بررسی شاخص‌های کمی و کیفی آزمون‌های مقطع کارآموزی دوره پزشکی عمومی

محمد گرک یراقی ، مریم آویژگان *، امراله ابراهیمی، الهام اسفندیاری، آتوسا اسماعیلی، شهرام شایان، وحید عشوریون، نیکو یمانی

چکیده

مقدمه: ارزشیابی پیشرفت تحصیلی بخش مهمی از فرآیند آموزشی است و برای تحقق آن از آزمون استفاده می‌شود. استفاده از سؤالات چندگزینه‌ای به عنوان یک روش آزمون عینی، متداول‌ترین روش ارزشیابی دانشجویان می‌باشد و طراحی درست آنها واجد اهمیت است. مطالعه حاضر به بررسی کمی و کیفی آزمون‌های مقطع کارآموزی دوره پزشکی عمومی در گروه‌های آموزشی پوست، روانپزشکی، زنان، چشم و نورولوژی دانشکده پزشکی دانشگاه علوم پزشکی اصفهان پرداخته است.

روش‌ها: مطالعه توصیفی- مقطعی بر روی آخرین آزمونهای مقطع کارآموزی دوره پزشکی عمومی در گروه‌های آموزشی پوست، روانپزشکی، زنان، چشم و نورولوژی در سال ۱۳۸۷ اجرا شد. شاخص‌های کمی و کیفی ۱۵ آزمون برگزار شده با استفاده از چک‌لیست‌هایی روا و پایا که برای تعیین سؤالات با ضریب تمیز نامناسب، سؤالات دشوار، سؤالات با انتخاب سه گزینه صفر، ضریب پایایی، توزیع سؤالات نسبت به سرفصل‌های اعلام شده، حیطه سؤالات بر اساس هدف‌ها، انواع سؤال بر اساس تاکسونومی بود، استخراج گردید. اطلاعات با نرم‌افزار SPSS و آمار توصیفی تجزیه و تحلیل شد.

نتایج: از نظر شاخص‌های کمی، میانگین سؤالات با ضریب تمیز نامناسب ۲۴ درصد، سؤالات دشوار ۴/۱ درصد و سؤالات با انتخاب سه گزینه صفر ۱۲/۹ درصد به دست آمد. میانگین توزیع سؤالات نسبت به سرفصل‌های اعلام شده و ضریب پایایی به ترتیب ۹۸ درصد و ۷۷/۹۲ درصد بود. از نظر شاخص‌های کیفی، بیشترین فراوانی نسبی سؤالات در حیطه علایم و نشانه‌ها (۳۰/۷۵ درصد) و تاکسونومی اول (۵۹ درصد) قرار داشت.

نتیجه‌گیری: با توجه به اهمیت ارائه سؤالات استاندارد و منطبق با قواعد طراحی سؤالات چندگزینه‌ای در برگزاری آزمون‌های مناسب، بازبینی مستمر سؤالات امتحانی و برگزاری دوره‌های آموزشی طراحی سؤالات امتحانی، ایجاد بانک سؤالات استاندارد و تأکید بر تهیه جدول دو بعدی آزمون‌ها قبل از برگزاری آزمون در ارتقای کیفیت سؤالات تاثیر دارد.

واژه‌های کلیدی: آزمون، سؤال چندگزینه‌ای، کارآموزی، ارزشیابی دانشجو

مجله ایرانی آموزش در علوم پزشکی (ویژه‌نامه توسعه آموزش) / زمستان ۱۳۸۹؛ ۱۰(۵): ۵۳۳ تا ۵۴۲

مقدمه

ارزشیابی پیشرفت تحصیلی دانشجویان یکی از اهداف عمده دانشگاه‌ها و مراکز آموزش عالی کشور است زیرا

* **نویسنده مسؤول:** مریم آویژگان، دانشجوی دکترای برنامه ریزی درسی، دانشکده علوم تربیتی و روانشناسی دانشگاه اصفهان؛کارشناس دفتر توسعه آموزش دانشکده پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران. avizhgan@yahoo.com

دکتر محمد گرک یراقی (دانشیار) گروه داخلی،قلب، دانشگاه علوم پزشکی اصفهان، اصفهان ،ایران. (garakyaraghi@med.mui.ac.ir)؛ امراله ابراهیمی، پزشک عمومی، کارشناس دفتر توسعه آموزش، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران. (ev_11@yahoo.com)؛ الهام اسفندیاری، کارشناس ارشد میکروبیولوژی، محقق گروه ژنتیک، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران. (elham_esfandiari@yahoo.com)؛ آتوسا اسماعیلی، پزشک عمومی، کارشناس دفتر توسعه آموزش، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران. (eatousa@yahoo.com)؛ شهرام شایان- دانشجوی دکترای مدیریت آموزش عالی و کارشناس مرکز تحقیقات آموزش پزشکی- دانشگاه علوم پزشکی اصفهان - shayan@edc.mui.ac.ir

وحید عشوریون(مربی)، مرکز تحقیقات آموزش پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران. (ashourioun@med.mui.ac.ir)؛ نیکو یمانی (استادیار)، مرکز تحقیقات آموزش پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران. (yamani@edc.mui.ac.ir)

این مقاله در تاریخ ۸۹/۱۱/۲ به دفتر مجله رسیده، در تاریخ ۸۹/۱۱/۲۷اصلاح شده و در تاریخ ۸۹/۱۲/۱ پذیرش گردیده است.

[Downloaded from ijme.mui.ac.ir on 2026-11-15]

بدین وسیله میزان یادگیری دانشجویان و در نهایت میزان دستیابی به اهداف آموزشی اندازه‌گیری می‌شود(۱). یک ارزشیابی مؤثر نه تنها در تمایز دانشجویان نقش به سزایی دارد بلکه به ارزیابی فعالیت‌های مدرس نیز کمک می‌کند زیرا که اگر نتوانیم چیزی را مورد سنجش قرار دهیم قادر به بهبود آن نخواهیم بود(۲). ارزشیابی یکی از جنبه‌های مهم در فرایندهای آموزشی بوده و این امکان را فراهم می‌سازد تا بر اساس نتایج آن نقاط قوت و ضعف آموزش شناسایی شده و با تقویت جنبه‌های مثبت و رفع نارسایی‌ها در ایجاد تحول و اصلاح نظام آموزشی گام‌های مناسبی برداشته شود(۳و٤). ارزشیابی مؤثر نه تنها باعث افزایش انگیزه در دانشجویان شده بلکه مدرس را نیز در ارزیابی فعالیت‌های خود کمک می‌کند(٥و٦).

مهم‌ترین وسیله‌اندازه‌گیری پیشرفت تحصیلی، آزمون است.آزمون‌هایی که عملکرد تحصیلی دانشجویان را اندازه می‌گیرند، به انشایی و عینی تقسیم می‌شوند. در خصوص این تصمیم که از کدام یک از این آزمون‌ها برای اندازه‌گیری عملکرد تحصیلی دانشجویان استفاده شود بین اساتید، توافق نظر وجود ندارد(۷). نحوه ارزیابی متداول برای دانشجویان پزشکی طرح سؤالات چند گزینه‌ای است(۸). آزمون‌های چند گزینه‌ای که به آزمون‌های تستی شهرت دارند یکی از انوع آزمون عینی هستند که اخیراً به دلیل محاسبه ویژگی‌های روان سنجی آن مورد توجه متخصصان آزمون‌سازی قرار گرفته‌اند. این نوع آزمون‌ها اگر با دقت و مهارت کافی ساخته شوند برای اندازه‌گیری تمام اهداف درس در تمام سطوح فکری و در تمام رشته‌های تحصیلی مناسب هستند و اکثر آزمون‌های استاندارد که در سطح بین‌المللی که برای گزینش دانشجو در سطوح مختلف تحصیلی انجام می‌پذیرد، از نوع چند گزینه‌ای می‌باشند(۹و۱۰) ولی بهترین شرایط استفاده از این روش همراه نمودن آن با روش‌های دیگر ارزیابی است(۱۱).

استفاده مؤثر از سؤال، به عنوان یک ترفند آموزشی، می‌تواند دانشجو را در بکارگیری اطلاعات در موقعیت‌های جدید و منحصر به فرد یاری دهد. سؤال کردن مؤثر به معنی طرح سؤالاتی چالش برانگیز و وادارکننده دانشجو به تجزیه و تحلیل، ارزیابی، قضاوت و تصمیم‌گیری است(۱۲). شواهد نشان می‌دهد فراگیران در طی دوره تحصیل یاد می‌گیرند که همه فعالیت‌های کلاس ارزش یکسانی ندارند، بنابراین، تنها آنچه را فرا می‌گیرند که در امتحانات عامل کسب موفقیت است(۱۳). چگونگی طرح این سؤالات در برگیرنده اهداف شناختی آموزشی عنوان شده در طرح درس مربوطه باشد که این اهداف به طور مطمئن قابل مشاهده و اندازه‌گیری است(١٤).

در بسیاری از موارد کلیه دانشجویان ضعیف و قوی صرف‌نظر از میزان آمادگی، نمرات بالایی در امتحان کسب می‌نمایند و یا از گذراندن موفقیت آمیز آن باز می‌مانند(١٥). این واقعیت که دانشجویان بتوانند سؤالات یک مبحث را به درستی پاسخ دهند، تضمینی بر این نیست که از دانش کافی در آن موضوع برخوردارند. آزمون‌هایی که فقط بر پایه‌ی یادآوری مطالب (تاکسونومی ۱) می‌باشند، از پایایی کمی برای سنجش فراگیری واقعی برخوردارند(١٦). اگر چه با استفاده از سؤالات چند گزینه‌ای می‌توان محتوای زیادی را مورد ارزیابی قرار داد، اما در صورتی که اصول صحیح طراحی سؤال رعایت نشود، از اعتبار این نوع سؤالات کاسته می‌گردد(١٧).

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 مسائل و مشکلاتی وجود دارد و بسیاری از مدرسان بدون اینکه آموزش کافی در سنجش و اندازه‌گیری و شیوه ساخت آزمون‌ها کسب کنند، آماده تدریس می‌شوند(۱۸و۱۹و۲۰). نتایج مطالعه دیگری در دانشگاه علوم پزشکی مشهد نیز

[Downloaded from ijme.mui.ac.ir on 2026-11-15]

نشان داد که روش‌های ارزشیابی موجود نه تنها در سنجش مهارت‌های بالینی، بلکه در سنجش دانش و محفوظات نیز از کفایت لازم برخوردار نبوده‌اند و لزوم تدارک دوره‌های آموزشی مناسب در زمینه طراحی سؤالات امتحانی برای اعضای هیأت‌علمی را ضروری مطرح نموده است(۲۰).

از زاویه ای دیگر انجام برخی مداخلات آموزشی باعث افزایش کیفیت آزمونها شده است(۲۱و۲۲و۲۳). پژوهشی در دانشکده پزشکی شاهرود نشان داد انعکاس نتایج تحلیل آزمون‌های قبلی دقت طراحان سؤال در تهیه سؤالات بعدی را افزایش داده است(۲٤). تحقیق دیگری در مقطع ICM دانشکده پزشکی اصفهان نشان داده که صد در صد اساتید بالینی بر این باورند که آگاهی از نتایج ارزشیابی سؤالات نقش مثبتی در بهبود کیفیت این آزمون‌ها دارد و با اینکه ۷۷درصد آنان در کارگاه اصول طراحی سؤال شرکت کرده بودند معتقد بودند نظارت کارشناسان خطاها را کاهش خواهد داد(۱٤).

آزمون چند گزینه‌ای در مقطع بالینی از شایع‌ترین موارد ارزشیابی و هنوز هم هسته مرکزی ارزشیابی دانشجویان پزشکی است. ولی متأسفانه با توجه به کیفیت پایین اغلب این آزمون‌ها باید تدابیری جهت ارتقای کیفیت آنها انجام داد زیرا این امر رابطه مستقیم و غیرمستقیم با اصلاح نظام ارزشیابی و سیستم آموزشی دارد. با توجه به مشکلات موجود در زمینه طرح سؤال و با عنایت به استانداردهای پزشکی عمومی مصوب شورای عالی برنامه‌ریزی پزشکی کشور و به تبع آن مصوبه‌ی شورای آموزشی – پژوهشی دانشکده مبنی بر تحلیل آزمون‌ها لذا این مطالعه به بررسی کمی و کیفی سؤالات آزمون‌های چندگزینه‌ای مقطع کارآموزی در گروه‌های آموزشی پوست، روانپزشکی، زنان، چشم و نورولوژی دانشکده پزشکی اصفهان پرداخته تا قدمی در جهت ارتقای کیفیت ارزشیابی دانشجو در سیستم آموزش بالینی دانشکده پزشکی اصفهان برداشته شود.

روش‌ها

مطالعه توصیفی مقطعی بر روی ۱۵ آزمون کارآموزی دانشکده پزشکی اصفهان در سال تحصیلی ۱۳۸۷ اجرا گردید. جامعه آماری شامل ۳ آزمون آخر گروه‌های آموزشی پوست، چشم، روانپزشکی، نورولوژی و زنان شامل ۷٦۵ سؤال دروس تخصصی دوره کارآموزی بود. معیار ورود داده‌ها به پژوهش کلیه آزمون‌های چند گزینه‌ای برگزار شده در فرجه زمانی تعیین شده و معیار خروج سؤالات از پژوهش نیز سؤالات تشریحی، Matching، سؤالاتی که توسط گروه حذف گردیده و یا سؤالاتی که در آنالیز کیفی قابل طبقه‌بندی در یک حیطه نبود.

پروژه شامل دو مرحله بود. مرحله اول شامل آنالیز کمی ۳ دوره آزمونهای قبل از مطالعه بود. کلیه سؤالات آزمون‌ها یک روز پس از برگزاری توسط گروه آموزشی مربوطه تحویل دفتر توسعه آموزش دانشکده می‌شد و آنالیز ماشینی اولیه شامل منحنی‌ها و جداول درصد پاسخگویی دانشجویان به سؤالات، ضریب تمیز و ضریب پایایی، یک هفته بعد از آن توسط اداره خدمات ماشینی حوزه معاونت آموزشی دانشگاه اخذ می‌گردید و بلافاصله تجزیه و تحلیل توسط کارشناسان آزمون و با کمک رابط آموزشی هر گروه در زمینه شاخص‌های کمی انجام می‌گرفت. در مرحله دوم شاخص‌های کیفی شامل سطح تاکسونومی سؤالات بر مبنای طبقه‌بندی بلوم توسط گروه افراد خبره در حیطه سؤالات طبق معیارهای تعریف شده با استفاده از جدول دو بعدی مورد بررسی قرار گرفت.

ابزار جمع‌آوری اطلاعات، سه چک‌لیست محقق ساخته بود که اولی شامل مشخصات سؤالات شامل ضریب تمیز، ضریب دشواری، سؤالات با انتخاب سه گزینه صفر و ضریب پایایی بود که با کمک اطلاعات مرکز خدمات ماشینی دانشگاه علوم پزشکی اصفهان با استفاده از نرم‌افزار تدوین شده (توسط صبوری و همکاران)

[Downloaded from ijme.mui.ac.ir on 2026-11-15]

محاسبه شد. دومین چک‌لیست شامل دو قسمت بود: قسمت اول مربوط به حیطه سؤالات شامل ۱- تعاریف و طبقه بندی‌ها ۲-پاتوفیزیولوژی بیماری ۳-علایم و نشانه‌های بیماری ٤- پاراکلینیک و نحوه تشخیص بیماری ٥- درمان بیماری و قسمت دوم انواع سؤال بر اساس تاکسونومی سؤالات بود. برای ساده تر کردن دسته بندی سؤالات چندگزینه‌ای در سطوح شناختی و نتیجه‌گیری بهتر، بر اساس طبقه‌بندی بلوم، سطوح شناختی سؤالات به سه دسته شامل تاکسونومی I (دانشی یا حفظ کردنی)، تاکسونومی II (فهمیدنی) و تاکسونومی III (به کار بستنی) تقسیم گردید. سومین چک‌لیست برای تعیین میزان تناسب موضوعات تئوری با تعداد و درصد سؤالات مطرح شده در آزمون یعنی توزیع سؤالات نسبت به سرفصل‌های اعلام شده دوره اختصاص یافت برای بررسی روایی چک‌لیست در رسیدن به اهداف پژوهش در اختیار سه تن از کارشناسان آزمون گذاشته شد و پس از اصلاح موارد پیشنهاد شده، قابلیت اعتماد آن مورد تأیید قرار گرفت. پایایی پرسشنامه با استفاده از مطالعه راهنما برروی پنج آزمون ارزیابی شد به این نحو که در فاصله دو هفته پرسشنامه به وسیله کارشناس آزمون و دو نفر هیات علمی درس تکمیل و ضریب همبستگی پاسخ‌های مربوطه اندازه‌گیری گردید و با ضریب همبستگی ۰/۸۵ پایایی نظرات کارشناسان طرح در ارزیابی سؤالات امتحانات مورد تأیید قرار گرفت. برای تجزیه و تحلیل داده‌ها از نرم‌افزار SPSS-11 استفاده شد. برای رعایت اصول اخلاقی در این تحقیق، سؤالات و با اجازه اساتید جمع‌آوری و مورد تجزیه و تحلیل قرار گرفت و برای حفظ امانت، نتایج تک تک دروس با نام گزارش نشد.

نتایج

نتایج نشان داد در کل، میانگین ضریب تمیز نامناسب ۲٤ درصد، سؤالات دشوار۱/ ٤درصد، سؤالات با انتخاب سه گزینه صفر ۱۲/۹ درصد و ضریب پایایی ۷۷/۹ درصد و توزیع سؤالات نسبت به سرفصل‌های اعلام شده با استفاده از جدول دو بعدی ۹۸ درصد بود. جدول یک توزیع فراوانی نسبی سؤالات با ضریب تمیز نامناسب، ضریب دشواری، طراحی گزینه‌های انحرافی توسط اساتید به طور غیر مستقیم، ضریب پایایی آزمون‌ها و سؤالات طرح شده بر مبنای سرفصل‌های اعلام شده بین گروه‌های مربوطه بر اساس درصد پرداخته است.

نتایج به دست آمده در مورد حیطه سؤالات شامل ۱٥/۳۲ درصد تعاریف و طبقه‌بندی‌ها، ۱۸/۷٥ درصد پاتوفیزیولوژی بیماری، ۳۰/۷٥ درصد علایم و نشانه‌های بیماری، ۲۰ درصد پاراکلینیک و نحوه تشخیص بیماری و ۱٤/٦۷ درصد درمان بیماری می‌باشد. جدول دو به میانگین حیطه سؤالات در گروه‌های مربوطه بر اساس درصد پرداخته است. از نظر تاکسونومی به ترتیب ٥۹، ۳٤ و ۸ درصد سؤالات در سطح تاکسونومی I، II، و III طراحی شده بود. جدول سه به تاکسونومی سؤالات در گروه‌های مربوطه بر اساس درصد پرداخته است.

جدول ۱: مقایسه شاخص‌های کمی بین ٥ گروه مربوطه بر اساس درصد

شاخصهای کمی	۱	۲	۳	۴	۵	میانگین
سؤالات با ضریب تمیز نامناسب (٪)	۱۸	۳۵	۲۵	۳۱	۱۰	۲۴
سؤالات دشوار(٪)	۴	۴	۳	۳/۵	۶	۴/۱
سؤالات با انتخاب سه گزینه صفر (٪)	۹/۵	۲۳	۱۲	۱۷	۳	۱۲/۹
ضریب پایایی	۸۴	۹۱/۶	۶۷	۶۲/۵	۸۴/۵	۷۷/۹۲
توزیع سؤالات نسبت به سرفصل‌های اعلام شده	۱۰۰	۹٦/۸	۹٥	۹۹	۹۷/٦	۹۸

[Downloaded from ijme.mui.ac.ir on 2026-11-15]

جدول ۲: مقایسه شاخص‌های کیفی (حیطه‌ها) بین ۵ گروه مربوطه بر اساس درصد

شاخص‌های کیفی (حیطه‌ها)	۱	۲	۳	۴	۵	میانگین
تعاریف و طبقه بندی‌ها	۶/۴	۱۸/۷۵	۵/۴۵	۲۴/۵	۲۱/۵	۱۵/۳۲
پاتوفیزیولوژی	۱۸	۲۶/۵	۲۳/۲۵	۱۳	۱۳	۱۸/۷۵
علایم و نشانه‌ها	۴۹	۲۹/۲۵	۳۹/۵	۱۴/۵	۲۱/۵	۳۰/۷۵
پاراکلینیک و نحوه تشخیص	۱۵	۱۱	۶	۲۵	۴۰	۲۰
درمان	۱۱/۴	۱۴	۲۵/۲۵	۲۱/۵	۲	۱۴/۶۷

جدول ۳: مقایسه شاخص‌های کیفی (تاکسونومی) بین ۵ گروه مربوطه بر اساس درصد

شاخص‌های کیفی	۱	۲	۳	۴	۵	میانگین
تاکسونومی یک(دانشی یا حفظ کردنی)	۶۶	۵۴	۸۲	۴۸	۴۲	۵۹
تاکسونومی دو (فهمیدنی)	۲۸	۳۵	۱۸	۳۸	۵۰	۳۴
تاکسونومی سه (به کار بستنی)	۶	۱۱	۰	۱۴	۸	۸

بحث

با عنایت به اهمیت پیشرفت تحصیلی و ارتباط راهبردهای آزمون و نهاده‌های یادگیری مورد استفاده گروه‌های آموزشی و میزان دستیابی به اهداف آموزشی آنان و نیز فقدان اطلاعات جامع از وضعیت این آزمون‌ها، این مطالعه به منظور بررسی کمی و کیفی آزمون‌های چهار گزینه‌ای در گروه‌های بالینی پوست، روانپزشکی، زنان،چشم و نورولوژی دانشکده پزشکی اصفهان پرداخته است.

در این مطالعه میانگین سؤالات دشوار۴/۱ درصد به دست آمد. در مطالعه کاظمی و همکاران نتایج نشان می‌دهد در ۷۳/۳ درصد امتحانات، سؤالات آسان بیشترین فراوانی نسبی را به خود اختصاص داده‌اند و تنها در دو مورد سؤالات از نظر ضریب دشواری مناسب بوده است(۲۵). در تحلیل سؤالات جامع آزمون نهایی پرستاری در دانشگاه علوم پزشکی شاهرود، ۵۷/۸ درصد سؤالات آسان و ۲۶/۹ درصد متوسط و ۱۵/۳ درصد سؤالات مشکل بوده است(۲۶). سطح دشواری سؤالات باید با هدف‌های تدریس و محتوای دروس و روش تدریس و توانایی فراگیران هماهنگ باشد. از آنجا که ارزشیابی ملاک مرجع، موقعیت هر یک از فراگیران را با توجه به میزان عملکرد یا سطح مطلوب پیشرفت تحصیلی تعیین می‌کند، تعیین کمیته‌های یادگیری برای فراگیران قبل از طرح سؤال ارزشمند خواهد بود. بنابراین، نتیجه‌گیری از یافته‌های به دست آمده و مناسبت ضریب دشواری هر کدام بستگی به هدف آزمون دارد(۲۷).

بررسی و مقایسه نتایج در مورد شاخص‌های کمی نشان داد که در کل، میانگین سؤالات با ضریب تمیز نامناسب ۲۴ درصد بوده است. صاحب‌نظران معتقدند سؤال‌های خوب، سؤال‌هایی هستند که ضریب دشواری متوسط و ضریب تمیز بالا دارند(۲۲). بنابراین، توصیه شده است پس از تعیین ضریب دشواری و شاخص تمیز سؤالات یک آزمون، طراحان سؤال به اصلاح سؤالات و بازنویسی آن سؤالاتی که نیاز به تغییر دارند اهتمام ورزند و سؤال‌هایی را که ضریب دشواری آنها بسیار بالا و یا بسیار پایین است، مورد تجدید نظر قرار دهند(۲۲). در مطالعه کاظمی و همکاران، ارزیابی شاخص تمیز سؤالات نشان داد در ۲۶/۷ درصد دروس بیش از ۵۰ درصد سؤالات شاخص تمیز بالا (سؤالات مناسب) را به خود اختصاص داده بودند. در بین مجموع سؤالات ارزیابی شده، ۴۴/۲ درصد سؤالات شاخص تمیز بالا، ۳۸/۹۶ درصد شاخص تمیز کم و ۱۷/۰۹ درصد شاخص تمیز

[Downloaded from ijme.mui.ac.ir on 2026-11-15]

منفی (نامناسب) داشتند(۲۵). در تحقیق دادگری، سؤالات جامع پرستاری شاهرود دارای ٤/٥ درصد شاخص تمیز بالا و ۷۲/۲ درصد شاخص تمیز کم و ۲۲/۳ درصد شاخص تمیز منفی (نامناسب) بوده است(۲٦). سؤالات با ضریب تمیز منفی نشان می‌دهد که در آن سؤال، گروه قوی بدتر از گروه ضعیف عمل کرده است. این گونه سؤال‌ها دارای معایب اساسی هستند که باید به کلی کنار گذاشته شوند و یا در آنها تجدید نظر اساسی صورت پذیرد(۲۲).

در این مطالعه میانگین سؤالات با انتخاب سه گزینه صفر ۱۲/۹ درصد به دست آمد، گزینه‌هایی که درصد انتخاب صفر یا بسیار کم دارند، گزینه‌های نامناسبی هستند و بود یا نبود این گزینه‌ها عملاً تأثیری در سؤال ندارد چرا که دانشجویان در اولین مرحله پاسخ به سؤال، گزینه‌های بسیار نامقبول را به سرعت کنار می‌گذارند. بنابراین در عمل این گزینه‌ها در مراحل تفکر و استدلال دانشجو وارد نمی‌شوند و دانشجو با یک سؤال سه گزینه‌ای یا گاهی دو گزینه‌ای روبرو است که احتمال شانس در چنین سؤالاتی افزایش یافته و در نتیجه قدرت تمیز بین دانشجویان قوی و ضعیف توسط این سؤال کاهش می‌یابد. نتایج بررسی کیفیت آزمون‌های پایان ترم فراگیران در دانشگاه علوم پزشکی بابل نشان داد طراحی آزمون‌ها وضعیت مناسب و راضی کننده نداشته و عموماً آزمون‌های مورد استفاده در فرایند ارزشیابی برایند آموزشی دانشجویان دارای اشکالات ساختاری اساسی بوده و به عنوان ابزار دقیق و معتبر عمل نمی‌نماید. چنانکه تفاوت‌های واقعی آزمودنی‌ها را نیز نمی‌سنجند(۲۸).

ضریب پایایی قابل قبول بالای ۸۰ درصد و ضریب پایایی مطلوب ۹۰ درصد در نظر گرفته شد. آزمون‌های دو گروه‌ها ضریب پایایی قابل قبول نداشته است و فقط یکی از گروه ضریب پایایی مطلوب داشته است. ضریب پایایی یک شاخص چندعاملی است و به زمان برگزاری آزمون، تعداد سؤالات، طول آزمون و... نیز بستگی دارد که بسیاری از آنها از عهده اساتید خارج است. پژوهشی در سال ۱۳۸۰ در مورد آگاهی اعضای هیأت‌علمی درباره‌ی روش‌های سنجش اعتبار و پایایی سؤالات امتحانی انجام شد و نتیجه گرفت که ۳٤ درصد آنان با نحوه پایایی آشنایی ندارند و در مجموع این مطالعه نشان داد اگر چه آگاهی اساتید از نحوه سنجش پایایی متوسط بود اما عملکرد آنها در این مورد ضعیف است(۲۹).

میانگین سؤالات نسبت به سرفصل‌های تدریس شده در هر پنج گروه بالای ۹۵ درصد بوده که تأکید بر طرح سؤال از منابع و طرح سؤالات امتحان پره انترنی از منابع رسمی موجب رعایت این نکته از سوی اساتید شده است. در تحقیق کمیلی نیز ٦۲/۷ درصد اساتید سؤالات خود را بر اساس اهداف آموزشی مطرح کرده‌اند(۳۰). از آنجا که دانشجویان برای فراگیری دروس تخصصی با مراجعه به طرح درس از انتظارات اساتید در ارزشیابی مطلع می‌شوند، عدم هماهنگی بین این دو با سردرگمی دانشجویان برای درک بهتر مطالب تئوریک همراه می‌شود همچنین این گونه مشکلات باعث می‌گردد دانشجویان برای درک بهتر وظایف آموزشی خود کمتر به طرح درس تدوین شده مراجعه نمایند. از دید آزمون‌شوندگان، تعداد ساعاتی که به یک محتوی اختصاص داده می‌شود و نیز تعداد سؤال از هر محتوی در ارزشیابی، نماینده اهمیت آن محتوی است و عدم هماهنگی بین این دو، این نگرش را ممکن است به وجود آورد که فراگیران وقت زیادی را صرف فراگیری محتوایی کرده‌اند که اهمیت چندانی نداشته است. طراحی سؤال بر اساس طرح درس می‌تواند از به وجود آمدن این اشکالات پیشگیری کند. از آنجا که معمولاً تدریس یک درس به وسیله چند استاد انجام می‌شود، برآورد تعداد سؤال هر محتوی به وسیله مسؤول درس و اعلام آن به کلیه اساتید ضرورت دارد. برخلاف نتایج این تحقیق در مطالعه کاظمی و همکاران این موضوع به وسیله طراحان سؤال کمتر مورد توجه قرار گرفته است(۲۵).

بررسی توزیع فراوانی نسبی سؤالات چندگزینه‌ای مورد

[Downloaded from ijme.mui.ac.ir on 2026-11-15]

نظر براساس حیطه نشان داد فراوانی سؤالات بیشتر در حیطه علایم و نشانه بوده است. بحث حیطه سؤالات بستگی به مهارت و رغبت اساتید در تهیه جدول دو بعدی دارد. جدول دو بعدی قبل از آزمون باید توسط اساتید تهیه شود و سپس آزمون اجرا گردد ولی در مطالعه حاضر با توجه به اینکه هنوز جدول دو بعدی توسط اساتید قبل از برگزاری آزمون تهیه نمی‌شد، به ناچار حیطه‌بندی پس از برگزاری آزمون تعیین شد.

اگر چه تاکنون نسبت خاصی برای سطوح مختلف شناختی سؤالات چند گزینه‌ای یک آزمون، پیشنهاد نشده است، اما انتظار می‌رود در یک آزمون، سؤالات امتحانی بتوانند جنبه‌های مختلف توانایی دانشجویان را مورد سنجش قرار داده و قابلیت‌های کسب شده‌ای که برای آن تربیت شده‌اند را به خوبی نمایان سازد(۳۱). در مورد تاکسونومی سؤالات فرض بر این بود که با توجه به هدف مقطع سوم دوره پزشکی عمومی یعنی کارآموزی بالینی توزیع فراوانی سؤالات بیشتر در تاکسونومی دو باشد. بررسی سؤالات مورد نظر نشان داد سؤالات بیشتر دارای تاکسونومی یک بوده است. در مطالعه بررسی مهارت‌های طراحی سؤال مربیان بالینی پرستاری، سطح ٤١/٢ درصد سؤالات مطرح شده در سطح درک، ٦/٧ درصد در سطح قضاوت، ١٦/٨ درصد در سطح دانش، ٨/٩ درصد در سطح کاربرد، ١١/٢ درصد در سطح تجزیه و تحلیل و ١٥/٣ درصد در سطح ترکیب قرار داشته است(۳۲). در تحقیق کاظمی و همکاران بیشترین فراوانی را در کلیه دروس، سؤالات سطح یک به خود اختصاص داده‌اند و تنها ٢٨/٧ درصد سؤالات در سطح سه قرار دارند(۲۵). تحقیقی نشان داد ٥٠ درصد سؤالات ارزیابی دانشجویان دوره کارشناسی دانشگاه آدالاید استرالیا در سطح یک طبقه‌بندی بلوم قرار داشته است(۳۳). یافته‌های پژوهش دیگری نشان داده که حدود ٩٠ درصد سؤالات امتحانی، سطوح پایین طبقه‌بندی بلوم (تاکسونومی I) را ارزشیابی نموده و تنها ١٠/٦ درصد سؤالات در سطوح بالاتر یعنی تفسیر و حل مسئله (تاکسونومی II و III) بوده‌اند(۳٤). در حالی که نتایج یک مطالعه در دانشکده پرستاری و مامایی گرگان میزان سؤالات طراحی شده در سطح پایین (تاکسونومی I) را ٥٧/٧ درصد گزارش شده است(۳٤). همچنین نتایج دو مطالعه در دانشگاه‌های هنگ کنگ و استرالیا نشان داد که به ترتیب ٩١/١ درصد و ٩١/٢ درصد سؤالات در سطوح پایین حیطه شناختی (تاکسونومیI) بوده است(۳۵و۳٦). نتایج مطالعات انجام شده در اصفهان و همدان و مازندران نشان داد که بیشتر سؤالات در سطح تاکسونومی I بوده‌اند(۳۸و۳۷و۱۸).

از مزایای این پژوهش محدود نکردن مقوله ارزشیابی آزمون صرفاً به item analysis (ارائه نتایج به شکل منحنی و درصدها و بدون تحلیل) است زیرا که تعدادی از سؤالات در این آنالیزها به نظر نامقبول می‌رسید ولی در مراحل بعدی و آنالیز کیفی، سؤال مناسب و استانداردی بود. متأسفانه ارسال روتین تعدادی جدول شامل عدد و رقم هر ماه که گنگ و مبهم نیز می‌باشد و خیلی از اساتید با وجود آموزش در این زمینه ولی به علت فراموشی مطلب و عدم استمرار آموزش آنان سردرگم شده و از دقت در آن سر باز می‌زنند.

انجام مطالعات مشابه در سایر گروه‌ها و تکرار در گروه‌های فعلی، برگزاری مستمر و منظم کارگاه‌های آموزشی با شرکت همه اساتید، دادن فیدبک آنالیز آزمون‌ها به گروه‌ها، بازبینی سؤالات، درست قبل از امتحان توسط استاد مسؤول آزمون (نه رزیدنت‌ها)، تهیه جدول دو بعدی برای هر آزمون و تعیین تاکسونومی و حیطه سؤالات و بودجه‌بندی آن، تربیت و آموزش کارشناسان در دانشکده جهت کمک به اساتید و استاندارد کردن سؤالات قبل از آزمون (تشکیل اتاق سؤال) و تخصیص درصدی از مجموع امتیازات سالانه ارتقای اعضای هیأت‌علمی به اساتیدی که در زمینه استانداردسازی سؤالات فعالیت می‌نمایند، پیشنهاد می‌گردد.

[Downloaded from ijme.mui.ac.ir on 2026-11-15]

محدودیت‌ها و مشکلات اجرائی مطالعه شامل مشکل اجماع در مورد حیطه‌بندی و تاکسونومی سؤالات توسط اساتید و کارشناسان، عدم وجود یک نفر کارشناس روان سنجی آزمون در بین اعضای هیأت‌علمی، کمبود کارشناس خبره کیفی آزمون، شفاف نبودن هدف دوره و هدف سؤالات، متغیر بودن تعداد گروه‌های دانشجویی به عنوان مثال در یک دوره ۸ نفر و در دوره بعد ۳۰ نفر بود که در زمینه تعیین بعضی میانگین‌ها به خصوص شاخص پایایی اختلال ایجاد می‌کرد و عدم توجه اساتید به تهیه جدول دو بعدی قبل از آزمون بود.

نتیجه‌گیری

آزمون‌ها بخش عمده‌ای از فرآیند آموزش هستندکه غالباً از شکل نظام‌مند و ساختاری منسجم و استاندارد برخوردار نبوده و در حاشیه قرار دارند لذا کسب اطلاعات با ارزش در این زمینه، تصمیم‌گیری‌ها و برنامه‌ریزی‌های بعدی را ممکن می‌سازد. کسب اطلاعات عموماً از رهگذر تجزیه و تحلیل سؤالات حاصل می‌شود. تجزیه و تحلیل سؤال، فرآیند بررسی نقادانه سؤالات آزمون‌ها بوده، هدف آن شناسایی و کاهش خطا است. باید طراحان سؤال به طور مستمر فرآیند تجزیه و تحلیل سؤالات را به عنوان معیار کیفیت و حذف سؤالات غیرقابل قبول، جایگزین ساختن آنها با سؤالات بهتر و نگه داشتن سؤالاتی که ملاک‌های مطلوب دارند را به کار گیرند تا به آن دسته از اهداف آموزشی که مهم می‌باشد نائل شوند. با عنایت به نتایج پروژه فعلی، فیدبک گروه‌های آموزشی و اساتید و استنباط کارشناسان در فرآیند اجرای پروژه این نکته واضح و مورد تأکید بود که مقوله آنالیز کمی و کیفی سؤالات مورد نیاز است و قطعاً اساتید در این موضوع نیاز به آشنایی بیشتر دارند کما اینکه نتایج مطالعات دیگر این موضوع را تأیید کرده است.

منابع

1. Guilbert JJ. Educational handbook for health personnel. 6th ed. Geneva: World Health Organization. 1987.
2. Newble DI, Jaeger K. The effect of assessment and examinations on the learning of medical students. Med Educ. 1983; 17(3):165-171.
3 Grauer GF, Forrester SD, Shuman C, Sanderson MW. Comparison of student performance after lecture-based and problem-based teaching in a case-based/large group. J Vet Med Educ. 2008; 35 (2):310-317.
4. Smith-Strøm H, Nortvedt MW. Evaluation of evidence-based methods used to teach nursing students to critically appraise evidence. J Nurs Educ. 2008; 47 (8): 372-375.
5. Yoshida M, Tanaka M, Mizuno K, Ishii A, Nozaki K, Urakawa A, et al. Factors influencing the academic motivation of individual college students. Int J Neurosci. 2008; 118 (10): 1400-1411.
6. Franko DL, Cousineau TM, Trant M, Green TC, Rancourt D, Thompson D, et al. Motivation, self-efficacy, physical activity and nutrition in college students: Randomized controlled trial of an education program. Prev Med internet-based. 2008; 47 (4): 369-377.
7. Sepasi H. [A comparison between and True-False test item in measuring multiple-choice Shahid Chamran University undergraduate students' academic achievement in measurement and evaluation course with the control of students' absences]. Journal of Education and Psychology. 1998; 5(1&2): 31-56. [Persian]
8. Mavis BE, Cole BL, Hoppe RB. A survey of student assessment in US medical schools: the balance of breadth versus fidelity. Teach Learn Med. 2001; 13(2):74-79.
9. Kaveh Tabatabaei MS, Bahreini Tousi MH, Modabber Azizi MJ, Ebrahimzadeh S, Bahreini Tousi V, Bahreini Tousi K. [Analyses survey on multiple choice question in medical school in Mashhad University of medical sciences in 2002]. Medical Journal of Mashhad University of Medical Sciences. 2003; 45 (76): 89-95. [Persian]
10. Holsgrove G, Elzubeir M. Imprecise terms in UK medical multiple-choice questions: What examiners think they mean. Med Educ. 1998; 32 (4): 343-350.

[Downloaded from ijme.mui.ac.ir on 2026-11-15]

11. A critical review of student assessment options. [cited 2010 Feb 22]. Available at: http://www.nmu.edu/ soa/ review. html.
12. Wink DM. Using questioning as a teaching strategy. Nurse Educ. 1993; 18(5): 11-15.
13 Marzano RJ, Brandt RS, Hughes CS, Jones BF, Presseisen BZ. Dimensions of thinking: a framework for curriculum and instruction. 1 st ed. Assn for Supervision & Curriculum. 1998.
14 Schultheis NM. Writing cognitive educational objectives and multiple-choice test questions. Am J Health Syst pharm. 1998; 15; 55(22):2397-2401.
15 Parson J, Fenewick T. Using objective testes to evaluate. Am J phys. 1999; 89: 45- 53.
16. Lowry S. Assessment of students. BMJ. 1993; 306:51-54.
17. 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): 133-143.
18. Moatamedi A, Seif A. Survey on evaluation methods of faculties of evaluation method course in Training Teacher Center and related factors. [dissertations].tehran: Allameh Tabatabaei University. 1994. [Persian]
19.Sepasi H. A study of cognitive domain and analysis of psychometric characteristics of test terms in final examinations of Arabic, calculus and biology courses of the third grade high school girls in different Scio-economic status in Khuzestan, Iran. Journal of Education and Psychology. 2006; 13(4): 57-78. [Persian]
20 Kaveh Tabatabaei MS, Bahreini Tousi MH, Modabber Azizi MJ, Ebrahimzadeh S, Bahreini Tousi V, Bahreini Tousi K.[Evaluation methods in clinical ward and medical student's views' in 2002]. Medical Journal of Mashhad University of Medical sciences. 2003; 45 (76): 97-103. [Persian]
21. Dadgostarnia M, Gholami V. [Barasi didgah asatid balini darbareh tasir eraeh fidbak be asatid tarah soal shakhsi baray asatid bar behbood kaifiat azmonha]. Special Supplement for 7th Nathional Congress on Medical Education; 2006. tabriz: University of Medical Science. 2006 14:52. [persian]
22. Saiff AA. [Andazegiri sanjesh va arzeshyabi amozeshi]. 17th ed. Tehran: Doran. 2005. [Persian]
23 Pasha sharifi H.[Osole ravansangi va ravanazmaei].Tehran: Entesharat roshd. 1374.
24. Fatahi A. [Tasire amozesh bar azaye heiate elmi dar tarh soalat chandgozinehei az nazar etebar sori va mohtavaei].Special Supplement for 7th Nathional Congress on Medical Education. tabriz: University of Medical Science . 2006; 14: 63-64. [persian]
25. Kazemi A, Ehsanpour S, Hassanzadeh A .[Investigating the Academic Achievement Evaluation of Specialized Theoretical Courses of Midwifery BS]. Iranian Journal of Medical Education. 2010; 9 (4):346-355. [Persian]
26. Dadgari A, Ebrahimi H, Hassani MR, Mashyekhi Y. [Analysis of the comprehensive nursing final exam in Shahrod faculty of medical sciences]. Shahrekord University of Medical Sciences Journal. 2003; 4(4): 4-10. [Persian]
27. Naderi E, Seyf Naraghi M. [Evaluation and measurement and basic analysis tools in cultural sciences and psychology]. 1st ed. Tehran: Misagh. 1997. [Persian]
28. Najar SH, Abedi P. [Baresi agahi va amalkard azaye heiate elmi daneshgah olom peseshki Babol dar mored raveshhayeh sanjesh etebar va payaei soalat emtehani sal 1380]. Iranian Journal of Medical Education. 2003; 7: 119-120. [Persian]
29. Azizi F. [Amozesh olom pezeshki, chaleshha va cheshmandazha]Tehran. Entesharat moavenat amozeshi vezarat behdasht. 1382.[Persian]
30. Komeile GR, Rezai A. [Methods of student assessment used by faculty members of basic medical sciences in medical university of Zahedan]. Iranian Journal of Medical Education 2001; 1(4): 52-57. [Persian]
31. 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.
32. Khoddam H, Jouybari L, Sanagoo A. [Nursing clinical instructors' skills in designing question]. Iranian Journal of Medical Education. 2005; 5 (2):191-192. [Persian]
33.Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. BMC Med Educ. 2007 Nov 28; 7: 49.

[Downloaded from ijme.mui.ac.ir on 2026-11-15]

34 Shakoornia A, Khosravi A, Mozafar A, Elhampoor H. [Barasi soalat emtehani azaye heiate elmi daneshgah olom peseshki jondi shapoor Ahvaz ba takid bar sakhtar soalat chand gozinehei dar sal 1386]. Hamhayeh toseh dar amozesh pezeshki. 2005; 6(2): 129-138. [Persian]
35. Tarrant M, Knierim A, Hayes SK, Ware J. The frequency of item writing flaws in questions used in high stakes multiple-choice nursing assessments. Nur Educ Today. 2006; 26 (8): 662-71.
36 Sellappah S, Hussey T, Blackmore AM, McMurray A. The use of questioning strategies by clinical teacher. J Adv Nurs. 1998; 28 (1): 142-148.
37 Garag Yaraghi M, Ebrahimi A, Ashirioun V, Azadbakht S, Avizhgan M, Esmaeili A, et al. [An analysis of multiple choice questions for undergraduate medical students and designing necessary steps to correct students' evaluation process]. Proceeding of the 9th National Congress of Medical Education; 2008. Yazd: Shaheed Sadoughi University of Medical Sciences. 2008: 27. [Persian]
38. Arab M. [Survey on defect of multiple choice questions in medical school of Hamadan Medical University in 2000]. Proceeding of the 4th National Congress of Medical Education; 2001. Tehran: University of Medical Sciences; 2001: 51. [Persian]

[Downloaded from ijme.mui.ac.ir on 2026-11-15]

Assessment of Qualitative and Quantitative Indexes of Clerkship Tests in General Medicine

Mohammad Garakyaraghi[1],Maryam Avizhgan[2],Amrollah Ebrahimi[3],Elham Esfandiari[4],Atoosa Esmaeili[5],Shahram Shayan[6],Vahid Ashourioun[7], Nikoo Yamani[8]

Abstract

Introduction: *Using multiple choice question tests, as an objective testing method, is the most common students evaluation procedure , and it is very important to design these tests properly.This study aimed to assess clerkship tests of general medicine courses in the training group of dermatology, psychiatry, gynecology, ophthalmology and neurology of medical college of Isfahan University of Medical Sciences both quantitatively and qualitatively.*

Methods: *This descriptive cross- sectional study was done on the latest tests of clerkship courses of general medicine including dermatology, psychiatry, gynecology, ophthalmology, and neurology in 2008. The quantitative and qualitative indexes of 15 previously held tests were extracted by using valid and reliable checklists used for determining questions with inappropriate question discrimination index, difficulty coefficient, question with 3 choices of zero, reliability coefficient, question distribution based on the announced topics, the scope of questions based on the objectives, the question types based on taxonomy. Data were analyzed through SPSS software using descriptive statistics.*

Results: *Among the quantitative indexes, mean of queations with inappopriate discrimination index was 24%, difficult questions 4.1%, and questions with 3 choices of zero was 12.9%. The average of questions distribution basedon the announced topics and the reliability coefficient were 98% and 77.92%, respectively. among the qualitative indexes, the highest relative frequency of questions were in the fild of signs and symptoms with 30.75 % and the first taxonomy with 59 %.*

Conclusion: *Considering the importance of using standard questions (which are according to the rules for designing multiple choise question) for holding appropriate exams, continuous review of test questions and administering training courses on test questions designing , creating question banks and putting emphasis on preparing two dimensional tables for tests before giving exams are effective on improvement of the quality of the questions.*

Key words: *Test, Multiple-choice Question, Student Evaluation, Clerkship*

Addresses:

[1] Associate professor, Department of cardiology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. Email: garakyaraghi@med.mui.ac.i

[2](✉)Ph.D Student in Course Planning .Educational development office expert, EDO, school of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran.E:mail: avizhgan@yahoo.com

[3] MD, Educational development office expert, EDO, school of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran.E:mail: ev_11@yahoo.com

[4] Master of Sciences,Researcher of genetic department, school of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran.E:mail: elham_esfandiari@yahoo.com

[5] Educational development office expert, EDO, school of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. Email: eatousa@yahoo.com

[6] Medical Education Research center expert, Isfahan university of medical sciences, Isfahan, Iran, Email: shayan@edc.mui.ac.ir

[7] Instructor, Medical Education Research center, Isfahan university of medical sciences, Isfahan, Iran, Email:ashourioun@med.mui.ac.ir

[8] Assistant professor, Medical Education Research center, Isfahan university of medical sciences, Isfahan, Iran, Email:yamani@edc.mui.ac.ir

[Downloaded from ijme.mui.ac.ir on 2026-11-15]