

مقایسه روش‌های آزمون چندگزینه‌یی، صحیح - غلط و کوتاه پاسخ در دانشجویان پزشکی

دکتر نیر رسائیان^۱، سهیلا نخعی^۲، دکتر نوید صادقی‌قندهاری^۳

Title: *Comparison of three exam techniques in medical students: multiple-choice, true-false and short answer questions.*

Authors: *Rassaian N, (MD, PhD); Nakhaei S, (MA); Sadeghi Ghandehari N, (MD).*

Abstract: *In most studies the validities of different methods for student assessment have independently been evaluated from each other.*

In this descriptive-analytical research, 1372 medical students were asked by the three types of questions, so that the result of each method can be compared to that of the other two for each student. Nine classes from different curricula were randomly selected and in each of the them, 30 questions for 10 selected topics were distributed, to a total of 270 questions.

The analysis of the students' scores after excluding questions with a discrimination index of less than 0.3, showed that the most valid assessment method was the short-answer question ($r=0.685$). The kappa coefficients of the short-answer method and the other two methods are in the "fair" range of agreement. The highest coefficient of contingency is between the multiple-choice and true-false questions (0.402). The percentage of correct answers is the lowest in short-answer questions (53%) and the highest in true-false questions (65%). The higher percentage of incorrect answers in multiple-choice (32%) compared to the true-false (26%) questions can be the result of the students being confronted with alternative choices in multiple-choice questions, which apparently encourages them to choose one without using the relevant knowledge.

In conclusion, it is recommended to use short-answer and true-false questions as the main components of examinations, instead of only multiple-choice questions, so that students' learning and recall are tested and they are not misled.

Keywords: *medical student assessment, short-answer, true-false, multiple-choice, examination question.*

- ۱- گروه فیزیولوژی، دانشکده پزشکی، دانشگاه علوم پزشکی شهید بهشتی
- ۲- روانشناسی آموزشی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی
- ۳- دانشکده پزشکی، دانشگاه علوم پزشکی ایران

Archive of SID

چکیده:

تا کنون اعتبار روش‌های مختلف برای ارزیابی دانش فراگیران، مستقل از یکدیگر بررسی و تعیین گردیده است. این تحقیق توصیفی-تحلیلی به گونه‌ی طراحی شده است که یک نکته علمی به سه روش چندگزینه‌یی، صحیح-غلط و کوتاه پاسخ (تکمیل جملات با کلمه مناسب) از ۱۳۷۲ دانشجوی پزشکی سؤال گردیده تا بتوان نتایج هر روش را با دو روش دیگر در همان افراد مقایسه نمود. هر یک از استادان ۹ کلاس با انتخاب تصادفی، ۳۰ سؤال با سه روش برای ۱۰ نکته علمی طرح نمودند که مجموعاً ۲۷۰ سؤال مورد بررسی قرار گرفت.

آنالیز فاکتور نمرات دانشجویان، پس از حذف سئوال‌ات با درجه تمایز کمتر از ۰/۳ نشان داد که معتبرترین روش مورد بررسی، سئوال‌ات کوتاه پاسخ است ($r = ۰/۶۸۵$). ضرایب کاپا برای روش کوتاه پاسخ و دو روش دیگر در محدوده توافقی "fair" قرار داشت. بیشترین ضریب توافقی بین روش‌های چندگزینه‌یی و صحیح-غلط بود (۰/۴۰۲). تعداد جواب‌های درست در روش کوتاه پاسخ کمترین (۰/۵۳) و در روش صحیح-غلط بیشترین (۰/۶۵) بود. تعداد جواب‌های غلط در روش چندگزینه‌یی (۰/۳۲) بیشتر از روش صحیح-غلط (۰/۲۶) بود که می‌تواند به علت مواجهه دانشجویان با تعداد زیاد پاسخ‌های انحرافی در روش چندگزینه‌یی باشد که بدون اتکا به دانش خود آنها را علامت زده‌اند.

پیشنهاد میشود برای ارزیابی میزان یادگیری و یادداری دانشجویان از یکطرف و گمراه نکردن آنان از طرف دیگر روش‌های کوتاه پاسخ و صحیح-غلط در ترکیب اصلی سئوال‌ات قرار گیرند و به سئوال‌ات چندگزینه‌یی اکتفا نگردد.

کل واژگان: ارزیابی دانشجویان پزشکی، کوتاه پاسخ، صحیح-غلط، چندگزینه‌یی، سؤال امتحانی.

مقدمه:

این نوع سئوال‌ات رعایت نمی‌شود (۵). سئوال‌اتی که در امتحانات مورد آمریکا استفاده می‌شود عمدتاً با روش چندگزینه‌یی است و توسط متخصصین مربوطه^۱ پس از بازنگری و بحث‌های موشکافانه طولانی انتخاب گردیده و بقیه آنها حذف می‌شوند (۶) که خود نشان‌دهنده مشکل بودن طراحی اینگونه سئوال‌ات است و از طرف دیگر با منسوخ شدن تخصیص نمره منفی به این نوع سئوال‌ات، ارتباط قبولی در امتحان با دانش فراگیران ضعیف به نظر می‌رسد. نکته حائز اهمیت آن است که گذرانیدن امتحانات حاصل از سؤال خوانی استقبال از کلاس‌های درس را که قاعدتاً محل انتقال اندیشه، تحلیل و تفسیر اطلاعات و حتی آموزش تحقیق است (۷ و ۸)، به طور نگران‌کننده‌یی کاهش داده است.

از آنجا که تغییر در شیوه ارزیابی و در نتیجه بازتاب آن در نحوه مطالعه و کاربرد مناسب از دانسته‌ها بسیار ضروری به نظر می‌رسد، این بررسی برای مقایسه روش‌های مختلف ارزیابی دانشجویان پزشکی از طریق تعیین تعداد پاسخ‌های درست، غلط

به نظر می‌رسد روش ارزیابی از نظر علمی تأثیر زیادی در نحوه مطالعه دانشجویان دارد و اطلاع آنان از شکل رایج سئوال‌ات امتحانی، سؤال خوانی را جانشین تعمق و تفکر در هنگام یادگیری نموده است. در سال ۱۹۹۸ Hammond و همکاران گزارش نمودند که سئوال‌ات چندگزینه‌یی بخشی از فرآیند ارزشیابی اولیه و نهایی دوره فوق تخصصی در کالج سلطنتی بیهوشی انگلستان را تشکیل می‌دهد. نتایج مطالعه بر روی داوطلبان شرکت کننده در کلاس مروری قبل از امتحان نشان داد که بخش مهمی از آمادگی فرد برای امتحان، ارزیابی او از روش آن است (۱). نحوه ارزیابی متداول برای دانشجویان پزشکی طرح سئوال‌ات چندگزینه‌یی است (۲). چگونگی طرح این سئوال‌ات باید دربرگیرنده اهداف شناختی آموزشی عنوان شده در طرح درس مربوطه باشد که این اهداف بطور مطمئن قابل مشاهده و اندازه‌گیری است (۳). ولی بهترین شرایط استفاده از این روش همراه نمودن آن با روش‌های دیگر ارزیابی است (۴).

با توجه به آزمون‌های تستی استاندارد مختلف و مزایای آنها که در کتاب‌های درسی آموزشی آمده است گاهی اصول طراحی

¹ - National Board of Medical Examiners

بالاست. انواع سئوالات عینی عبارتند از: سئوالات چندگزینه‌یی^۵،

سئوالات صحیح- غلط^۶ و سئوالات تطبیقی^۷.

با تحلیل سئوالات می‌توان اطلاعات آماری و عددی درباره هر سؤال موجود در تست به دست آورد، که مهمترین آنها عبارتند از:

الف: شاخص آسانی سؤال^۸: درصد دانشجویانی که به سؤال مورد نظر پاسخ درست داده‌اند شاخص درجه آسانی سؤال است. سئوالاتی که شاخص مزبور در آنها بین ۳۰٪ تا ۷۰٪ باشد سئوالات مطلوبی بوده و سئوالات کمتر از ۳۰٪ سخت و بیش از ۷۰٪ آسان می‌باشند.

ب: شاخص تمایز^۹: در این تحلیل پاسخ سؤال مورد نظر توسط دانشجویان خوب که ۲۵٪ بالای نمره کل امتحان را کسب نموده‌اند با پاسخ همان سؤال توسط دانشجویان ضعیف که در گروه ۲۵٪ پائین آن قرار دارند مقایسه می‌شود. سئوالاتی که به آن تعداد کمی از دانشجویان خوب و تعداد زیادی از دانشجویان ضعیف پاسخ داده‌اند قاعداً سئوالاتی نبوده است که بوسیله آن بتوان دانشجویان خوب را از ضعیف تمیز داد که خود یکی از مهمترین اهداف یک ارزیابی است. فرمول درجه تمایز سؤال (DI) به قرار زیر است:

$$DI = \frac{a}{n_1} - \frac{b}{n_2}$$

a = تعداد پاسخهای صحیح در ۲۵٪ بالای نمره کل، n_1 = تعداد دانشجویان در گروه ۲۵٪ بالا

b = تعداد پاسخهای صحیح در ۲۵٪ پائین نمره کل، n_2 = تعداد دانشجویان در گروه ۲۵٪ پائین

سئوالاتی که درجه تمایز آنها کمتر از ۰/۳ باشد سئوالات خوبی نبوده و نباید مجدداً از آنها استفاده شود

روش‌هایی که در این تحقیق مورد بررسی و مقایسه قرار گرفته‌اند آزمون‌های چندگزینه‌یی، صحیح- غلط و کوتاه پاسخ بود و برای طرح سئوالات با هماهنگی قبلی به حضور استاد هر درس رسیده و پس از توضیح درباره اهداف طرح و چگونگی طرح سؤال، ۱۰ نکته علمی به سه شکل مجموعاً ۳۰ سؤال از نوع تاکسونومی ۲ نوشته شد به طوری که پاسخ هر نکته در بدنه

و بدون پاسخ، مدت زمان تصحیح اوراق و نمرات دانشجویان در سه روش طراحی گردیده است.

روش کار:

این بررسی با روش توصیفی- تحلیلی و با استفاده از سئوالات امتحانی دوره‌های علوم پایه، فیزیوپاتولوژی و بالینی دانشکده پزشکی دانشگاه‌های علوم پزشکی و خدمات بهداشتی درمانی شهیدبهشتی و ایران در امتحانات ترم دوم سال تحصیلی ۷۶-۱۳۷۵ با حضور ۱۳۷۲ نفر دانشجوی شرکت کننده انجام شد. نحوه انتخاب دانشجویان منوط به موافقت دانشگاه‌های مورد نظر و اعلام آمادگی استادان ارجمند جهت صرف وقت و همکاری در طراحی سؤال طبق اهداف و روش بررسی بود. لذا پس از کسب مجوز از معاونین محترم آموزشی دانشگاه‌های ذیربط و با دسترسی به برنامه‌های امتحانی دوره‌های تحصیلی مورد نظر، بطور تصادفی مجموعاً امتحانات سه درس از هر دوره انتخاب شد. گزینش روش‌های آزمون و مزایای سئوالات طراحی شده براساس مطالبی است که در کتب منتشر شده آموزشی ذکر گردیده است (۹ و ۱۰).

آزمون‌های استاندارد:

روش‌های استاندارد طرح سؤال به چند گروه تقسیم شده‌اند:

۱- سئوالات تشریحی^۱ که با وجود اعتبار بیشتر آن از سایر انواع سئوالات به علت وقت زیادی که برای تصحیح اوراق طلب می‌کند تقریباً برای ارزیابی دانشجویان پزشکی به کار نمی‌رود.

۲- آزمون‌های کوتاه پاسخ^۲ که بدنه سؤال نسبتاً طولانی ولی پاسخ آن در یک یا چند کلمه می‌باشد این نوع سئوالات نیز با وجود اعتبار زیاد خیلی کم استفاده می‌شوند.

۳- سئوالات ساختاری^۳ که مهارت حل مسئله را در دانشجو می‌سنجد و مهمترین آنها Management Problem (PMP) Patient و Modified Essay Question (MEQ) می‌باشند که گاهی در امتحانات بالینی از آنها استفاده می‌شود.

۴- آزمون‌های عینی^۴ که هدف آنها مشخص نمودن پاسخ‌های مورد نظر توسط دانشجو می‌باشد. زمان تصحیح آنها کوتاه بوده و در ارزیابی دانشجویان پزشکی کاربرد بسیار دارند زیرا می‌توانند شامل بخش زیادی از محتوای درس باشند و اعتبار آنها نیز نسبتاً

⁵ - multiple-choice

⁶ - true-false

⁷ - matching

⁸ - facility index (FI)

⁹ - discrimination index (DI)

¹ - essay

² - short answer

³ - structural

⁴ - objective

از ۱ تا ۱۰، صحیح- غلط از ۱۱ تا ۲۰ و کوتاه پاسخ از ۲۱ تا ۳۰ کدگذاری شد. با توجه به ترتیب رعایت شده پاسخ‌های با کد ۳، ۱۳ و ۲۳ مربوط به یک نکته علمی مشترک و کدهای ۴، ۱۴ و ۲۴ در ارتباط با نکته علمی دیگر بود که به ترتیب با روش‌های چند گزینه‌یی، صحیح- غلط و کوتاه پاسخ طرح گردیده بودند. برای تجزیه تحلیل داده‌ها از آزمون‌های آماری Kappa, t-test, Duncan's multiple range test و آنالیز فاکتور استفاده شده است.

یافته‌ها:

توصیف جامعه مورد بررسی در جدول ۱ آورده شده است. توصیف سئوال‌ات از نظر درجه سختی و آسانی در جدول ۲ آمده است و در آن از ۹۰ سئوال در هر دوره، بیشترین تعداد سئوال سخت در علوم پایه (۲۵ عدد) و کمترین آنها در دوره بالینی (۶ عدد) بوده است و از طرف دیگر کمترین تعداد سئوال آسان در علوم پایه (۱۸ عدد) و بیشترین آنها در دوره بالینی (۴۵ عدد) دیده شد.

توصیف سئوال‌ات با درجه تمایز کمتر از ۰/۳ در نمودار ۱ آورده شده است که بیشترین تعداد آنها در دوره بالینی با روش صحیح- غلط بوده و هیچ سئوالی با درجه تمایز غیر قابل قبول با روش چند گزینه‌یی برای دانشجویان این دوره طرح نشده است.

سئوال همان نکته در روش‌های دیگر آورده نشده و در معلومات دانشجو مداخله نکند. ترتیب سئوال‌ات به نحوی بود که اولین

سئوال چندگزینه‌یی، اولین سئوال صحیح - غلط و اولین سئوال کوتاه پاسخ درباره یک مطلب علمی مشترک بوده و این ترتیب تا دهمین سئوال با هر روش رعایت گردید.

برای کاستن میزان احتمال استفاده از شانس و قابل مقایسه گردیدن پاسخ‌های دو روش دیگر با روش کوتاه پاسخ، در توضیح سئوال‌ات چندگزینه‌یی و صحیح - غلط داشتن نمره منفی ذکر گردیده و برای روش کوتاه پاسخ نداشتن نمره منفی کتباً به دانشجویان یادآوری شد. سئوال‌ات تایپ شده در دو صفحه پایانی برگه‌های امتحانی دروس مربوطه قرار داده شد. علاقه استادان بر این بود که نمره این بخش از امتحان به عنوان امتیاز اضافی و یا بخشی از نمره نهایی دانشجویان محاسبه گردد.

پس از برگزاری هر امتحان، تصحیح اوراق با استفاده از کلید سئوال‌ات توسط دو نفر از همکاران انجام شده و بوسیله یک نفر مجدداً کنترل گردید. از مصححین درخواست شد که تعداد پاسخ‌های غلط، موارد بدون جواب و مدت زمان تصحیح سئوال‌ات دهگانه با هر روش را در محل ثابتی از برگه امتحانی هر دانشجو یادداشت نمایند. نمرات دانشجویان در اسرع وقت طی لیستی تحویل آموزش دانشکده مربوطه شد و طبق مجوز قبلی اوراق امتحانی جهت تحلیل‌های آماری از متعلقات طرح بوده و نزد مجری می‌باشد. داده‌های این بررسی به ترتیب سئوال‌ات با هر روش جمع‌آوری گردید و در هر کلاس، پاسخ سئوال‌ات چندگزینه‌یی

جدول ۱- توصیف جامعه مورد بررسی از نظر دوره تحصیلی، شماره و نام امتحان، تعداد دانشجویان و نام دانشگاه ترم دوم سال تحصیلی ۱۳۷۵-۷۶

دوره	شماره کلاس و نام امتحان درس	تعداد دانشجو	نام دانشگاه علوم پزشکی
علوم پایه	۱- فیزیولوژی کلیه و تعادل اسید و باز	۲۵۶	شهید بهشتی
	۲- ایمونولوژی	۹۶	ایران
	۳- فیزیولوژی گوارش	۱۸۱	ایران
فیزیوپاتولوژی	۴- پاتولوژی اختصاصی	۷۳	شهید بهشتی
	۵- گوارش	۱۲۸	شهید بهشتی
	۶- روماتولوژی	۱۲۱	ایران
بالینی	۷- زنان و زایمان	۱۳۶	شهید بهشتی
	۸- جراحی اعصاب	۲۳۲	شهید بهشتی
	۹- داخلی اعصاب	۱۴۹	شهید بهشتی

جدول ۲- توزیع فراوانی سئوال‌ات با ضریب آسانی (FI) آنها در دوره های تحصیل پزشکی بر حسب روش ارزیابی (۳۰٪ < سئوال‌ات سخت و ۷۰٪ > سئوال‌ات آسان).

روش ارزیابی	علوم پایه		فیزیوپاتولوژی		بالینی		جمع	
	> ۷۰٪	< ۳۰٪	> ۷۰٪	< ۳۰٪	> ۷۰٪	< ۳۰٪	> ۷۰٪	< ۳۰٪
چند گزینه‌یی	۸	۵	۷	۸	۲	۱۳	۱۴	۲۹
صحیح- غلط	۷	۹	۳	۷	۰	۲۲	۱۲	۳۶
کوتاه پاسخ	۳	۱۱	۷	۶	۴	۱۰	۲۲	۱۹
جمع	۱۸	۲۵	۱۷	۲۱	۶	۴۵	۴۸	۸۴

شده (۹۰ سؤال برای هر روش) محاسبه گردید و سئوال‌ات نامناسب (با درجه تمایز کمتر از ۰/۳) از داده‌های این بخش، از تجزیه و تحلیل نتایج حذف شد. مسلماً چنانچه لازم بود سئوالی درباره یک نکته علمی، مثلاً به روش چندگزینه‌یی، حذف گردد لاجرم سئوال‌ات مربوط به آن نکته علمی که به دو روش دیگر نیز طرح شده بود حذف شد چون در متدولوژی این تحقیق، هر روش کنترل همان نکته علمی با روش‌های دیگر است و حذف یکی به حذف دو سؤال دیگر نیز می‌انجامد. آمار دقیق داده‌ها در جدول ۳ آورده شده است.

تعداد پاسخ‌های درست، غلط و بدون پاسخ در روش‌های مختلف ارزیابی پس از حذف سئوال‌ات با $DI < 0/3$ در نمودار ۳ ارائه گردیده و نشان می‌دهد که با روش صحیح- غلط دانشجویان بیشترین درصد پاسخ درست (۶۵٪) و کمترین درصد پاسخ غلط (۲۶٪) را داشته‌اند، ۹٪ سئوال‌ات را پاسخ نداده و در دو روش دیگر ۸٪ سئوال‌ات بدون پاسخ بوده است.

نمودار ۱- درصد سئوال‌ات با $DI < 0/3$ در دوره‌های تحصیل پزشکی بر حسب روش ارزیابی

میانگین زمان تصحیح اوراق برای هر روش ارزیابی محاسبه و از طرق آزمون‌های Duncan و t-test با هم مقایسه گردید. نتیجه آن در نمودار ۲ آورده شده و نشان داده است که کمترین زمان صرف تصحیح اوراق باروش صحیح- غلط و بیشترین مدت را روش کوتاه پاسخ به خوداختصاص داده است ($p < 0/0001$).

نمودار ۳- توزیع فراوانی پاسخ‌های ۱۳۷۲ نفر دانشجویان پزشکی در ۹ امتحان برگزار شده پس از حذف سئوال‌ات با درجه تمایز کمتر از ۰/۳ - دانشگاه‌های علوم پزشکی شهید بهشتی و ایران ترم دوم ۷۶-۱۳۷۵

جالب توجه اینکه پاسخ‌های مشترک بین روش کوتاه پاسخ با دو روش چندگزینه‌یی و صحیح- غلط توسط آزمون Kappa مقایسه شده است. در جدول ۴ مشاهده می‌شود که تعداد

نمودار ۲- میانگین مدت زمان تصحیح ۱۰ سؤال مربوط به هر یک از روش‌های ارزیابی مورد بررسی در امتحانات دانشجویان پزشکی - دانشگاه‌های علوم پزشکی شهید بهشتی و ایران - ترم دوم ۷۶-۱۳۷۵

با توجه به اینکه کیفیت سئوال‌ات، یک متغیر مداخله‌گر در تحلیل پاسخ‌هاست لذا درجه تمایز هر یک از ۲۷۰ سؤال طرح

داشت. از طرف دیگر چون هدف از این بررسی مقایسه روش‌های ارزیابی است و باید مستقل از اشکالاتی که به طور معمول در طراحی برخی سئوالات بوده انجام شود لذا مقایسه میانگین نمرات دانشجویان پس از حذف سئوالات با $DI < 0/3$ از امتیاز ۲۰ نیز لازم به نظر رسید. نتایج نشان داد که روش کوتاه پاسخ دارای بیشترین همبستگی بوده و به عبارت دیگر از بالاترین اعتبار برخوردار است ($r = 0/685$)، و روش چند گزینه‌یی دارای ضریب

پاسخ‌های درست مشترک بین روش چند گزینه‌یی و کوتاه پاسخ معادل روش صحیح - غلط و کوتاه پاسخ است (۰/۷۸).
 ضمناً پاسخ‌های درست مشترک بین سئوالات چند گزینه‌یی و صحیح - غلط ۴۲۲۷، جواب‌های غلط مشترک بین این دو روش ۱۲۱۲ و بدون پاسخ مشترک ۳۳۵ و ضریب کاپا در این مقایسه $p < 0/000001$ و $0/269$ بوده است.

جدول ۳- توزیع فراوانی سئوالات امتحانی و کل پاسخ‌ها در هر روش پس از حذف سئوالات با درجه تمایز کمتر از ۰/۳

جمع	بالینی			فیزیوپاتولوژی			علوم پایه			
شماره کلاس	۹	۸	۷	۶	۵	۴	۳	۲	۱	
سئوال در هر روش	۶	۸	۷	۷	۷	۷	۶	۵	۷	
کل پاسخها در هر روش	۹۳۱۴	۱۸۹۴	۱۸۵۶	۹۵۲	۸۴۷	۸۹۶	۵۱۱	۱۰۸۶	۴۸۰	۱۷۹۲

جدول ۴- آزمون کاپا برای مقایسه پاسخهای مشترک بین روش‌های چند گزینه‌یی و صحیح - غلط با روش کوتاه پاسخ. درصد پاسخ‌ها در هر روش نسبت به روش کوتاه پاسخ محاسبه شده است.

p	Kappa	ضریب توافقی	بدون پاسخ	غلط	درست	پاسخ‌های مشترک روشهای مورد مقایسه
۰/۰۰۰۰۰۱	۰/۳۳	۰/۳۹۵	۲۱۴ ٪۲۸	۱۸۵۳ ٪۵۱	۳۸۰۱ ٪۷۸	چند گزینه‌یی با کوتاه پاسخ
۰/۰۰۰۰۰۱	۰/۲۵۳	۰/۳۴۳	۲۵۳ ٪۳۳	۱۴۲۶ ٪۳۹	۳۸۱۵ ٪۷۸	صحیح - غلط با کوتاه پاسخ

تعداد کل پاسخها در هر روش ۹۳۱۴

جدول ۵- مقایسه میانگین نمرات خام ۱۳۷۲ نفر دانشجوی پزشکی از امتیاز ۲۰ با سه روش مورد بررسی قبل و بعد از حذف سئوالات با $DI < 0/3$ ، ($p < 0/0001$). ضریب همبستگی هر روش با نمرات دانشجویان در آنالیز فاکتور پس از حذف سئوالات نامطلوب نیز ارائه شده است.

نمره دانشجویان روش ارزیابی	قبل از حذف سئوالات	بعد از حذف سئوالات	
		نمره	ضریب همبستگی در آنالیز فاکتور
کوتاه پاسخ	۱۰/۲	۱۰/۵	۰/۶۸۵
چند گزینه‌یی	۱۱/۷	۱۱/۹	۰/۶۷۷
صحیح - غلط	۱۳	۱۲/۹	۰/۵۶۰

همبستگی میانگین نمرات به آن است ($r = 0/677$) که منعکس کننده اهمیت سئوالات چند گزینه‌یی است. در این آنالیز روش صحیح - غلط در مرتبه اول قرار دارد ($r = 0/560$). میانگین نمرات خام دانشجویان حاصل از هر روش بسیار نزدیک میانگین نمرات پس از حذف سئوالات نامطلوب است (جدول ۵).

بحث:

از آنجا که تا کنون اعتبار سئوالات امتحانی به هر روشی مستقل از روشهای دیگر تعیین گردیده است و با توجه به تعداد زیاد دانشجویان پزشکی و متداول گردیدن سئوالات چند گزینه‌یی که عملاً موجب بروز اشکالاتی از نظر طرح سؤال، گرایش دانشجویان به سؤال خوانی و بخصوص احتمال اجرای برخی ارزیابی‌های ناموفق از معلومات دانشجویان را مطرح نموده است، این تحقیق طراحی گردید که در آن دانشجویان در باره هر مطلب علمی به سه روش چند گزینه‌یی و صحیح - غلط در

پس از حذف سئوالات با درجه تمایز کمتر از ۰/۳، نمرات دانشجویان از امتیاز ۲۰ محاسبه شد و از آنالیز فاکتور نمرات حاصله یک فاکتور به دست آمد که عامل مشترک بین اجزاء تشکیل دهنده آن فاکتور مربوط به دانش دانشجویان تلقی میگردد و مسلماً متغیری که ضریب همبستگی بالاتری با آن داشته باشد اهمیت بیشتری در ارزیابی میزان این معلومات خواهد

نفر دانشجوی پزشکی گزارش نموده است و ۸۵/۳٪ آنان با روش صحیح- غلط بیشترین نمره را کسب نموده‌اند (۱۱)، منطبق بوده است. البته اگر در تفسیر آن، این نکته از ذهن بگذرد که در روش صحیح- غلط با استفاده از شانس تعداد پاسخ‌های صحیح بیشتر شده است باید گفت که بیشترین سئوالات بدون پاسخ (۹٪) نیز در این روش دیده شد. لذا به نظر می‌رسد کم بودن تعداد پاسخ‌های درست در روش چندگزینه‌یی در ارتباط با مشاهده عبارات غلط با تواتر سه برابر تعداد جملات درست است و درصد سئوالات بدون پاسخ نیز در این روش معادل روش کوتاه پاسخ است که نداشتن نمره منفی در آن محرز بوده است. بنابراین مشکلی که در روش چندگزینه‌یی می‌توان عنوان کرد این است که با وجود داشتن نمره منفی، تعدد گزینه‌های غلط دانشجوی را از ندادن پاسخ باز نمی‌دارد و با استفاده از شانس آنها را نیز به عنوان جواب صحیح انتخاب می‌کند بدون اینکه در حقیقت پاسخ صحیح را بداند.

با وجود نامأنوس بودن دانشجویان مورد بررسی با روش کوتاه پاسخ، نظر به اینکه این آزمون معتبرترین روش برای ارزیابی معلومات دانشجوی نسبت به دو روش دیگر است (۹-۱۱)، لذا این روش به عنوان کنترل در نظر گرفته شده است و تعداد پاسخ‌های درست، غلط و بدون پاسخ که با روش‌های چندگزینه‌یی و صحیح- غلط مشترک می‌باشند از طریق آزمون Kappa تعیین گردیده است که تفسیری جالب را در پی دارد و آن معادل بودن تعداد پاسخ‌های درست مشترک بین دو روش با روش کوتاه پاسخ است و تفاوت اساسی بین دو روش اخیر یکی در زیادتیر بودن پاسخ‌های غلط مشترک در روش چندگزینه‌یی نسبت به روش صحیح- غلط بوده (۵۱٪ در برابر ۳۹٪) و دیگری بیشتر بودن تعداد سئوالات بدون پاسخ مشترک با روش صحیح- غلط (۳۳٪ در برابر ۲۸٪) می‌باشد. ضرایب کاپا اختلاف کمی با یکدیگر داشته و هر دو در محدوده توافقی "fair" قرار دارند، مهم اینکه در آزمون کاپا به نظر می‌رسد تفاوت‌های مقادیر کاپا در وجه اشتراک پاسخ‌های هر یک از دو روش با روش کوتاه پاسخ ناشی از پاسخ‌های صحیح که قاعدتاً معیار ارزیابی معلومات دانشجویست نمی‌باشد، و پاسخ‌های غلط و سئوالات بدون پاسخ در این نتیجه دخالت داشته‌اند (جدول ۴). انجام آزمون کاپا نشان می‌دهد که روش‌های چندگزینه‌یی و صحیح- غلط بیشترین ضریب توافقی را با یکدیگر دارا می‌باشند (۰/۴۰۲). در تحقیق Wass نیز همبستگی روش کوتاه پاسخ با چندگزینه‌یی ۰/۵۶ و با صحیح- غلط ۰/۴۶ گزارش گردیده است (۱۱)، ولی نظر ۸۱٪ دانشجویان دانشکده پزشکی نیوکاسل انگلستان درباره انواع آزمون‌های انجام

مقایسه با روش کوتاه پاسخ به عنوان کنترل ارزیابی گردیدند و تعداد پاسخ‌های درست، غلط و بدون پاسخ در هر یک از این روش‌ها حاصل شد. لذا به سبب جدید بودن اهداف این بررسی از نظر متدولوژی، مقایسه نتایج با اصول موجود در کتاب‌های ارزیابی آموزشی در بحث اعتبار روش‌ها امکان‌پذیر است کما اینکه آزمون‌های کوتاه پاسخ از اعتبار بالایی برخوردار بوده و روش‌های عینی دیگر از جمله سئوالات چندگزینه‌یی و صحیح- غلط در درجه بعد به طور نسبی معتبرند (۱۰ و ۹) که در این تحقیق نیز همین نتیجه حاصل گردید.

با توجه به اینکه سئوالات امتحانی در ۹ کلاس درس مربوط به سه دوره تحصیل پزشکی چه از نظر طراحی توسط استادان مربوطه و چه از نظر دانشجویان یک روند کاملاً عادی و رسمی را دارا بود لذا در بررسی کیفیت سئوالات مشاهده شد که درجه آسانی ($FI > 70\%$) و سختی سئوالات ($FI < 30\%$) در دوره‌های مختلف علوم پایه، فیزیوپاتولوژی و بالینی با هم متعادل نبوده و در جدول ۲ مشاهده می‌شود که در دوره بالینی نیمی از سئوالات (۴۵ از ۹۰) آسان بوده که بیشتر آنها به روش صحیح- غلط بود. لذا این نکته هر چند در ارزیابی کلی چندان مؤثر نبود ولی توجه به آن در هنگام طرح سئوال به روش فوق ضروری به نظر می‌رسد بخصوص که این روش بیشترین تعداد سئوال را با درجه تمایز کمتر از ۰/۳ نیز در دوره بالینی تشکیل داده است (۸ از ۹۰)، (نمودار ۱).

با اینکه پاسخگویی به آزمون‌های عینی سریع است (۱۰)، همانطور که پیش بینی می‌شد میانگین مدت زمان لازم برای تصحیح اوراق نشان داد که روش صحیح- غلط کمترین زمان را به خود اختصاص داده (۲/۳ ± ۵/۲ دقیقه) و از نظر آماری اختلاف معنی‌داری با روش چندگزینه‌یی (۵/۴ ± ۱۳/۱ دقیقه) داشت که خود با روش کوتاه پاسخ (۷/۹ ± ۱۴/۱ دقیقه) اختلاف معنی‌داری داشت و همانطور که انتظار می‌رفت روش اخیر بیشترین زمان را مصروف تصحیح اوراق نمود ($p < 0.001$)، (نمودار ۲).

قابل ذکر است که پاسخ به سئوالات کوتاه پاسخ نسبتاً مشکل بوده و در ارتباط با نداشتن نمره منفی بیشترین جواب غلط را دارد ولی آنچه شایان توجه است اینکه تعداد غلط در سئوالات با روش صحیح- غلط ۲۶٪ ولی در سئوالات چندگزینه‌یی ۳۲٪ می‌باشد و به عکس تعداد پاسخ‌های درست در سئوالات صحیح- غلط ۶۵٪ است که بیشترین میزان پاسخ درست می‌باشد و با نتایج بررسی Wass که در آن پاسخ سئوالات طراحی شده با روش‌های مختلف راطی ۸ ساعت امتحان از ۲۱۴

بهمین تجاسب که در تحلیل آماری این تحقیق نهایت همفکری را فرموده‌اند نیز تقدیر می‌گردد.

انجام این تحقیق با همکاری بی‌شائبه اساتید گرانقدر دانشگاه‌های علوم پزشکی شهید بهشتی و ایران سرکار خانم

دکتر مریم السادات حسینی و جناب آقایان دکتر حسین الیاسی، دکتر پرویز پاکزاد، دکتر علیرضا سالک‌مقدم، دکتر بهروز شفقی، دکتر منوچهر شیروانی، دکتر جعفر فرقانی‌زاده، دکتر کوروش قره‌گزلی و دکتر همایون همایون‌فر امکان پذیر گردید که با صرف وقت در طرح سئوالات و پشتیبانی از رسمیت دادن به اجرای تحقیق به آن اعتبار بخشیدند متواضعانه قدردانی نموده و بخصوص از جناب آقای پروفیسور هانری والتون از دانشگاه ادینبورگ انگلستان و پرزیدنت سابق فدراسیون جهانی آموزش پزشکی نیز برای نقد پیش‌نویس انگلیسی این مقاله و توصیه‌های ارزشمندی که فرموده‌اند بسیار سپاسگزاریم.

شده از آنان این بوده است که روش‌های عینی را عادلانه‌ترین انواع سئوال دانسته‌اند بدون اینکه بین آنها تمایزی قایل شوند (۱۲).

نتیجه‌گیری:

با توجه به بالاترین ضریب توافقی بین روش‌های صحیح-غلط و چند گزینه‌یی و اعتبار قابل توجه روش کوتاه پاسخ، پیشنهاد می‌شود با استفاده همزمان از این سه روش، اعتبار ارزیابی دانشجویان را ارتقا بخشیده تا در عین حال بتوان اثرات ناخواسته سئوالات چند گزینه‌یی را به حداقل رسانید و میزان یادگیری و یادداری دانشجویان را تماماً ارزیابی نمود.

تشکر و قدردانی:

وظیفه داریم از معاونت محترم پژوهشی وزارت بهداشت، اعضای محترم شورای پژوهشی وزارت متبوع، معاونین محترم آموزشی دانشگاه‌های علوم پزشکی شهید بهشتی و ایران و مسئولین محترم دایره امتحانات دانشگاه‌های مزبور جهت تصویب، پی‌گیری، صدور مجوزها و نظارت در حسن انجام امتحانات صمیمانه سپاسگزاری نمائیم. از دقت نظر جناب آقای

References: منابع

- 1- Hammond EJ, McIndoe AK, Sansome AJ, et al. Multiple-choice examinations: adopting an evidence-based approach to exam technique. *Anaesthesia* 1998; 53(11): 1105-8.
- 2- 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): 74-9.
- 3- Schultheis NM. Writing cognitive educational objectives and multiple-choice test questions. *Am J Health Syst Pharm* 1998; 15; 55(22): 2397-401.
- 4- A critical review of student assessment options. Available at: <http://www.nmu.edu/soa/review.html>.
- 5- Specific Student Assessment Techniques, in *Student Evaluation: A Teacher Handbook*; December 1991. Available at: <http://www.sasked.gov.sk.ca/docs/policy/studeval/>.
- 6- Frohlich ED. *Medical Qualifying Examination: Rypins' Questions & Answers for Basic Sciences Review*, 4th ed. Philadelphia.: Lippincott Williams & Wilkins; 2001: 4.



- 7- Rassaian N, Ghandehari NS, Nakhaei S, et al. Attitude and academic performance of medical students in research-centered teaching method. *Med J Islam Rep Im* 2000; 14(3): 253- 60.
- 8- Rassaian N. Long-term memory and learning through the use of research-centered teaching method. *J Med Edu* 2001; 1(1): 38-42.
- 9- Cox KR, Ewan CE. *The Medical Teacher*, Churchill Livingstone; 1982: 193-218.
- 10- McAleer S. Objective testing: Dent JA, Harden RM (editors), *A Practical Guide for Medical Teachers*. London: Churchill Livingstone; 2001: 314-25.
- 11- Wass V, McGibbon D, Van der Vleuten C. Composite undergraduate clinical examinations: How should the components be combined to maximize reliability? *Med Edu* 2001; 35(4): 326-30.
- 12- Duffield KE, Spencer JA. A survey of medical students' views about the purposes and fairness of assessment. *Med Edu* 2002; 36(9): 879-86.

^۱ - دکترای ویروس شناسی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران

^۲ - گواهی عالی بهداشت، اداره کل پیشگیری و مبارزه با بیماریها

Archive of SID