

تجزیه و تحلیل کیفیت سؤالات آزمونهای چندگزینه‌ای رشته کارشناسی ناپیوسته مدارک پزشکی

سعید حسینی تشنیزی^۱ دکتر شهرام زارع^۲ دکتر سیدمهداد صولتی^۳

^۱ مربی گروه آمار حیاتی، ^۲ دانشیار گروه پزشکی اجتماعی، ^۳ استادیار گروه داخلی، دانشگاه علوم پزشکی هرمزگان

مجله پزشکی هرمزگان سال چهاردهم شماره سوم پاییز ۸۹ صفحات ۱۸۳-۱۷۷

چکیده

مقدمه: روشهای مختلفی برای ارزشیابی علمی دانشجویان وجود دارد. یکی از رایج‌ترین آنها طراحی سؤالات چندگزینه‌ای می‌باشد که چنانچه به درستی طراحی شوند روش مناسبی برای سنجش آموخته‌های دانشجویان می‌باشد. با توجه به گسترش استفاده از آزمونهای چندگزینه‌ای، این مطالعه بمنظور بررسی کیفیت آزمونهای چندگزینه‌ای رشته مدارک پزشکی دانشگاه علوم پزشکی هرمزگان اجرا شده است.

روش کار: در یک مطالعه توصیفی تحلیلی تمامی آزمونهای سؤالات چندگزینه‌ای (MCQ) پایان ترم مربوط به دوره کارشناسی رشته تحصیلی رشته مدارک پزشکی دانشگاه علوم پزشکی هرمزگان شامل کاربرد رایانه، زبان انگلیسی، سیستمهای رایج طبقه‌بندی، اصطلاحات پزشکی^۳، روش تحقیق و شاخصها و تحلیل داده‌های بهداشتی انتخاب شدند. سپس شاخص دشواری و تمیز سؤالات با استفاده از نرم‌افزارهای Excel2007 و SPSS16 و با کمک روشهای آمار توصیفی و آزمونهای آنالیز واریانس و آنالیز رگرسیون تجزیه و تحلیل شدند.

نتایج: نتایج نشان داد از ۱۲۸ سؤال، ۱۰۵ سؤال (۵۷/۷٪) دارای ضریب دشواری مناسب که کمترین آن مربوط به روش تحقیق با ۰/۵۹ و بیشترین آن مربوط به اصطلاحات پزشکی ۳ با ۰/۷۸ است. قدرت تمیز ۱۳۳ سؤال (۷۳/۱٪) در حد ضعیف تا خیلی ضعیف بطوری که زبان انگلیسی (۰/۰۵-) و سیستمهای رایج طبقه‌بندی (۰/۰۳-) ضعیف‌ترین قدرت تمیز و روش تحقیق (۰/۲۳) و کاربرد رایانه (۰/۱۸) بالاترین قدرت تمیز را داشتند. نتایج آزمون آماری نشان داد میزان سختی و قدرت تمایز در همه دروس با هم تفاوت معنی‌داری دارند ($P < 0/05$) همچنین بین ضریب دشواری و تمیز سؤالات رابطه‌ای مشاهده نشد.

نتیجه‌گیری: هر چند ضریب دشواری برای تمامی آزمونهای از میزان قابل قبول و مناسب برخوردار بود ولی نتایج شاخص تمیز نشان داد تقریباً برای تمام سؤالات آزمونها، کیفیت در حد ضعیف تا خیلی ضعیف می‌باشد. بنابراین بهتر است این سؤالات از بانک سؤالات حذف شوند یا بطور اساسی مورد تجدید نظر قرار گیرند.

کلیدواژه‌ها: مدارک پزشکی - دانشجویان - ارزشیابی

نویسنده مسئول:

سعید حسینی تشنیزی
دانشکده پرستاری و مامایی دانشگاه
علوم پزشکی هرمزگان
بندرعباس - ایران
تلفن: ۰۹۸۷۶۱۶۶۶۶۶۶
پست الکترونیکی:
shosseini@hums.ac.ir

دریافت مقاله: ۸۹/۷/۵ اصلاح نهایی: ۸۹/۷/۱۰ پذیرش مقاله: ۸۹/۷/۱۳

مقدمه:

میزان دقت و نارسائی آنها مورد بررسی قرار می‌گیرد تا نهایتاً نقاط قوت و ضعف و کیفیت آن آزمون تعیین گردد (۱). سؤوالهای چند گزینه‌ای (MCQ) از یک سؤوال یا مسئله به عنوان پایه یا بدنه سؤوال و تعداد چهار یا پنج پاسخ به عنوان گزینه‌ها که فقط یکی از آنها صحیح است، ساخته می‌شوند. پاسخ گزینه‌های صحیح به عنوان جوابهای درست و سایر پاسخها تحت عنوان گزینه‌های انحرافی معرفی می‌شوند (۲).

ارزشیابی از مهمترین بخشهای هر فرآیند آموزشی است. انجام صحیح آن می‌تواند باعث ایجاد انگیزه یادگیری در دانشجویان و بازخورد روند آموزش به مدرس شود. تحلیل آزمون و آمار مربوط به آزمون‌سازی یکی از مهمترین مراحل ارزشیابی آموزشی است و آن تک تک سؤوالها به منظور تعیین

تا ۰/۷ و ضریب تمیز بیشتر از ۰/۲ را برای آزمونهای معلم ساخته مناسب پیشنهاد می‌کنند (۱۰) ولی جانستون (۲۰۰۳) و مکالپین (۲۰۰۲) برای اساتید دانشگاه ضریب دشواری بین ۰/۳ تا ۰/۸ را بعنوان سطح دشواری و درجه تمیز بالایی ۰/۲ را قابل قبول و مناسب پیشنهاد نموده‌اند (۷). بطور کلی ضریب دشواری ۰/۳ تا ۰/۷ حداکثر اطلاع را درباره تفاوت بین آزمودنیها را بدست می‌دهد ولی در شرایط خاص، ضریب دشواری فرق می‌کند. مثلاً آزمونی برای کارشناسی ارشد یا دکتری برای انتخاب بهترین داوطلبان بکار می‌رود و حدوداً ۱۰ درصد بالایی داوطلبان را برمی‌گزیند، باید شامل سؤالهای بسیار دشوار باشد و برعکس آن آزمونی که برای انتخاب کودکان برای یک دوره آموزش اصلاحی باید شامل سؤالهای خیلی آسان باشد (۱).

یکی از مهمترین وظایف اعضای هیأت علمی، ارزیابی علمی دانشجویان در درسی که تدریس می‌کنند، می‌باشد. این وظیفه مهم نیاز به مهارت های قابل توجهی دارد، زیرا آزمون‌های دانشگاهی به عنوان یک شاخص کلیدی و حساس‌ترین نقطه فرآیند آموزش و یادگیری در تصمیم‌گیری برای ارتقاء یا رد دانشجویان به کار می‌روند (۱۱).

از آنجایی که سؤالات چندگزینه‌ای در ارزیابی سطح علمی دانشجویان دانشکده‌ها به کرات مورد استفاده قرار می‌گیرد، اطلاع از سطح کیفیت این سؤالات و ارائه بازخورد به اساتید و دانشجویان بسیار حائز اهمیت می‌باشد و به همین منظور این مطالعه در راستای بررسی کیفیت آزمونهای MCQs برگزار شده برای دانشجویان رشته مدارک پزشکی طراحی شده است.

روش کار:

در این مطالعه توصیفی تحلیلی کلیه دروسی آزمونهای MCQ که برای رشته کارشناسی ناپیوسته مدارک پزشکی دانشکده پیراپزشکی علوم پزشکی هرمزگان در نیمسال اول تحصیلی ۸۹-۸۸ برگزار شده بود و شامل ۶ آزمون کاربرد رایانه (۲۰ برگه)، زبان انگلیسی (۲۰ برگه)، سیستمهای رایج طب‌بندی (۲۰ برگه)، اصطلاحات پزشکی ۳ (۱۳ برگه)، روش تحقیق (۱۷ برگه) و شاخص‌ها و تحلیل داده‌های بهداشتی (۲۰

اینگونه سؤالات به کرات در آزمونهای پایان ترم دانشجویان دانشکده‌ها یا آزمون جامع دانشجویان پزشکی و همچنین ممکن است برای شناسایی نقاط قوت و ضعف دانش‌آموزان و ارائه بازخورد به معلمان جهت اقدامات آموزشی، مورد استفاده قرار گیرند (۳). سؤالاتی که در آزمونهای اکثر دانشکده‌های علوم پزشکی مورد استفاده قرار می‌گیرند عمدتاً به روش چند گزینه‌ای بوده و توسط متخصصین پس از بازنگری و بحث‌های موشکافانه طولانی انتخاب شده و بقیه آنها حذف می‌شوند (۴). نوشتن MCQ حتی برای افراد حرفه‌ای و آموزش دیده دشوار و وقت گیر است، بطوری که گفته می‌شود همین افراد حرفه‌ای برای نوشتن یک سؤال خوب بیش از یک ساعت وقت می‌گذارند (۵) و از آنجایی که هر روزه نیاز به مواد آموزشی در حوزه پزشکی رو به افزایش است، نیاز به افراد ماهر و با تجربه در طرح سؤالات چند گزینه‌ای جهت ارزیابی آزمون شونده‌گان نیز رو به افزایش است (۶) و این در وضعیتی است که معمولاً در دانشکده‌ها فرد متخصص و ماهر جهت ارزشیابی امتحانات وجود ندارد و بطور دوره‌ای سؤالات آزمونها از جهت استاندارد بودن مورد تجزیه بررسی قرار نمی‌گیرد (۷).

در مطالعات مختلف انجام شده در دانشگاههای علوم پزشکی کشور، مشخص شد که ارزشیابی‌های پایان ترم به عمل آمده از دانشجویان عمدتاً با استفاده از سؤالات چند گزینه‌ای انجام می‌گیرند و نقص‌های قابل ملاحظه‌ای دارند و مهارت طرح سؤال استاندارد در حال حاضر به عنوان یک مسأله قابل بحث آموزشی مطرح می‌باشد (۸). در تحقیقی که اشرفپور و همکاران در دانشگاه علوم پزشکی بابل در سال ۱۳۸۲ انجام دادند، نشان دادند که طراحی آزمونهای پایان ترم وضعیت مناسب و راضی‌کننده نداشته و عموماً آزمونهای مورد استفاده در فرآیند ارزشیابی برآیندهای آموزشی دانشجویان، دارای اشکال ساختاری اساسی بوده و این آزمونها به عنوان ابزار دقیق و معتبر عمل نمی‌نمایند و تفاوت‌های واقعی آزمودنی‌ها را نمی‌سنجند (۹).

بررسی دشواری و قدرت تمیز توسط متخصصان آزمون‌سازی همچون مهرنزلهمان (۱۹۸۴)، ایبل و فریزی (۱۹۹۱)، تورندایک و هیگن (۱۹۸۶) ضریب دشواری بین ۰/۳۰

برگه) بودند، به عنوان نمونه مورد مطالعه انتخاب شدند. برگه‌های این آزمونها پس از دریافت مجوز از ریاست و آموزش دانشکده برای استخراج اطلاعات مورد نیاز از آنها تحویل گرفته شدند. برای انتخاب برگه‌هایی که اطلاعات آنها لازم بود استخراج شود مطابق با روش ارائه شده توسط دکتر علی اکبر سیف به صورت زیر عمل شده است:

- برای هر کدام از شش آزمون، برگه‌های امتحانات به ترتیب نمره آنها از کمترین نمره تا بیشترین نمره (بطور صعودی) مرتب شده‌اند.

- برای تعیین تعداد برگه‌های گروه بالا و پایین با توجه به تعداد برگه‌های آزمون آنها بصورت زیر عمل شده است.

برای تعداد ۴ آزمون کاربرد رایانه، زبان انگلیسی، سیستم‌های رایج طبقه‌بندی و شاخصها و تحلیل داده‌های بهداشتی که بین ۲۰ تا ۴۰ برگه داشته‌اند پس از مرتب کرده آنها، ۱۰ برگه از گروه بالا (شروع از بیشترین نمره) و ۱۰ برگه هم از گروه پایین (شروع از کمترین نمره) انتخاب و مابقی کنار گذاشته شدند و برای اصطلاحات پزشکی ۳ و روش تحقیق که تعداد برگه‌های آنها کمتر از ۲۰ نفر ($N < 20$) بوده‌اند، پس از مرتب نمودن برگه‌ها بر اساس ترتیب صعودی نمرات، به دو قسمت مساوی (در حالتی که تعداد برگه‌ها فرد بود یکی از گروهها یک برگه بیشتر از گروه دیگر) تقسیم شدند.

برای محاسبه شاخص‌های کمی سؤالات از فرمولهای ضریب دشواری (Difficulty index) و ضریب تمیز (Discrimination index) سؤالات چندگزینه‌ای استفاده شده است. ضریب دشواری عبارتست از درصدی از افراد که به یک سؤال جواب درست داده‌اند، آن را با P نشان می‌دهند و مقداری بین ۰ و ۱ دارد و در تفسیر آن، چنانچه $P < 0/3$ باشد، سؤال سخت، $0/8 < P < 0/3$ سؤال مناسب و $P \geq 0/8$ سؤال خیلی ساده محسوب می‌شود. ضریب تمیز نیز شاخصی است که قدرت سؤال در جدا کردن افراد گروه قوی (که در آزمون نمره بالا گرفته‌اند) و ضعیف (که در آزمون نمره پایین گرفته‌اند) را مشخص می‌کند. آن را با D نمایش می‌دهند و مقداری بین +۱ تا -۱ می‌تواند داشته باشد و هرچه این ضریب بزرگتر باشد، بهتر است. در تفسیر آن چنانچه $D \leq 0/19$ باشد قدرت تمیز ضعیف تا بد، $0/3 < D < 0/2$ در حد مناسب،

خوب می‌باشد. $0/39 < D < 0/3$ خوب و $D \geq 0/4$ قدرت تمیز سؤال خیلی خوب می‌باشد.

برای محاسبه شاخصهای دشواری و تمیز در نرم‌افزار Excel برنامه‌ای برای محاسبه شاخصها نوشته شد و سپس اطلاعات هر کدام از برگه‌های آزمونها را با در نظر گرفتن عدد ۱ برای پاسخهای صحیح و عدد صفر برای پاسخهای غلط به سؤالات، وارد برنامه طراحی شده نموده و پس از محاسبه شاخصهای دشواری و تمیز هر سؤال به کمک نرم‌افزار آماری SPSS و با استفاده از روشهای آمار توصیفی و آزمونهای آنالیز واریانس و آنالیز رگرسیون و نمودار پراکنش تجزیه و تحلیل شدند و برای تمام آزمونهای آماری $P < 0/05$ بعنوان معنی‌دار لحاظ گردید.

نتایج:

در این مطالعه، ۱۸۲ سؤال MCQ که ۲۸ سؤال (۱۵/۴٪) مربوط به آزمون کاربرد رایانه، ۲۰ سؤال (۱۱٪) زبان انگلیسی، ۳۰ سؤال (۱۶/۵٪) سیستمهای رایج طبقه‌بندی، ۵۹ سؤال (۳۴/۴٪) اصطلاحات پزشکی ۳، ۳۱ سؤال (۱۷٪) روش تحقیق و ۱۴ سؤال (۷/۷٪) هم مربوط به درس شاخصها و تحلیل داده‌های بهداشتی برای تحلیل بکار گرفته شدند.

مطابق با جدول شماره ۱، از ۱۸۲ سؤال، ۱۳ سؤال (۷/۱٪) سخت تا خیلی سخت بودند که از این میان بیشترین نسبت سؤالات سخت را آزمونهای کاربرد رایانه (۱۰/۷٪) و روش تحقیق (۹/۷٪) و کمترین نسبت را هم دروس سیستمهای طبقه‌بندی و شاخصها و تحلیل بهداشتی بدون سؤال سخت در برداشته‌اند. تعداد ۶۴ سؤال (۳۵/۲٪) آسان وجود داشت که بیشترین نسبت آنها را درس اصطلاحات پزشکی (۵۴/۳٪) و کمترین نسبت را هم دروس روش تحقیق (۱۹/۳٪) و کاربرد رایانه (۲۱/۴٪) در بر داشته‌اند. بطور کلی ۱۰۵ سؤال (۵۷/۷٪) از ۱۲۸ سؤال مناسب بودند که بیشترین نسبت این سؤالات در بین سؤالات درس روش تحقیق (۷۱٪) و سیستمهای رایج (۷۰٪) و کمترین نسبت را هم درس اصطلاحات پزشکی (۳۹٪) داشته است. همچنین مطابق با اطلاعات جدول شماره ۱، کمترین سطح دشواری مربوط به درس روش تحقیق با متوسط ۰/۵۹ و بیشترین آن هم مربوط به درس اصطلاحات پزشکی ۳ با

متوسط ۰/۷۸ بدست آمده است. نتایج آزمون آنالیز واریانس یک طرفه نشان داده بین سطح دشواری سؤالات این ۶ درس تفاوت معنی‌داری وجود دارد ($P < ۰/۰۵$).

بررسی جدول شماره ۲ نشان می‌دهد، اکثریت سؤالات یعنی ۱۳۳ سؤالی (۰/۷۳/۱) دارای قدرت تمیز ضعیف که در این میان دروس سیستمهای رایج طبقه‌بندی (۰/۹۳/۳) و زبان انگلیسی (۰/۹۰) بیشترین نسبت سؤالات با قدرت تمیز ضعیف و درس روش تحقیق (۰/۳۵/۵) کمترین نسبت را داشتند. بر اساس این مطالعه سؤالاتی که دارای ضریب تمیز بین ۰/۲ تا ۰/۳ بودند، به عنوان سؤالات با قدرت جداسازی در حد متوسط محسوب شده‌اند. مطابق با جدول شماره ۲، برای ۱۸۲ سؤالی این ۶ آزمون، تعداد ۲۲ سؤالی (۰/۱۲/۱) با قدرت تمیز در حد متوسط که روش تحقیق (۰/۱۹/۴) و اصطلاحات پزشکی (۰/۱۵/۳) بیشترین نسبت و بقیه دروس با اندکی تفاوت تقریباً به نسبت یکسان سؤالات با قدرت تمیز متوسط را در بر داشته‌اند. تنها ۱۱ سؤالی (۰/۶) از سؤالات دارای ضریب تمیز خوب بودند که ۷ سؤالی آن مربوط به درس روش تحقیق، ۳ سؤالی در درس کاربرد رایانه و ۱ سؤالی هم مربوط به درس شاخصها و تحلیل

بهداشتی بوده است. نهایتاً ۱۶ سؤالی (۰/۸/۸) هم دارای ضریب تمیز خیلی خوب بودند که روش تحقیق با ۷ سؤالی و اصطلاحات و کاربرد رایانه نیز با ۴ سؤالی بیشترین تعداد سؤالی با قدرت تمیز خیلی خوب را داشتند. در مقایسه میانگین قدرت تمیز دروس، دروس زبان انگلیسی و سیستمهای رایج طبقه‌بندی به ترتیب با متوسط ضریب تمیز ۰/۰۵- و ۰/۰۳- ضعیف‌ترین قدرت تمیز و دروس روش تحقیق و کاربرد رایانه نیز به ترتیب با میانگین ۰/۲۳ و ۰/۱۸، بالاترین قدرت جداسازی دانشجویان قوی و ضعیف را داشته‌اند. همچنین نتایج آزمون آنالیز واریانس نشان داد بین میانگین ضریب تمیز این ۶ درس تفاوت معنی‌داری وجود دارد ($P < ۰/۰۵$). در بررسی نمودار پراکنش و نتایج آزمون رگرسیون (شکل شماره ۱). مشخص شد رابطه خطی یا غیرخطی بین ضریب دشواری و تمیز ۱۸۲ سؤالی این ۶ آزمون وجود ندارد ($R^2 = 0$) اما با کمی دقت به نمودار، یک شکل گنبدی مانند را می‌توان دید و می‌توان یک منحنی درجه دوم به آن برازش داد. به عبارت دیگر تقریباً یک رابطه غیرخطی درجه دوم بین این دو شاخص مشهود است.

جدول شماره ۱- مقایسه میانگین و توزیع فراوانی دشواری سؤالات آزمونها

درس	$P < ۰/۳$ (سخت)		$۰/۳ < P < ۰/۸$ (مناسب)		$P < ۰/۸$ (آسان)		متوسط دشواری
	تعداد	درصد	تعداد	درصد	تعداد	درصد	
کاربرد کامپیوتر	۳	۱۰/۷٪	۱۹	۶۷/۹٪	۶	۲۱/۴٪	۰/۶۲
زبان انگلیسی	۳	۱۵٪	۱۱	۵۵٪	۶	۳۰٪	۰/۷۲
سیستم‌های طبقه‌بندی	۰	۰٪	۲۱	۷۰٪	۹	۳۰٪	۰/۷۲
اصطلاحات پزشکی ۴	۴	۶/۸٪	۲۳	۳۹٪	۳۲	۵۴/۳٪	۰/۸۷
روش تحقیق	۳	۹/۷٪	۲۲	۷۱٪	۶	۱۹/۳٪	۰/۵۹
شاخص‌ها و تحلیل بهداشتی	۰	۰٪	۹	۶۴/۳٪	۵	۳۵/۷٪	۰/۶۹
کل	۱۳	۷/۱٪	۱۰۵	۵۷/۷٪	۶۴	۳۵/۲٪	$F = ۴/۰۷$ و $P = ۰/۰۰۲$

جدول شماره ۲- مقایسه میانگین و توزیع فراوانی درجه تمیز سؤالات آزمونها

درس	$D < ۰/۲$ (ضعیف)		$۰/۲ < D < ۰/۳$ (متوسط)		$۰/۳ < D < ۰/۴$ (خوب)		$D > ۰/۴$ (خیلی خوب)		متوسط تمیز
	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	
کاربرد کامپیوتر	۱۹	۶۷/۹٪	۲	۷/۱٪	۳	۱۰/۷٪	۴	۱۴/۳٪	۰/۱۸
زبان انگلیسی	۱۸	۹۰٪	۲	۱۰٪	۰	۰٪	۰	۰٪	۰/۰۳
سیستم‌های طبقه‌بندی	۲۸	۹۳/۳٪	۲	۶/۷٪	۰	۰٪	۰	۰٪	۰/۰۵
اصطلاحات پزشکی ۴	۴۶	۷۸٪	۹	۱۵/۳٪	۰	۰٪	۴	۶/۸٪	۰/۰۹
روش تحقیق	۱۱	۳۵/۵٪	۶	۱۹/۴٪	۷	۲۲/۶٪	۷	۲۲/۶٪	۰/۲۳
شاخص‌ها و تحلیل بهداشتی	۱۱	۷۸/۶٪	۱	۷/۱٪	۱	۷/۱٪	۱	۷/۱٪	۰/۱۵
کل	۱۳۳	۷۳/۱٪	۲۲	۱۲/۱٪	۱۱	۶٪	۱۶	۸/۸٪	$F = ۴/۰۷$ و $P = ۰/۰۰۲$

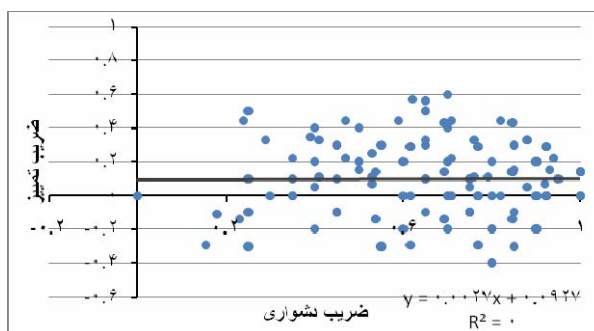
در یک مطالعه مداخله که شبان و همکاران (۱۱) در دانشگاه علوم پزشکی تهران انجام دادند، نشان دادند قبل از آموزش ۵۶/۵٪ سؤالات آزمونهای انتخابی اعضای هیئت علمی و بعد از آموزش از طریق یک جزوه درسی هم ۵۹/۳٪ آنها دارای سطح دشواری قابل قبول بودند.

در مطالعه دیگری که توسط اشرفپور و همکاران (۹) بر روی دروس تخصصی و پایه رشته پرستاری دانشگاه علوم پزشکی بابل در نیمسال تحصیلی دوم ۷۸-۷۹ انجام داده بودند، سطح دشواری قابل قبول را فاصله عدد ۳ تا ۰/۷ در نظر گرفته بودند و نتایجی که بدست آوردند بدین صورت بود که پایین‌ترین سطح دشواری برای درس پرستاری ویژه با میزان ۰/۵۹ و بیشترین میزان آن هم برای درس دروس تغذیه و پرستاری بهداشت کودک و مادر با میزان ۰/۷۸ همچنین نتایج این مطالعه نشان داد ۱۴۸۵ سؤال تنها ۳۸٪ آنها دارای ضریب دشواری مناسب بودند.

بطور کلی یک آزمون MCQ باید به نسبت مناسب هم سؤال سخت داشته باشد و هم سؤال آسان ولی برخی مواقع تعداد زیاد سؤالات ساده آزمونهای یک مدرس موجب تلاش کمتر دانشجویان برای یادگیری محتوای درسی می‌شود و این موضوع بخصوص برای دانشجویان رشته پزشکی و پیراپزشکی نتایج زیانباری را متوجه جامعه خواهد ساخت (۱۳). بطور مثال تحلیل آزمونهای ارتقاء رشته های پزشکی در تیرماه ۷۷ نشان داد که ۳۶/۳٪ از سؤالات دارای ضریب دشواری در حد آسان، ۱۷/۳٪ در حد سخت و تنها ۵۴٪ آنها دارای ضریب سختی در حد قابل قبول بودند (۹) که این مسئله می‌تواند منجر به ارتقاء نامناسب دانشجویان رشته پزشکی شود.

شاخص تمیز، قدرت یک سؤال را در تمایز یا جداسازی آزمون‌شوندگان (دانشجویان) ضعیف از قوی نشان می‌دهد (۳). در تعیین کیفیت سؤالات آزمونها علاوه بر سطح دشواری، شاخص تمیز نیز نقش مهمی را دارد بطوری که تلفیق این دو شاخص در شناسایی سؤالات مناسب می‌تواند کمک بزرگی به طراحان سؤالات آزمونها داشته باشد.

در این مطالعه سؤالاتی دارای درجه تمیز زیر ۰/۲ دارای قدرت تمیز ضعیف تا خیلی ضعیف، درجه تمیز ۰/۲ تا ۰/۳ دارای



نمودار شماره ۱- رابطه بین سطح دشواری و تمیز برای ۱۸۲ سؤال ۶ آزمون

بحث و نتیجه‌گیری:

تجزیه و تحلیل سؤالات آزمونها یک فرآیندی است که در آن پاسخ دانشجویان به تک تک سؤالات جهت ارزیابی کیفیت هر سؤال و بطور کلی کیفیت آزمون بکار گرفته می‌شود. این فرآیند می‌تواند منجر به بهبود کیفیت سؤالات و مهارت مدرس سازنده سؤال در آزمونهای بعدی شود (۱۲). از جمله شاخصهای کمی که به وسیله آنها کیفیت سؤالات آزمونهای چندگزینه‌ای و حتی تشریحی را مورد بررسی قرار می‌دهند، شاخصهای دشواری و تمیز می‌باشند (۱) که در این مطالعه نیز برای بررسی کیفیت سؤالات چند گزینه‌ای آزمونهای رشته کارشناسی مدارک دانشگاه علوم پزشکی هرمزگان از آنها استفاده شده است.

نظرات متفاوتی در خصوص سطح دشواری مناسب یک سؤال وجود دارد ولی بطور کلی سطح دشواری مناسب و قابل قبول بین ۰/۳ تا ۰/۸ است و اگر کمتر از ۰/۳ سؤال را سخت تا خیلی سخت و اگر به عدد ۱ نزدیک شود آن را آسان تا خیلی آسان نامند. در این مطالعه کمترین میانگین ضریب دشواری ۰/۵۹ (روش تحقیق) و بیشترین آن ۰/۷۸ (اصطلاحات پزشکی ۳) بود. مطالعات متعددی پیرامون بررسی کیفیت آزمونها با استفاده از این دو شاخص انجام شده که در مطالعه راجا (۳) که به بررسی رابطه بین سطح دشواری چند درس علوم پایه طی سالهای ۲۰۰۱ تا ۲۰۰۴ در دانشگاه علوم پزشکی مالایای مالزی پرداخته بود، کمترین میزان سطح دشواری را ۰/۵۹ برای بخش C در سال تحصیلی ۲۰۰۱-۲۰۰۲ و بیشترین میانگین دشواری را ۰/۶۷ برای بخش A در سال تحصیلی ۲۰۰۱-۲۰۰۲ بدست آورد.

قدرت تمیز در حد متوسط، $0/3$ تا $0/4$ در حد خوب و بالاتر از $0/4$ هم قدرت جداسازی یا تمیز خیلی خوب معرفی شده‌اند.

از طرفی دیگر چنانچه ضریب تمیز سئوالی عدد منفی مفهوم آن این است که افراد ضعیف بهتر از افراد قوی توانسته‌اند به آن سئوال جواب دهند، در این حالت اگر سئوال اشکال ساختاری و فنی داشته باشد باید به کلی کنار گذاشته شود ولی چنانچه سئوال اشکال فنی نداشته باشد دلیل آن می‌تواند این باشد که افراد گروه بالا یا آن سئوال را به طور کامل یاد نگرفته‌اند یا آن را به غلط آموخته‌اند ($13/14$). در بررسی ضریب تمیز مطالعه ما مشخص شد که بجز درس روش تحقیق که دارای ضریب تمیز نسبتاً مناسب ($0/23$) بود برای سایر دروس، آزمونها از درجه تمیز ضعیف تا خیلی ضعیف برخوردار بودند. به طوری که برای دو درس سیستم‌های رایج طبقه‌بندی با متوسط قدرت تمیز $0/05$ - که $93/3\%$ سؤالات آن قدرت تمیز کمتر از $0/2$ و درس زبان انگلیسی با قدرت تمیز $0/03$ - و 90% سؤالات قدرت تمیز کمتر از $0/2$ نه تنها قدرت تمایز افراد ضعیف و قوی را به درستی نشان ندادند، بلکه سؤالات این آزمون‌ها به گونه‌ای بوده که افراد ضعیف بهتر از افراد قوی پاسخ‌های درست را انتخاب نموده‌اند و در وضعیتی است که مطابق با اصول تجزیه و تحلیل آزمون‌ها، تقریباً تمامی سؤالات این دو آزمون باید کنار گذاشته شوند و از این سؤالات در آزمونهای بعدی استفاده نکرد. در حالت کلی از تعداد ۱۸۲ سئوال این شش آزمون 27% از سؤالات دارای ضریب تمیز متوسط تا خیلی خوب و 73% دیگر قدرت تمیز ضعیف تا خیلی ضعیف داشتند. همچنین نتایج آزمون آنالیز واریانس نشان داد از نظر آماری تفاوت معنی‌داری بین میانگین قدرت تمیز این شش درس وجود دارد.

در مطالعه اشرفیور و همکاران (۹) متوسط درجه تمیز برای تمامی سؤالات $0/14$ بدست آمده که از میزان مطلوبی برخوردار نبود به طوری که حداکثر سطح دشواری برای دو درس مدیریت پرستاری ($0/22$) و روانپرستاری ($0/21$) و کمترین آن هم برای درس فوریتهای پرستاری ($0/09$) بدست آمد و نتیجه گرفته شده که سؤالات از قدرت تمیز مناسبی برخوردار نمی‌باشند و علت آن را هم شیوه نامناسب طراحی سؤالات مطرح کرده بود.

در مطالعه شبیان و همکاران (۱۱) مشخص شد قدرت تمیز $0/6$ تا $0/4$ سؤالات بعد از مداخله در حد ضعیف یعنی کمتر از $0/2$ بوده است.

در مطالعه راجا و همکاران (۳) نیز کمترین میزان درجه تمیز ($0/2$) برای سؤالات بخش B در سال تحصیلی $2003-2002$ و بیشترین میزان آن ($0/33$) برای سؤالات بخش A سال در تحصیلی $2002-2001$ بدست آمد.

بطور کلی یافته‌های این مطالعه نشان داد تقریباً ضریب سختی تمامی آزمونها در حد مناسب و قابل قبولی قرار دارد ولی باید توجه داشت که سطح کیفیت یک آزمون را باید با توجه به هر دو شاخص کمی یعنی ضریب دشواری و ضریب تمیز مورد ارزیابی قرار داد (15). از آنجایی که تمامی سؤالات آزمونها از درجه تمیز نامناسب ضعیف تا خیلی ضعیف برخوردار هستند، بهتر است سؤالات این آزمونها حذف یا بطور اساسی مورد تجدید نظر قرار گیرند.

رابطه بین ضریب دشواری و تمیز در این مطالعه نشان داد که یک رابطه خطی بین دو شاخص وجود ندارد ($R^2=0$)، اما ممکن است بتوان یک رابطه غیرخطی از نوع درجه دوم را به آن برآزش نمود. در صورتی که در مطالعه‌ای که راجا (۳) بر روی ۲۵۰ سئوال MCQ دو سال 2001 تا 2002 دروس علوم پایه سه بخش A، B و C دانشگاه علوم پزشکی مالایای مالزی انجام داد به این نتیجه رسید که برای هر سه بخش بین سطح دشواری و تمیز نمرات یک رابطه غیر خطی درجه دوم وجود دارد. شاید با افزایش تعداد سؤالات به 250 یا بیشتر از آن رابطه (خطی یا غیر خطی) بین دو شاخص دشواری و تمیز در این مطالعه نیز مشهود می‌شد.

سپاسگزاری:

لازم می‌دانم از معاونت محترم پژوهشی دانشگاه علوم پزشکی هرمزگان، ریاست محترم دانشکده، مسئولین آموزش و کارشناس مرکز توسعه آموزش دانشکده پیراپزشکی هرمزگان که در تصویب طرح، پیگیری و صدور مجوزها همکاری نمودند، صمیمانه سپاسگزاری نمایم.

References

منابع

1. Seaf AA. Educational Measurement, Assessment and evaluation. 3thed, Tehran: Dowran company; 1976. [Persian]
2. Mehrens WA, Lehmann IJ. Measurement and evaluation in education and psychology. 4th ed. New York, Wadsworth Publishing; 1991.
3. Raja MS, Rasiah I, Relationship Between Item Difficulty and Discrimination Indices in True/False-Type Multiple Choice Questions of a Para-clinical Multidisciplinary Paper. 2006,35:67-71
4. schultheis NH. Writing cognitive educational objectives and multi choice test questions. *Am J Health Syst Pharm.* 1998,55:2397-2401.
5. Van Hoozer H. The teaching process: theory and practice in nursing. Norwalk, CT: Appleton-Century-Crofts, 1987,11:279-280.
6. Writing Multiple Choice Questions for Continuing Medical Education Activities and Self-Assessment Modules. [cited 2010 Apr 01]. Available from: URL:<http://www.arrs.org/uploadedFiles/ARRS/Publications/writingMultipleChoiceHandout.pdf>.
7. Crisp G, Palmer E. Engaging Academics with a Simplified Analysis of their Multiple-Choice Question (MCQ) Assessment Results. *Journal of University Teaching and Learning Practice.* 2007;4:88-106.
8. Salemi A, Rafiei M. Effects of workshops on teaching methods of using skill of design questions among teachers. Special proceedings of the seventh national conference on medical education: 2004; Tehran, Iran.
9. Ashrafpour M, Beheshti Z, Molookzadeh F. Quality of final examination in students of Babol medical university, 1999-2000. *Journal of Babol University of Medical Sciences.* 2003;0:42-47. [Persian].
10. Sepasi. Analysis of indicators of levels of cognitivedomain and psychometrics final exam Arabic lessons, biology and calculus girl students of the third school in three different regional economic – social Khuzestan province. *Iranian Journal of Psychology & Cultural Science of Shahid Chamran University of Ahvaz.* 2006;13:57-78. [Prsian]
11. Shaban M, Ramazani Badr F. Effect of test item analysis of summative exams on quality of test designing. *The Journal of Faculty of Nursing & Midwifery.* 2007;13:5-15. [Persian]
12. Item Analysis. Available from: URL: <http://www.washington.edu/oea/score1.htm>: [cited 2010 Apr 05].
13. Rassaian N, Nakhaei S, Sadeghi Ghandehari N. Comparison of three exam teaching in medical students: Multi-choice, true-false and short answer questions. *Hakim Research Journal.* 2003;5:271-278. [Persian]
14. Rezae H. Assessment techniques in medicine Tehran: Amad press; 1998. [Persian]
15. Vazquez Alonso A, Manassero MA, Acevedo JA. Quantitative analysis of complex multiple-choice items in science technology and society: Item scaling. *Revista Electrónica de Investigación Educativa.* 2005;7:1-29.

Quality analysis of multiple choice questions(MCQs) examinations of non-continuous undergraduate medical records

S. Hosseini Teshnizi, MSc¹ S. Zare, PhD² M. Solati, MD³

Instructor Department of Biostatistics¹, Associate Professor Department of Community Medicine², Assistant Professor Department of Internal Medicine³, Hormozgan University of Medical Sciences, Bandar Abbas, Iran.

(Received 27 Sep, 2010 Accepted 5 Oct, 2010)

ABSTRACT

Introduction: There are Different methods for evaluating student stuff. One of the most commons is multiple choice questions (MCQs). If properly designed, it is a good way to measure student knowledge. Due to expansion using MCQs, This study was designed in order to review the quality of multiple choice question exam types of medical records students of Hormozgan University of Medical Sciences.

Methods: In a descriptive and analytical study, all end of semester MCQ type examinations related to undergraduate degree of medical records course of Hormozgan University of Medical Sciences, including computer applications, English language, common classification systems, medical terminology 3, research methodology and data analysis of health indicators were selected. Then difficulty and discrimination index of all questions were calculated. For statistical analysis of the data, by taking advantage of Excel and SPSS software's, descriptive statistics, analysis of variance and regression methods were used.

Results: The results showed that, from 128 questions, 105 questions (57.7%) had a suitable difficulty index, which the lowest rate was related to the research methods course (0.59) and the highest rate was related to the medical terminology3 (0.78). 133 questions (73.1%) were in weak and very weak discrimination index category, so that the lowest rate was related to English language (-0.05) and common systems (-0.03) and the most rate was related to the research methodology (0.23) and use of computer application (0.18). Results of analysis of variance showed a significant difference, between difficulty and discrimination index of six exams ($P < 0.05$). There was no significant correlation between difficulty and discrimination indexes.

Conclusion: Although, the difficulty index for all examinations were acceptable and appropriate but the results of discrimination index showed that, almost for all questions, the quality level is poor to very poor. So, it is better to remove these questions from the bank of questions or basically revise all the questions.

Key words: Medical Records – Students – Evaluation

Correspondence:
S. Hosseini Teshnizi, MSc.
Nursing Midwifery School.
Hormozgan University of
Medical Sciences.
Bandar Abbas, Iran
Tel: +98 761 6666365
Email:
shosseini@hums.ac.ir