش‌ها:** در دوره مقدمات پزشکی بالینی، ۳۹۷۳ سؤال 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 تمیز مورد بررسی قرار گرفتند.

**نتایج:** نتایج گویای آن بود که ۶۶/۹ درصد از سؤالات دارای حداقل یک خطا در ساختار خود بودند. ۱/۷۹ درصد سؤالات در سطوح پایین شناختی طراحی شده بودند. سؤالاتی که دارای سطح پایین شناختی بودند به طور معناداری دارای خطاهاست و ساختاری بیشتری بودند. ۳/۴۷ درصد سؤالات دارای ضریب تمیز مناسب و ۸/۲۸ 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 تمیز مناسب رابطه معناداری داشتند.

**نتیجه‌گیری:** برای بهبود کیفیت آزمون‌ها و استفاده از نتایج تجزیه و تحلیل کمی و کیفی سؤالات می‌بایست راه کارهای 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) : ۱۱۴۶-۱۱۵۲

### مقدمه

می‌گیرند(۱-۴). 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 فرآگیران، آزمون‌های چند گزینه‌ای است. سؤالات چندگزینه‌ای خوب طراحی شده، فراتر از یک یادآوری ساده اطلاعات هستند. چنین سؤالاتی آزمون‌شونده را مجبور می‌کند که برای یک مشکل کلینیکی، به تحلیل و نتیجه‌گیری از مطالب تئوری ذهن خود بپردازد(۵). این سؤالات می‌توانند برای سنجش اغلب سطوح یادگیری به کار بردند. همچنین توانایی اندازه‌گیری بیشتر اهداف آموزشی را دارا

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

**نویسنده مسؤول:** دکتر محمد دادگسترنسی، کارشناس ارشد آموزش پزشکی و مسؤول دوره مقدمات پزشکی بالینی دانشکده پزشکی، دانشگاه علوم پزشکی، اصفهان، ایران.

(dadgostar@med.mui.ac.ir)

دکتر وجیهه و فامهر، 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.

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 ۱۳۸۴ تا ۱۳۸۸ در دوره مقدمات پزشکی بالینی (ICM) برگزار شده بود، مورد بررسی قرار گرفت. تعداد ۸ آزمون از هریک از مباحث هفتگانه پاتوفیزیولوژی (قلب، کلیه، تنفس، خون، غدد، گوارش، روماتولوژی) و یا به عبارت دیگر در کل ۵۶ آزمون مورد بررسی قرار گرفت. تمامی این آزمون‌ها، آزمون‌های پایانی و به صورت چند گزینه‌ای MCQ بوده‌اند. در این ارزیابی کلیه سؤالات چهارگزینه‌ای از نظر ضریب دشواری، ضریب تمیز، ۱۰ مورد مربوط به آنالیز کیفی و سطح شناختی مورد بررسی قرار گرفتند. برای بررسی کیفیت، نقص‌های زیر اندازه‌گیری شده است: (۱) فراوانی و نوع شاخص‌های زیر اندازه‌گیری شده است: (۲) فراوانی سؤالات با ضریب تمیز مناسب (۳) فراوانی سؤالات با ضریب دشواری مناسب (۴) سطح شناختی سؤالات (۵) ارتباط بین سطح شناختی سؤالات و وجود نقص در سؤال (۶) ارتباط بین سطح شناختی سؤالات و ضریب تمیز و دشواری سؤالات. از میان ۳۹۷۳ سؤال، ۶۴ سؤال تکراری (سؤالاتی که دقیقاً ساخته 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

می‌باشد(۶). از طرف دیگر آزمون‌های چهارگزینه‌ای توانایی ارزیابی تعداد زیادی دانشجو با بکارگیری حداقل نیروی انسانی را دارند (۴). 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(۷).

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 عل 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 و در 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، بازخورده درباره سؤالات مورد اشکال و اعتراض، دریافت نمی‌کنند و به این ترتیب گاهی

آگاهی و دریافت از سطوح پایین شناختی و کاربرد و تجزیه و تحلیل از سطوح بالای شناختی در سوالات ۴ گزینه‌ای می‌باشد<sup>(۹)</sup>. بنابراین در این مطالعه برای اینکه پایایی بین نمره دهندگان افزایش پیدا کند، طبفه بندی بلوم ساده شده و سوالات چهارگزینه‌ای به دو گروه  $T_1$  و  $T_2$  تقسیم شدند<sup>(۱۰)</sup>. یک سوال با تاکسونومی  $T_1$  فقط یادآوری واقعیت‌ها یا دریافت مسائل پایه‌ای را سنجیده است. یک سوال 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 ۱۰-۱۵ درصد سوالات در گروه  $T_2$  قرار داشته باشند<sup>(۲)</sup>.

فرآیند ارزیابی سوالات به این‌گونه بوده است که با بررسی بانک سوال 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 استادی طراح سوال بوده و اکاواه محتوا و ساختار سوالات این دوره بر عهده ایشان می‌باشد، می‌توان از تبحر در زمینه محتوای سوالات و تبحر در ساختن سوالات چهارگزینه ای اطمینان داشت. لازم به ذکر است که بانک سوال دوره مقدمات پزشکی بالینی توسط دو پزشک کارشناس ارشد آموزش پزشکی مدیریت می‌شود. در این بانک سوال، ضریب تمیز و دشواری سوال، درصد انتخاب گزینه‌ها و رعایت یا عدم رعایت ۱۰ مورد مشخصات کیفی سوال که در این مطالعه به آن پرداخته شده، وجود دارد. از آنجا که سطح شناختی سوال در بانک سوال مشخص نشده بود، در یک بررسی اولیه، هر پژوهشگر درباره سطح شناختی ۲۰۰ سوال قضاؤت نمود. سپس در یک جلسه، با بررسی قضاؤت‌های انجام شده، دو پژوهشگر در قضاؤت درباره سطح شناختی سوال، به توافق رسیدند.

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

در آنالیز کمی سوالات، ضریب دشواری و تمیز هر سوال 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 ۰/۰ مناسب، بیشتر از ۰/۷ آسان و کمتر از ۰/۳ دشوار در نظر گرفته شده است. ضریب تمیز بالاتر از ۰/۰ مناسب، صفر تا ۰/۲ کم و زیر صفر نامناسب در نظر گرفته شده است. از آنجا که در آزمونهای پیشرفت تحصیلی سوالات بر اساس جدول دو بعدی آزمون طراحی می‌شوند، لذا میزان مطلوب از پیش تعیین شده ای نمی‌توان برای تعداد سوالات آسان و مناسب و دشوار در نظر گرفت. در این آزمونها میزان پایایی آزمون اهمیت دارد<sup>(۸)</sup>، برای بررسی کیفی سوالات از یک جدول با ۱۰ معیار مربوط به آنالیز کیفی سوالات استفاده شد. طی چند سال آنالیز کمی و کیفی سوالات در دفتر ICM مشخص گردیده که از میان آیتم‌های موجود در راهنمای طراحی سوال<sup>(۹)</sup>، ۱۰ مورد بیش از همه مورد نقض قرار می‌گیرد. بنابراین این ۱۰ مورد(جدول ۲) برای آنالیز کیفی سوالات در این مطالعه مورد استفاده قرار گرفته است.

از طرف دیگر سوالات براساس دو سطح شناختی طبقه بندی شدند. از آنجا که تاکسونومی بلوم برای طبقه بندی سوالات در ارزیابی‌های پایانی کتبی به آسانی قابل استفاده است<sup>(۹)</sup>، از این تاکسونومی استفاده گردید. اگر چه تاکسونومی بلوم دارای ۶ سطح است (آگاهی، دریافت، کاربرد، تجزیه و تحلیل، ترکیب و ارزشیابی)، اما فقط ۴ سطح اول قابل ارزیابی در سوالات ۴ گزینه‌ای می‌باشد.

۱۸۶(۳/۸)	وجود گزینه‌های همه موارد، هیچ کدام، گزینه‌های ترکیبی
۱۰(۰/۲)	وجود گزینه‌های متضاد
۲۴(۰/۵)	عدم همبستگی گزینه‌ها از نظر طول و ساختاربندی
۱۵۷(۳/۲)	کاربرد عبارات تکراری در گزینه‌ها
۸۳(۱/۷)	عدم مطابقت سؤال با اهداف دوره
۱۷۶(۳/۶)	عدم وجود فقط یک گزینه صحیح

از میان ۳۹۷۳ سؤال مورد بررسی سؤالات آسان (۴۲/۲ درصد)، 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٪ سؤالات بترتیب ضریب تمیز مناسب و نامناسب بودند.

در بررسی تاکسونومی سؤالات بیشتر سؤالات (۷۹/۱ درصد ۳۱۴۲ (n=۳۱۴۲) در سطح T<sub>1</sub> (یادآوری و دریافت مسائل پایه‌ای) و بقیه (۹/۰/۹ درصد (n=۸۳۱) در سطح T<sub>2</sub> ارزیابی شدند. ۷۷/۹ درصد (n=۲۴۴۷) از سؤالات با سطح T<sub>1</sub> در مقایسه با ۲۵/۴ درصد (n=۲۱۱) از سؤالات با سطح T<sub>2</sub> دارای نقص بودند (P<0.001) (جدول ۳).

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

Sig chi-square	وجود نقص در ساختار سؤال		نوع نقص
	بلی	خیر	
P<0.001	۶۹۵(۰/۲۲/۱)	۲۴۴۷(۰/۷۷/۹)	T <sub>1</sub>
	۶۲۰(۰/۷۴/۶)	۲۱۱(۰/۲۵/۴)	T <sub>2</sub>

اگرچه ارتباط معناداری بین دشواری مناسب سؤال و سطح شناختی وجود نداشت، اما سؤالات با سطح شناختی T<sub>2</sub> به طور معناداری دارای درصد بالاتری از سؤالات با ضریب تمیز مناسب بودند.

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

## نتایج

با بررسی ۳۹۷۳ سؤال چهار گزینه‌ای مشخص گردید که ۶۶/۹ درصد) سؤال حداقل دارای یک نقص و ۴۲/۳ درصد سؤالات بیش از یک نقص در ساختار خود دارند (جدول ۱). در کل ۴۸۹۳ نقص در طراحی ۳۹۷۳ سؤال یافت شد (برخی از سؤالات 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

نوع نقص	n	(%)	N=۳۹۷۳
بدون نقص	۱۲۱۵	(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)	(۴۳/۲)	
وجود گزینه منفی برای ساقه منفی	۲۰۸(۰/۶)	(۰/۶)	
وجود لغات منفی در ساقه سؤال	۱۷۰۳(۳۴/۸)	(۳۴/۸)	
وجود لغات منفی در گزینه‌ها	۱۲۲(۰/۷)	(۰/۷)	

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 سوالات با ضریب دشواری مناسب زیاد نبودند (۲۸/۸ درصد) اما به نظر می‌رسد که فراوانی سوالات با ضریب تمییز مناسب در حد خوبی باشد (۴۷/۳ درصد). فراوانی سوالات با سطح شناختی T2 در این مطالعه در سطح مطلوبی بوده است. این فراوانی (۲۵/۴ درصد) نسبت به مطالعه تارانت (۱۰ درصد) بالاتر و تقریباً مشابه با مطاله ماستر (۲۸/۳ درصد) است (۱۰ و ۱۳). جالب است که در مطاله ما، سوالات با سطح شناختی T2 نسبت به طور معناداری دارای قدرت تمییز بالاتری بوده‌اند. همچنین سوالات 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 سوالات اذعان نموده است (۱۰). شاید بتوان نتیجه گرفت که طراحی سوال در سطوح شناختی بالاتر باعث کاهش نقص‌های ساختاری آن می‌گردد.

استفاده روز افزون از سؤالات چند گزینه ای در ارزشیابی ها ضرورت آنالیز دقیق سؤالات توسط اساتید قبل و بعد از آزمون را بیشتر می کند (۴). برای این کار لازم است بر تهیه بانک سؤال و آنالیز سؤالات موجود در بانک سؤالات دانشکده های پژوهشی 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 جزئی و برخی از آنها دارای اشکالات اساسی بوده است. آنها اساتید را تشویق می کنند که قبل از استفاده از سؤالات بانک سؤال، باند هم سؤال موعد نظر دا به دقت مورد بررسی قرار

## جدول ٤: ارتباط بین سطح شناختی و ضریب دشواری و قدرت تمیز سؤالات

ضریب تمییز سؤالات		ضریب دشواری سؤالات		سطح شناختی	
مناسب	نامناسب	مناسب	دشوار	آسان	
۱۲۴۵	۱۶۷۴	۱۲۳	۸۱۱	۸۸۴	۱۴۴۷
					$T_1$ (n= 3142 )
۵۳۴	۲۲۹	۶۸	۳۳۳	۲۶۷	۲۳۱
					(n= 831 ) $T_2$
$P<./.001$		$P=.0241$		Sig chi-square	

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 (۶۶/۹ درصد) بیشتر از مطالعه مشابهی است که تارانت (Tarrant) در سال ۲۰۰۶ (۴۶/۲ درصد) بر روی سؤالات پرستاری انجام داده است (۱۰). دونینگ (Downing) نیز در ۲۰۰۶ نتایج مشابهی با مطالعه ترن特 در مورد سؤالات دانشکده پزشکی به دست آورده است (۱۱). اما نتایج این مطالعه با میزان نقص‌های موجود در بانک سؤال روانشناسی (۶۰ 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

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(۲ و ۳). در نهایت پیشنهاد می‌شود که مطالعات مشابهی بر روی سایر آزمون‌های دانشکده پزشکی انجام گرفته و استادی طراح سؤال از ترتیب آن آگاه گرددند.

دهند(۱۲). چرا که item-analysis اطلاعات مفیدی را برای بهبود کیفیت و صحت سؤالات چندگزینه‌ای در اختیار آنان قرار می‌دهد(۱۴ تا ۱۶).

### 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.

### 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 achievement tests in medical sciences. *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 choice 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 choice 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

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

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

Vajiheh Vafamehr<sup>1</sup>, Mohammad Dadgostarnia<sup>2</sup> \*

## 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 to 2009. 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 and 79.1% 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

<sup>1</sup>Master of Medical Education, evaluation and tests expert, school of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. E mail: vgrenani@yahoo.com

<sup>2</sup>(✉)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

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.