

ارتقای کیفیت طراحی سؤالات چند گزینه‌ای آزمون‌های ارتقای دستیاری از طریق ارایه بازخورد

اعظم معیاری^۱، مهدی بیگلرخانی^{۱*}

۱. پزشک عمومی، کارشناس ارشد، گروه آموزش پزشکی، مرکز مطالعات توسعه آموزش پزشکی، دانشگاه علوم پزشکی همدان، ایران
۲. پزشک و کارشناس ارشد آموزش پزشکی، مربی گروه آموزش عمومی دانشکده پزشکی، دانشگاه علوم پزشکی همدان، ایران
• دریافت مقاله: ۹۱/۴/۱۹ آخرین اصلاح مقاله: ۹۱/۷/۱۱ • پذیرش مقاله: ۹۱/۷/۱۸

زمینه و هدف: هدف از انجام این مطالعه، ارتقای کیفیت طراحی سؤالات آزمون‌های ارتقای دستیاری از طریق به کارگیری تکنیک‌های ارایه بازخورد مورد نظر در علم آموزش پزشکی به طراحان سؤالات چند گزینه‌ای بود.

روش کار: در این مطالعه نیمه تجربی، سؤالات آزمون‌های ارتقای دستیاری پزشکی در سال ۱۳۸۸ از نظر ساختار طراحی با استفاده از چک‌لیست روا و پایای Millman مورد بررسی قرار گرفت. همچنین توزیع سطح دانش مورد سنجش در هر یک از سؤالات نیز مشخص گردید. بر اساس اطلاعات حاصل از بازخورد مداخله‌ای با رعایت خصوصیات فنی آن در علم آموزش پزشکی یعنی عینی، به موقع، مستقیم، محترمانه، بدون قضاوت از عملکرد و همراه با توصیه‌هایی جهت بهبود، به طراحان آزمون که همگی سابقه شرکت در کارگاه‌های آموزشی MCQ (Multiple choice question) و تجربه طراحی سؤال در سال‌های قبل را داشتند، ارایه شد. بعد از برگزاری آزمون ارتقای سال ۱۳۸۹ نیز سؤالاتی که طراحان در هر دو سال در طراحی آن‌ها نقش داشتند، انتخاب شدند و میزان رعایت اصول استاندارد Millman و سطح تاکسونومی به کار رفته در سؤالات توسط طراحان قبل و بعد از مداخله بازخوردی مورد مقایسه قرار گرفت.

یافته‌ها: از مجموع ۲۵۰۰ سؤال، ۲۰۳۵ سؤال وارد مطالعه شد. قبل از مداخله ۶۵/۲ درصد از سؤالات از نظر ساختار کلی بدون اشکال بود و پس از انجام مداخله این میزان به ۸۲/۸ درصد رسید که این تفاوت معنی‌دار بود ($P < ۰/۰۰۱$). از نظر سطح دانش مورد ارزیابی در سؤالات نیز در سال ۱۳۸۸، ۳۸ درصد از سؤالات و در سال ۱۳۸۹، ۵۳/۱ درصد از سؤالات با تاکسونومی بالا طراحی شده بود که تفاوت آن‌ها نیز معنی‌دار بود ($P < ۰/۰۰۱$). نتیجه‌گیری: انجام مداخله به صورت بازخورد و منطبق با اصول فنی آن در آموزش پزشکی حتی جهت طراحان آموزش دیده و با تجربه در طراحی سؤالات نیز می‌تواند در ارتقای کیفیت طراحی سؤالات چند گزینه‌ای مؤثر باشد.

کلید واژه‌ها: سؤالات چندگزینه‌ای، دستیاری، آموزش پزشکی، تاکسونومی، قواعد Millman، بازخورد

*نویسنده مسؤول: دانشکده پزشکی، دانشگاه علوم پزشکی، بلوار شهید فهمیده، همدان، ایران

مقدمه

استفاده از ارزیابی در فعالیتهای آموزشی به اندازه خود فعالیتهای آموزشی قدمت دارد. در کشور ما آزمونهای چند گزینه‌ای، شایعترین فرم سؤالات را در امتحانات کتبی مربوط به گروههای پزشکی تشکیل داده است. نتایج نشان می‌دهد که در همه دانشکده‌ها بیشتر مدرسان در امتحان پایان ترم از سؤالات چند گزینه‌ای استفاده می‌کنند. این آزمون همانند دیگر آزمون‌ها دارای معایب و محاسنی است که با علم به آنها و تلاش در رفع معایب آن می‌توان تا حد زیادی آن را بهبود بخشید. از جمله محاسن آن یکسان بودن سؤالات برای تمام دانشجویان، عملی بودن، بالا بودن میزان پایایی، سهولت تصحیح پاسخنامه‌ها و عینی بودن آن است. گرچه استفاده از این نوع سؤالات در بررسی سطح دانش در گروههای بزرگ از آزمون شوندگان مفید است، ولی عدم آشنایی با فنون طراحی آنها به طور قابل ملاحظه‌ای از قدرت و عملکرد این نوع آزمون خواهد کاست. از محدودیتهای این آزمون‌ها می‌توان به دشواری طراحی سؤال طبق قواعد علمی و با تاکسونومی مناسب، سر نخ دادن و ... اشاره نمود. برای طراحی بهتر این گونه سؤالات رعایت نکاتی که اولین بار توسط Millman ذکر شده است، ضروری می‌باشد. مطالعات نشان می‌دهد که اغلب طراحان سؤال مهارت چندانی در طرح سؤال ندارند و شاید سؤالاتی طراحی کنند که آزمون شوندگان فاقد علم و دانش کافی، با توجه به سر نخ‌هایی که داده شده است، بتوانند گزینه درست را انتخاب کنند (۱-۳).

همچنین به منظور تسهیل در یادگیری و سنجش آموخته‌ها، اهداف آموزشی یا اهداف یادگیری تدوین و مورد سنجش قرار می‌گیرد که با توجه به آنها تاکسونومی سؤالات طراحی شده تعیین می‌شود. تاکسونومی سؤال، طبقه‌بندی است که در آن سطح دانش آزمون شونده سنجیده می‌شود و مراحل آن عبارت است از یادآوری ساده مطالب (تاکسونومی ۱)، تفسیر اطلاعات (تاکسونومی ۲) و حل مسأله و کاربرد مطالب (تاکسونومی ۳) و در این راستا طراحی سؤالات با تاکسونومی بالاتر درک

عمیق‌تر آزمون شوندگان را مورد ارزیابی قرار می‌دهد، از اهمیت بیشتری برخوردار خواهد بود.

از سوی دیگر، در هر سیستم آموزشی، ارایه بازخورد مناسب به آموزش دهندگان، طراحان و مسؤولین آموزشی، یکی از اجزای مهم در ارتقای کیفیت آموزش است. استفاده از بازخورد به مفهوم فنی و علمی آن در آموزش پزشکی یعنی نقد عینی، آگاهانه و بدون قضاوت از عملکرد فرد یا افراد محدود به یک عملکرد خاص و به موقع که با هدف بهبود مهارت‌ها به صورتی مستقیم و محترمانه، ارایه می‌شود (۴)، ممکن است که در رفع نواقص و ایرادات طراحی آزمون‌های چند گزینه‌ای در نوبت‌های بعدی مؤثر واقع شود. از این رو در این تحقیق سعی شد که از این تکنیک در جهت ارتقای کیفیت طراحی آزمون‌ها توسط استادان هم از منظر ساختاری و هم از منظر تاکسونومی سؤالات استفاده گردد.

تحقیقات متعددی در خصوص آزمون‌های برگزار شده در دانشگاه‌ها و مؤسسات آموزش عالی داخل و خارج از کشور صورت گرفته است. نتایج این بررسی‌ها نشان داده است که در این آزمون‌ها از سؤالات مناسب کمتر استفاده می‌شود و در اجرا و ساخت آزمون‌ها مسایل و مشکلاتی وجود دارد. از جمله در تحقیقی که توسط واعظی و همکاران در بررسی سؤالات آزمون‌های پایان ترم دانشگاه علوم پزشکی یزد انجام شد، تنها ۲۴/۶ درصد از سؤالات دارای وضعیت خوب بودند (۵). در مطالعه حق‌شناس و همکاران در مورد آزمون‌های دانشکده پزشکی مازندران بیش از سه چهارم سؤالات در تاکسونومی ۱ مطرح شده بود (۶). در مطالعه بیگلرخانی و همکاران نیز ۳۵/۹ درصد از سؤالات آزمون ارتقای دستیاری پزشکی در سال ۱۳۸۷ و ۳۱/۷ درصد از سؤالات در سال ۱۳۸۶ با تاکسونومی بالا طراحی شده بوده‌اند (۷). مطالعات دیگر حاکی از آن است که با انجام مداخلات آموزشی وضعیت طراحی سؤالات بهبودی قابل توجهی می‌یابد. به طور مثال، مطالعه معیاری و همکاران تفاوت آماری معنی‌داری را در ارتقای کیفیت طراحی سؤالات از دو منظر تاکسونومی و

هر دو سال نقش داشتند، انجام شد. این رشته‌ها عبارت از رشته‌های داخلی، جراحی، بیهوشی، پاتولوژی، کودکان، زنان و زایمان، پوست و روان‌پزشکی بود. به دنبال جمع‌آوری اطلاعات مربوط به تک‌تک سؤالات سال ۱۳۸۸ و مشخص شدن نقاط ضعف و قوت آن‌ها، مداخله لازم به صورت رایج بازخورد مورد نظر در علم آموزش پزشکی به طراحان سؤالات مربوط انجام شد و متعاقب طراحی سؤالات آزمون ارتقای سال بعد توسط آنان، تک‌تک این سؤالات نیز با همان پرسش‌نامه مورد بررسی و نتایج حاصل از انجام مداخله بازخوردی با سال قبل از مداخله مورد مقایسه قرار گرفت. با خارج کردن سؤالاتی که طراحان آن‌ها فقط در یکی از این دو سال در طراحی سؤال نقش داشتند، تعداد سؤالاتی که وارد مطالعه شد، در سال ۱۳۸۸ از ۱۲۵۰ سؤال به ۹۳۸ سؤال و در سال ۱۳۸۹ از ۱۲۵۰ سؤال به ۱۰۹۷ سؤال رسید. به عبارت دیگر، در مجموع در هر دو سال تعداد ۲۰۳۵ سؤال وارد مطالعه شد. ابزار سنجش در این مطالعه پرسش‌نامه‌ای سه بخشی بود که بخش اول آن حاوی پرسش‌هایی بود که شناسه‌های سؤال شامل رشته آموزشی، سال آزمون، کد طراح و شماره سؤال را بررسی می‌کرد. بخش دوم، چک‌لیستی مبتنی بر اصول Millman و چندین منبع دیگر بود (۱۲، ۱۱، ۳، ۲)، که روایی آن با توجه به استاندارد بودن چک‌لیست و بومی‌سازی و استفاده مکرر توسط محققین داخلی برآورده شده بود (۱۷-۱۳، ۹، ۶، ۵). چک‌لیست مزبور مشتمل بر ۱۸ آیتم در خصوص رعایت قواعد ساختاری در سه بخش تنه (۳ آیتم)، گزینه‌ها (۹ آیتم) و کلیات سؤال (۶ آیتم) بود و بر اساس رعایت یا عدم رعایت هر یک از آیتم‌ها (که در پرسش‌نامه به صورت بلی یا خیر قید شده بود)، نمره ۱ یا صفر به آن تعلق می‌گرفت. بدین ترتیب تنها در صورتی یک سؤال بدون اشکال ساختاری قلمداد می‌شد که از مجموع امتیازات آیتم‌ها، امتیاز ۱۸ را کسب می‌کرد و به همین ترتیب ملاک وجود یا عدم وجود اشکال ساختاری در قسمت تنه، گزینه‌ها و کلیات سؤال، دریافت یا عدم دریافت نمره کامل از هر قسمت بود. قسمت سوم این

ساختار پس از برگزاری کارگاه آموزشی جهت استادان جوان طراح سؤالات در آزمون‌های ارتقای دستیاری دندان‌پزشکی نشان داد (۸). وحیدشاهی و همکاران نیز نشان دادند که با انجام مداخله آموزشی از نوع رایج بازخورد و متون تحصیلی و آموزشی به دبیران و اعضای هیأت‌های ممتحنه، می‌توان تأثیر مثبتی بر طراحی سؤالات آزمون‌های ارتقای دستیاری گذاشت (۹).

در میان آزمون‌های سنجش پزشکی، بی‌تردید آزمون ارتقای دستیاری چه از نظر بر عهده داشتن وظیفه تأیید صلاحیت علمی دستیاران و چه از نظر سطح تحصیلی آزمون‌شوندگان و آزمون‌گیرندگان، اهمیتی ویژه و وظیفه‌ای خطیر در قبال نظام سلامت و همچنین نظام آموزش پزشکی بر عهده داشته است، از این رو توجه به کیفیت آن ضرورت کامل دارد. همچنین از آن جا که در سال‌های اخیر طراحی و اجرای آزمون‌های ارتقای دستیاری رشته‌های تخصصی پزشکی به عهده دانشکده‌های پزشکی نهاده شده است و بر اساس توصیه‌های وزارت متبوع، استفاده هر چه بیشتر از سؤالات با سطوح تاکسونومی بالاتر مورد تأکید قرار گرفته است، توجه و اهتمام به برگزاری و انتخاب آزمون‌های چند گزینه‌ای استاندارد، امری ضروری می‌باشد. این در حالی است که نتایج مطالعات مختلف حاکی از وجود نقص در طراحی سؤالات چند گزینه‌ای آزمون‌های ارتقای دستیاری بوده‌اند (۱۰، ۹، ۷)، از این رو، این پژوهش به منظور تعیین نقاط قوت و ضعف آزمون‌های ارتقای دستیاری بالینی سال ۱۳۸۸ دانشکده پزشکی همدان و در نتیجه رایج بازخورد نقاط قوت و ضعف آزمون‌های مزبور به همان افرادی انجام شد که در طراحی سؤالات آزمون سال ۱۳۸۹ نیز نقش داشته‌اند.

روش کار

این پژوهش یک مطالعه نیمه تجربی بود که به صورت قبل و بعد از مداخله و به شیوه سرشماری بر روی کلیه سؤالات سال‌های ۱۳۸۸ و ۱۳۸۹ که طراحان آن‌ها در تدوین سؤالات

مداخله لازم به صورت ارایه بازخورد نقاط قوت و ضعف سؤالات طراحی شده توسط آنان در سال قبل (سال ۱۳۸۸) به یکایکشان ارایه شد. همچنین از آن جا که تمرکز اصلی مداخله در مطالعه بر انجام بازخورد بود، ارایه بازخورد اصول فنی و مشخص در علم آموزش پزشکی، به طور کامل مورد رعایت قرار می‌گرفت. بدین شکل که به صورت عینی، مستقیم، محترمانه، محدود به عملکرد خاص فرد در طراحی همان سؤال و بدون قضاوت از وی، در موقع مناسب همراه با ارایه توصیه‌هایی برای رفع نواقص و بهبود آن فعالیت نیز همراه بود (۱۸).

در مرحله بعدی طرح و پس از مداخله اشاره شد که به صورت بازخورد ارایه گردید. متعاقب طراحی سؤالات آزمون ارتقا توسط طراحان و برگزاری آزمون ارتقای دستیاری، یک نسخه از کلیه سؤالات آزمون و کلید آن‌ها اخذ شد و با استفاده از همان پرسش‌نامه‌های کددار توسط پژوهشگران مورد بررسی قرار گرفت و اطلاعات مربوط به کلیات، تنه، گزینه‌ها و تاکسونومی سؤالات مربوط به این سال نیز استخراج شد. دوباره مانند سری قبل، جلسه مشترکی بین محققین به منظور رفع ابهامات و ایرادات، تشکیل و توافق نظر لازم حاصل گردید. داده‌های حاصل بدون ذکر نام افراد طراح سؤال به نرم‌افزار SPSS نسخه ۱۵ (SPSS Inc., Chicago, IL) (version 15, وارد گردید و با استفاده از آمار توصیفی (جدول توزیع فراوانی، شاخص‌های آماری و نمودار) و آزمون تحلیلی مقایسه نسبت‌ها (آزمون Z) با خطای آلفا کمتر از ۰/۰۵، سؤالات بدون اشکال ساختاری و سؤالات با تاکسونومی ۲ و ۳ (تاکسونومی بالا) در دو سال ۱۳۸۸ و ۱۳۸۹ (قبل و بعد از مداخله) مورد بررسی و مقایسه قرار گرفت. اصول اخلاقی در مطالعه، مورد رعایت قرار گرفت؛ بدین صورت که بازخورد نقاط ضعف در طراحی سؤالات به صورت کتبی و فقط به خود طراح ارایه شد و همچنین نتایج بدون ذکر نام طراحان منتشر گردید.

پرسش‌نامه نیز حاوی یک سؤال بسته در مورد تعیین سطح تاکسونومی سؤال بود؛ به طوری که سؤالات دارای تاکسونومی ۱ را به عنوان تاکسونومی پایین، از سؤالات دارای تاکسونومی ۲ و ۳ به عنوان تاکسونومی بالا، تفکیک و متمایز می‌نمود. تکمیل کنندگان پرسش‌نامه‌ها را تیمی از استادان و صاحب‌نظران در آموزش پزشکی تشکیل می‌داد که سابقه شرکت در کارگاه‌های کشوری طراحی آزمون‌ها و تدریس در مجموعه کارگاه‌های ارزشیابی دارا بودند.

جهت تعیین پایایی مطالعه، ابتدا جلسات هماهنگی بین پژوهشگران (که همگی آن‌ها کارگاه طراحی سؤال را گذرانده بودند و برخی از آنان مدرس کارگاه مزبور هستند) برگزار شد و توافق نظر اولیه در مورد نحوه بررسی سؤالات و تکمیل پرسش‌نامه ایجاد شد. به دنبال آن مجموعه سؤالات طراحی شده سال ۱۳۸۸ همراه با کلید آن‌ها استخراج و مورد بررسی قرار گرفت. با بررسی حدود ۱۰ درصد از سؤالات (۵۰ سؤال) توسط پژوهشگران، میزان همخوانی نتایج بین ارزیابی کننده‌ها (Interrater) ۰/۸۵ و در سطح (Intrater) پس از حداقل یک هفته تست مجدد ۰/۹۲ برآورد شد. هر یک از سؤالات توسط دو فرد به طور جداگانه بررسی می‌شد و موارد مبهم در جلسات مشترک مطرح می‌گردید و بین محققین اشتراک نظر لازم حاصل می‌شد؛ بدین ترتیب با پایایی قابل قبول حاصل، بررسی کلیه سؤالات مرحله قبل از مداخله یعنی سؤالات آزمون ارتقای سال ۱۳۸۸ انجام گرفت. به منظور رفع مشکل Observer bias قبل از بررسی سؤالات در مرحله قبل از مداخله، به هر یک از طراحان سؤال کد خاصی داده شد تا بررسی کنندگان از نام طراحان سؤال بی‌اطلاع باشند. در نتیجه با تجزیه و تحلیل چک‌لیست‌های تکمیل شده مربوط به تک‌تک سؤالات و مشخص شدن نقاط قوت و ضعف آزمون مزبور، سه ماه قبل از طراحی و برگزاری آزمون ارتقای سال ۱۳۸۹، کدها شکسته شد و با مشخص شدن طراحان، در جلسات جداگانه‌ای اعضای هیأت علمی هر یک از گروه‌ها به دانشکده دعوت شدند و توسط تکمیل کنندگان پرسش‌نامه

یافته‌ها

کلیه ۲۵۰۰ سؤال طراحی شده در هر دو سال ۱۳۸۸ و ۱۳۸۹ مربوط به رشته‌های داخلی، جراحی، کودکان، بیهوشی، آسیب‌شناسی، زنان و زایمان، روان‌پزشکی و پوست مورد بررسی قرار گرفت که از این تعداد، ۶۶۵ سؤال که طراحان آن‌ها فقط در طراحی سؤالات یک سال نقش داشتند، از مطالعه خارج شدند و ۲۰۳۵ سؤال باقیمانده (۹۳۸ سؤال مربوط به سال ۱۳۸۸ و ۱۰۹۷ سؤال مربوط به سال ۱۳۸۹) به عنوان جامعه آماری پذیرفته شد و تجزیه و تحلیل‌های بعدی روی آن‌ها انجام گرفت. استادان طراح که در تدوین سؤالات هر دو سال مشارکت داشتند، سابقه شرکت در کارگاه آموزشی آزمون‌های چند گزینه‌ای و تجربه کاری مربوط به طراحی سؤالات آزمون‌های ارتقای دستیاری قبلی را داشتند.

از مجموع ۹۳۸ سؤال بررسی شده مربوط به قبل از مداخله یعنی سال ۱۳۸۸، ۶۱۲ سؤال (۶۵/۲ درصد از سؤالات) در کل آیتم‌ها بدون اشکال ساختاری بود که این رقم از مجموع ۱۰۹۷

سؤال بررسی شده مربوط به بعد از مداخله یعنی سال ۱۳۸۹ به ۹۰۸ سؤال (۸۲/۸ درصد از سؤالات) رسید، به عبارت دیگر در سال قبل از مداخله اصول طراحی سؤالات در ۶۵/۲ درصد از سؤالات و در سال بعد از مداخله در ۸۲/۸ درصد از سؤالات به طور کامل رعایت شده بود که بر اساس آزمون آماری مقایسه نسبت‌ها -آزمون Z- بین سؤالات بدون اشکال در سال ۱۳۸۸ (قبل از مداخله) و ۱۳۸۹ (پس از مداخله) اختلاف آماری معنی‌داری وجود داشت ($P < 0/001$ ، جدول ۱).

از نظر رعایت اصول ساختاری در قسمت تنه سؤالات (۳ آیتم) در سال ۱۳۸۸، ۹۱۲ سؤال (۹۷/۲ درصد از سؤالات) و در سال ۱۳۸۹، ۱۰۷۳ سؤال (۹۷/۸ درصد از سؤالات) بدون اشکال بود که با استفاده از آزمون آماری مقایسه نسبت‌ها مشخص شد که بین این دو اختلاف آماری معنی‌داری وجود ندارد ($P > 0/01$ ، جدول ۱)

جدول ۱. وضعیت رعایت اصول ساختاری سؤالات قبل و بعد از ارایه بازخورد

P	سال ۱۳۸۹		سال ۱۳۸۸			
	درصد	فراوانی	درصد	فراوانی		
< 0/001	۸۲/۸	۹۰۸	۶۵/۲	۶۱۲	بدون اشکال	وضعیت کلی سؤال (کل ۱۸ آیتم)
	۱۷/۲	۱۸۹	۳۴/۸	۳۲۶	دارای اشکال	
0/240	۹۷/۸	۱۰۷۳	۹۷/۲	۹۱۲	بدون اشکال	وضعیت تنه سؤالات
	۲/۲	۲۴	۲/۸	۲۶	دارای اشکال	
< 0/001	۹۱/۸	۱۰۰۷	۷۷/۸	۷۳۰	بدون اشکال	وضعیت گزینه‌های سؤالات
	۸/۲	۹۰	۲۲/۲	۲۰۸	دارای اشکال	

این مورد نیز نشان دهنده اختلاف آماری معنی‌داری بود ($P < 0/001$ ، جدول ۱).

در بین ۱۸ آیتم مورد نظر از لحاظ رعایت اصول طراحی صحیح سؤال در مجموع سه قسمت تنه، گزینه‌ها و کلیات سؤال، بیشترین موردی که رعایت آن قبل از مداخله مد نظر

همچنین از نظر رعایت اصول ساختاری در قسمت گزینه‌های سؤالات (۹ آیتم)، ۷۳۰ سؤال (۷۷/۸ درصد از سؤالات) در سال ۱۳۸۸ و در سال ۱۳۸۹، ۱۰۰۷ سؤال (۹۱/۸ درصد از سؤالات) بدون اشکال ساختاری بود. آزمون Z در

سؤال بود که تنها در ۸۳۰ سؤال (۸۸/۵ درصد از سؤالات) رعایت شده بود و این میزان بعد از مداخله به ۱۰۳۰ مورد (۹۳/۹ درصد از سؤالات) رسیده است (جدول ۲).

قرار گرفت (در ۹۳۵ سؤال یا ۹۹/۷ درصد از سؤالات)، عدم استفاده از علائم اختصاری با ظاهری مشترک بوده است که پس از مداخله رعایت آن در تمامی سؤالات (۱۰۹۷ سؤال) مورد عنایت قرار گرفته است. همچنین کمترین موردی که قبل از مداخله مورد توجه قرار گرفت، صحیح بودن شکل ظاهری

جدول ۲. توزیع و درصد فراوانی میزان رعایت اصول طراحی ساختاری سؤالات قبل و بعد از مداخله در سال ۱۳۸۸ و ۱۳۸۹

سال ۱۳۸۹ (از ۱۰۹۷ سؤال)		سال ۱۳۸۸ (از ۹۳۸ سؤال)		وضعیت اصول کلی طراحی
درصد	فراوانی	درصد	فراوانی	
وضعیت اصول کلی طراحی				
۹۳/۹	۱۰۳۰	۸۸/۵	۸۳۰	۱- صحیح بودن شکل ظاهری سؤال
۹۹/۵	۱۰۹۲	۹۹/۶	۹۳۴	۲- اجتناب از آیشاری بودن سؤالات
۹۹/۶	۱۰۹۳	۹۸/۴	۹۲۳	۳- خودداری از به کارگیری اشاره‌های نامربوط
۹۸/۵	۱۰۸۰	۹۴/۷	۸۸۸	۴- سنجش یک موضوع واحد
۱۰۰	۱۰۹۷	۹۹/۷	۹۳۵	۵- عدم استفاده از علائم اختصاری با ظاهر مشترک و متفاوت در مفهوم
۹۹/۲	۱۰۸۸	۹۸/۵	۹۲۴	۶- پرهیز از به کارگیری منفی مضاعف
تنه سؤال				
۱۰۰	۱۰۹۷	۹۸/۷	۹۲۶	۷- آوردن موضوع اصلی سؤال در قسمت تنه
۱۰۰	۱۰۹۷	۹۹/۵	۹۳۳	۸- پرسشی بودن تنه سؤال و یا درج جای خالی در انتهای تنه (در موارد غیر پرسشی)
۹۷/۸	۱۰۷۳	۹۸/۸	۹۲۷	۹- متمایز نمودن عبارات منفی
گزینه‌های سؤال				
۹۹/۵	۱۰۹۲	۹۹/۵	۹۳۳	۱۰- مناسب بودن شکل و چیدمان ظاهری گزینه‌ها در ذیل تنه
۹۹/۳	۱۰۸۹	۹۸	۹۱۹	۱۱- منطقی بودن ترتیب ظاهری در گزینه‌های عددی و رتبه‌ای
۹۷/۱	۱۰۶۵	۹۲/۴	۸۶۷	۱۲- یکسان بودن زبان مورد استفاده در گزینه‌ها
۹۹/۷	۱۰۹۴	۹۸/۳	۹۲۲	۱۳- پرهیز از تکرار مطالب در گزینه‌ها
۹۹/۹	۱۰۹۶	۹۷/۷	۹۱۶	۱۴- پرهیز از توضیح اضافی در گزینه‌ها
۹۹/۷	۱۰۹۴	۹۸/۱	۹۲۰	۱۵- اجتناب از به کارگیری دو گزینه متضاد، که یکی صحیح و دیگری غلط باشد
۹۸/۶	۱۰۸۲	۹۷/۸	۹۱۷	۱۶- پرهیز از به کارگیری عباراتی همچون همه موارد، الف و ب، هیچ کدام
۹۸/۲	۱۰۷۷	۹۵/۷	۸۹۸	۱۷- یکسانی نسبی طول گزینه صحیح با سایر گزینه‌ها
۹۹/۴	۱۰۹۱	۹۴/۸	۸۸۹	۱۸- برخورداری گزینه‌های انحرافی از ظاهر صحیح و منطقی

سال ۱۳۸۸ (قبل از مداخله) و ۱۳۸۹ (پس از مداخله) اختلاف آماری معنی‌داری وجود داشت ($P < 0/001$)، جدول ۳).

از نظر تاکسونومی سؤالات، ۳۵۶ سؤال (۳۸ درصد از سؤالات) در سال ۱۳۸۸ و ۵۸۲ سؤال (۵۳/۱ درصد از سؤالات) در سال ۱۳۸۹ با تاکسونومی بالا (۲ و ۳) طراحی شده بود که با استفاده از آزمون آماری مقایسه نسبت‌ها (آزمون Z) مشخص شد و بین تعداد سؤالات با تاکسونومی بالا در

انجام شده بود، ۶۰/۲۸ درصد از سؤالات در سال ۱۳۸۹ بدون اشکال ساختاری بود که در مقایسه با سال قبل از آن (۶۰/۶) درصد) از نظر آماری تفاوت معنی‌داری نداشت (۱۰) که این عدم تفاوت آماری می‌تواند به دلیل عدم مداخله آموزشی در گروه هدف و به عبارتی، عدم تلاش در جهت ارتقای کیفیت این آزمون‌ها باشد.

نتایج مطالعه کنونی نشان داد که ارایه بازخورد در زمینه طراحی سؤال مؤثر است که شاید دلیل آن رعایت صحیح اصول علمی ارایه بازخورد یعنی عینی، مستقیم، محترمانه، بدون قضاوت از فرد، محدود به عملکرد خاص، در زمان مناسب، همراه با توصیه‌هایی برای بهبود طراحی سؤال مربوط باشد. دیگر مطالعات انجام شده نشان دهنده وجود اشکال ساختاری در حدود نیمی از سؤالات گروه‌های مشابه بوده است (۱۵، ۱۴، ۶). برای حل این مشکل به نظر می‌رسد که نیازمند مداخلاتی به طور ترجیحی آموزشی هستیم.

همچنین این مطالعه نشان داد که با انجام مداخله آموزشی از نوع ارایه بازخورد جهت آن دسته از استادان طراحی که از سابقه کاری بیشتری برخوردار هستند و سابقه شرکت در کارگاه‌های آموزشی آزمون‌های چند گزینه‌ای و مطالعات مرتبط در این زمینه را دارا می‌باشند، با تأکید مستقیم بر نقاط ضعف و قوت طراحی سؤالات و ارایه راه‌کارهای ارتقای کیفیت طراحی سؤالات (بدون پرداختن به مقدمات و حاشیه‌ها) می‌توان اثربخشی بهتری را به دست آورد. در همین راستا معیاری و همکاران در مطالعه خود نشان دادند که برگزاری کارگاه آموزشی جهت طراحان آزمون‌های ارتقای دستیاری دندان‌پزشکی که اغلب استادانی جوان بوده‌اند و کارگاه مربوط را طی نکرده بودند، می‌تواند به ارتقای معنی‌دار کیفیت طراحی این گونه سؤالات منجر شود (۸). به عبارت دیگر، در این مطالعه مشخص شد که با وجود آن که طراحان سؤال همگی سابقه حضور در کارگاه‌های آموزشی آزمون‌های چند گزینه‌ای و نیز طراحی سؤال در آزمون‌های ارتقای دستیاری سال‌های قبل را دارا بودند، همچنان در طراحی سؤال دارای ایراداتی

جدول ۳. سطح تاکسونومی سؤالات قبل و بعد از ارایه بازخورد

P	۱۳۸۹		۱۳۸۸		
	درصد	فراوانی	درصد	فراوانی	
۰/۰۰۱ <	۵۳/۱	۵۸۲	۳۸	۳۵۶	تاکسونوم ی بالا
	۴۶/۹	۵۱۵	۶۲	۵۸۲	تاکسونوم ی پایین
	۱۰۰	۱۰۹۷	۱۰۰	۹۳۸	جمع

بحث و نتیجه‌گیری

با وجود نگرانی‌هایی که در مورد تناسب کاربرد سؤالات چند گزینه‌ای در آزمون‌های ارتقای دستیاری وجود داشته است، ارزش به کارگیری این گونه سؤالات به خصوص برای سنجش دانش در این آزمون‌ها در مطالعات مختلف از جمله مطالعه Downing (۱۹) و نیز Smith و همکاران (۲۰) مورد تأکید قرار گرفت. از سوی دیگر با توجه به آن که آزمون‌هایی نظیر آزمون دستیاری که در قبال سلامت جامعه پاسخگو است، از اهمیت خاصی برخوردار است، طراحی صحیح و درخور این نوع آزمون‌ها همراه با تجزیه و تحلیل آن‌ها با برنامه‌ریزی برای تقویت نقاط قوت و رفع نقاط ضعف و برگزاری دوره‌ها، کارگاه‌های آموزشی و جلسات توجیهی جهت طراحان سؤال ضروری می‌باشد. در این مطالعه با مداخله و ارایه بازخورد به طراحان سؤال، سعی شد تا تأثیر آن بر دو جنبه مهم از عواملی که در کیفیت سؤالات چند گزینه‌ای نقش تعیین کننده‌ای دارند، یعنی اول این که رعایت اصول ساختاری طراحی سؤال و دوم این که سطح دانش مورد ارزیابی در سؤالات (تاکسونومی) مورد بررسی قرار گیرد.

بر اساس نتایج این مطالعه پس از ارایه بازخورد، درصد سؤالاتی که بر اساس رعایت اصول ساختاری طراحی شده بود، به طور معنی‌داری افزایش یافت، در صورتی که بر اساس گزارش رسمی منتشر شده از سوی وزارت متبوع، در تحلیل مقایسه‌ای ساختار و نتایج آزمون‌های دانش‌نامه تخصصی پزشکی دوره ۵۷ که در کل کشور (در سطح وزارت متبوع)

نشان داده است که به دلیل اهمیت بیشتر یادگیری عمیق‌تر، در سال‌های اخیر سهم بیشتری از سؤالات به سطوح عمیق‌تر دانش (تاکسونومی ۲ و ۳) اختصاص یافته‌اند.

نتیجه‌گیری

با توجه به نتایج این مطالعه، جهت آن دسته از طراحان سؤالات چند گزینه‌ای که سابقه شرکت در کارگاه‌های آزمون‌های چند گزینه‌ای را داشته‌اند و از مطالعات و تجارب قبلی در این زمینه برخوردار بودند، ارایه بازخورد به شیوه‌ای فنی و متناسب با اصول آموزش پزشکی می‌تواند به عنوان روشی مؤثر در ارتقای کیفیت طراحی آزمون‌های دستیاری و همچنین سایر مقاطع و رشته‌های علوم پزشکی قابل به کارگیری باشد. به عبارت دیگر، در کنار برگزاری کارگاه‌های آموزشی طراحی سؤالات آزمون‌های چند گزینه‌ای نیاز به تجزیه و تحلیل آزمون‌های طراحی شده توسط استادان وجود دارد و به کارگیری روش‌های مؤثر بعدی همچون ارایه بازخورد جهت ارتقای هر چه بیشتر کیفیت طراحی سؤالات راه‌گشا خواهد بود.

سپاسگزاری

این پژوهش با حمایت مالی معاونت تحقیقات و فناوری دانشگاه علوم پزشکی همدان انجام شده است. بدین‌وسیله از اعضای محترم هیأت علمی گروه‌های بالینی دارای دستیار که در انجام طرح با ما همکاری نمودند، تشکر و قدردانی می‌گردد.

References:

1. Bazargan A. Educational assessment. 5th ed. Tehran: Samt; 2007. [In Persian]
2. Gharib M, Khakbazan Z, et al. Educational Process in Medicine. 1th ed. Tehran: Tehran University of Medical Sciences; 2005. [In Persian]
3. SabooriKashani A, Mohamadi M, Zeinaloo A, Razavi M. Standardization of student

بودند که در این شرایط انجام مداخله به صورت ارایه بازخورد در حد قابل توجهی توانسته است تا در رفع اشکالات مؤثر باشد. در همین زمینه جهت استادان دارای تجربه طراحی سؤال و سابقه شرکت در کارگاه‌های آموزشی چند گزینه‌ای، انجام مداخله به صورت بازخورد و با ارایه نمونه کارهای قبلی خودشان در زمانی مناسب، به صورتی مستقیم، عینی، به موقع، محترمانه و به دور از قضاوت از عملکردشان و همچنین همراه با ارایه توصیه‌هایی جهت رفع ایرادات و تقویت نقاط قوت طراحی تک‌تک سؤال‌های خود می‌تواند اثربخش باشند و به ارتقای کیفیت طراحی سؤالات توسط آن‌ها بیانجامد.

از این رو موارد فوق می‌تواند مؤید لزوم انجام مداخله آموزشی متناسب با شرایط و نیازهای هر گروه باشد. این در حالی است که تاکسونومی آزمون برگزار شده در سال‌های ۱۳۸۸ و ۱۳۸۹ در سایر دانشگاه‌ها طبق گزارشی که دبیرخانه شورای آموزش پزشکی و تخصصی وزارت بهداشت مبنی بر تحلیل نتایج و ساختار آزمون دانش‌نامه تخصصی دوره ۵۷ منتشر نمود، ۵۶٪ درصد از سؤالات در سال ۱۳۸۹ در تاکسونومی‌های ۲ و ۳ طراحی شده است که در مقایسه با سال گذشته آن (۵۶ درصد)، تفاوت آماری معنی‌داری را نشان نمی‌دهد (۱۰) که این نیز می‌تواند به دلیل عدم مداخله آموزشی مناسب در گروه هدف در این مدت باشد.

در این زمینه مطالعات خارجی، در حالی که شواهد قدیمی‌تر مانند مطالعه Huxham و Naeraa یافته‌های مشابه مبنی بر غلبه کاربرد تاکسونومی ۱ در آزمون‌های پیشرفت تحصیلی را نشان دادند (۲۱)، اما مطالعات جدیدتر مانند مطالعه McCourbie (۲۲) و Tarrant و همکاران (۲۳)

evaluation: Two testing theories. 1th ed. Tehran: Tehran sadra; 2004. [In Persian]

4. Zolfaghari B, Adibi N, Darakhshanfar S, Tansaz M. Academic achievement tests in medical sciences. 1st ed: Isfahan: Medical University Educational Development center Publisher; 2001. [In Persian]
5. Vaezi AA, Hosseini M, Aghily H, Mirjalili MR. The investigation of applying multiple-choice question making rules in final exams

- of the faculty members of ShahidSadoughi Yazd University of Medical Sciences. The Journal of Medical Education & Development Center of Yazd. 2007; 2(1)31-38. [In Persian]
6. Haghshenas M, Vahidshahi K. Evaluation of multiple choice questions in the school of medicine, Mazandaran University of medical sciences, the first semester of 2007. Strides in Development of Medical Education 2008; 5(2):120-27. [In Persian]
 7. Bighlarkhani M, Meyari A, Falah M, Zamanian A. The study of Multiple Choice Questions designed in residency annual exams in Hamedan University Medical Sciences 2007&2008. Final Report of research project 10856, Hamedan university of Medical sciences. 2009. Available From: <http://dme.umsha.ac.ir/uploads/MCQ%20Term...%20pezesh1.7cross-sectional.pdf> [In Persian]
 8. Meyari A, Beiglarkhani M. The effect of education on improvement of multiple choice questions' designing in annual residency exams of dental faculty. Iranian Journal of Medical Education 2012; 12(1):36-45. [In Persian]
 9. Vahidshahi K, Mohagheghi MA. Comparison of some quality criteria of multiple choice questions in residency annual exam 2007 & 2008. Available From: http://cgme.behdasht.gov.ir/uploads/264_781_N4_Moghayese.html
 10. Ministry of health, Treatment and Medical Education. Iranian council for graduate medical education.[Analysis of structure &results of Residency annual exam 57;2010. Available From:http://cgme.behdasht.gov.ir/uploads/264_781_N10_TahlilBord.pdf [In Persian]
 11. Millman J. The new handbook of teacher evaluation. Assessing elementary and secondary school teachers. New York: Sage publication;1981.
 12. Saif AA. Educational Measurement, Assessment and Evaluation. Fourth edition, 25th revision. Tehran: Nashredoran; 2007. [In Persian]
 13. Kaveh M, Bahreini M. Analytical assessment of multiple-choice tests at Mashhad Medical School. Medical Journal of Mashhad University of Medical Sciences 2002; 45(76):89-95. [In Persian]
 14. Shakurnia A, KhosraviBorojeni A, Mozaffari A, Elhampour H. An evaluation of exam questions designed by faculty members, emphasizing on the multiple choice question structure, Ahvaz Jundishapur University of Medical Sciences, 2007. Strides in Development of Medical Education 2010; 6(2):129-138. [In Persian]
 15. Arab M. The study of Multiple Choice Questions flaws in Hamedan medical school 1999. Journal of Tehran University of Medical Sciences: Abstract of 4th National Congress of Medical Education 2000: 51. [In Persian]
 16. Doripour F, KarimiSh, Pakaein J. Evaluation of MCQ'S of nursing students' health courses 2006. Strides in Development of Medical Education. Abstract of 8th National Congress of Medical Education 2006. [In Persian]
 17. Rasoulinejad A, Vakili Z. Comparison of taxonomy of Kashan medical University's residency exams. Strides in Development of Medical Education. Abstract of 8th National Congress of Medical Education 2006. [In Persian]
 18. Ende J. Feedback in Clinical Medical Education. The Journal of the American Medical Association 1983; 250(6):777-81.
 19. Downing SM. Assessment of knowledge with written test formas. International Handbook of research in medical education; 2: 647-72.
 20. Smith JD, Prideaux D. Developing the accredited postgraduate assessment program for Fellowship of the Australian College of Rural and Remote Medicine.Rural and Remote Health 2007; 7(4):805.
 21. Huxham GJ, Naeraa N. Is Bloom's taxonomy reflected in the response pattern to MCQ items? Medical Education 1980; 14(1):23-6.
 22. McCourbie P. Improving the fairness of multiple choice questions: A literature review. Medical Teacher 2004; 26(8):709-12.
 23. 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 Education in Practice 2006; 6(6):354-63.

Improvement of Design of Multiple Choice Questions in Annual Residency Exams by Giving Feedback

Azam Meyari¹, Mahdi Beiglarkhani^{2*}

1.General Practitioner, M.Sc. in Medical Education, Medical Education Development Center, Hamedan University of Medical Sciences, Hamedan, Iran

2.General Practitioner, M.Sc. in Medical Education, Instructor, Department of General Education, School of Medicine, Hamedan University of Medical Sciences, Hamedan, Iran

• Received: 9 Oct, 2012

• Received Corrected Version: 2 Oct, 2012

• Accepted: 9 Jul, 2012

Background & Objective: The aim of this study was to survey the effect of presenting feedback on improvement of multiple choice questions (MCQ) design in annual residency exams.

Methods: In this semi-experimental study, the structure and taxonomy in MCQs of annual residency exams in 2009 were analyzed through Millman's checklist. Based on this data, an interventional feedback on strengths and weaknesses was presented to question designers who had experience of attending MCQ workshops and designing questions in the past. The principles of giving technical feedback in medical education such as objectivity, right timing, and being direct, respectful and without judgment were considered. After the annual residency exam in 2010, the questions of designers in both years were compared in respect to Millman's principals and taxonomy before and after the intervention.

Results: From 2500 questions, 2035 questions were entered into the study. 65.2% of the questions developed in 2009 had no structural flaw; this increased to 82.8% after the intervention, which was a significant difference ($P < 0.001$). Moreover, there was a significant increase in the percentage of questions with a high taxonomy; from 38% in 2009 to 53.1% in 2010 ($P < 0.001$).

Conclusion: Giving interventional feedback, based on its techniques in medical education, to experienced designers of MCQs can affect the improvement of designing quality of questions.

Key Words: Multiple choice questions, Residency, Medical education, Taxonomy, Millman's principles, Feedback

*Correspondence: School of Medicine , Hamedan University of Medical Sciences, Shahid Fahmide Blvd., Hamedan, Iran

• Tel: (+98) 811 838 0246

• Fax: (+98) 811 838 0238

• Email: dmbiglar@yahoo.com