



Quantitative and Qualitative Evaluation of Four Choice Questions In Shahroud University of Medical Sciences During 2021-2022

S. Paryab (MSc)¹, R. Zarouj Hosseini (MSc)², S. Zamani (PhD)³, M. Lashkari (MSc)⁴, M. Yousefi
(MSc)⁴, K. Zarouj Hosseini (BS)⁵, O. Garkaz (MSc)^{6*}

1. School of Nursing and Midwifery, Shahroud University of Medical Sciences, Shahroud, Iran.
2. School of Allied Medical Sciences, Shahroud University of Medical Sciences, Shahroud, Iran.
3. Student Research Committee, School of Medicine, Shahroud University of Medical Sciences, Shahroud, Iran.
4. Vice Chancellor for Education, Shahroud University of medical sciences, Shahroud, Iran.
5. Student Research Committee, School of Allied Medical Sciences, Shahroud University of Medical Sciences, Shahroud, Iran.
6. Department of Epidemiology, School of Public Health, Shahroud University of Medical Sciences, Shahroud, Iran.

Article Info

ABSTRACT

Article Type:

Research Article

Received:

12 April 2023

Revised:

30 April 2023

Accepted:

11 July 2023

Published online:

16 August 2023

Background and Objective: Any test as a measurement tool must have sufficient validity and reliability to measure the desired attribute. Multiple-choice tests are the most common types of tests in medical education, which have a high degree of reliability, and this study was conducted with the aim of quantitative and qualitative evaluation of four-choice questions in Shahroud University of Medical Sciences in the academic year of 2021-2022.

Methods: This descriptive study was conducted on all General and specialized courses in the two academic years of 2021-2022 at the Shahroud University of Medical Sciences. For the quantitative evaluation of the questions (difficulty index and discrimination index), the Hamava system was used and for the qualitative evaluation, the 14-item checklist of Millman was used.

Findings: This study showed that out of 5000 examined questions, 2569 questions (51.4%) had a simple difficulty coefficient and 2500 (50%) had a suitable clean coefficient. And the average standard deviation of the degree of difficulty was 0.700 ± 0.314 and the degree of cleanliness was 0.625 ± 0.247 , also 330 questions (48.5%) of taxonomy two and 190 questions (27.9%) were without defects. The most types of defects were the absence of most of the information in the question stems of 196 questions, the presence of negative words in the question stems of 136 questions, and the presence of options for all items in 121 questions, none in 45 questions, and combined options in 34 questions. And in the structure of question 398 (58/1%) was the most type of defect.

*Corresponding Author: O. Garkaz(MSc)

Address Shahroud University of Medical Sciences, Shahroud, Iran.

Tel: +98 (02332395054) E-mail: omid.garkaz23@gmail.com

Conclusion: To improve the quality of tests and use the results of quantitative and qualitative analysis of questions, practical solutions should be thought of, and professors should be trained in designing and using appropriate questions and preparing for-choice question banks.

Keywords: *Analysis, Quantitative, Qualitative, and Four-Option Questions*

Cite this article: S. Paryab, R. Zarouj Hosseini, S. Zamani, M. Lashkari, M. Yousefi, K. Zarouj Hosseini, O. Garkaz. Quantitative and Qualitative Evaluation of Four Choice Questions In Shahroud University of Medical Sciences 2021-2022. Medical Education Journal. 2023; 12: e2.



© The Author(s).

Publisher: Babol University of Medical Sciences

ارزیابی کمی و کیفی سؤالات چهارگزینه‌ای در دانشگاه علوم پزشکی شاهرود (۱۳۹۹-۱۴۰۰)

سحر پاریاب^۱(MSc)، ربابه زروج حسینی^۲(MSc)، سپهر زمانی^۳(PhD)، معصومه لشکری^۴(MSc)، مریم یوسفی^۵(MSc)، کیمیا زروج حسینی^۶(MSc)؛ امید گرکز^۶(MSc)

۱. دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی شاهرود، شاهرود، ایران.

۲. دانشکده پیراپزشکی، دانشگاه علوم پزشکی شاهرود، شاهرود، ایران.

۳. کمیته تحقیقات دانشجویی، دانشکده پزشکی، دانشگاه علوم پزشکی شاهرود، شاهرود، ایران.

۴. معاونت آموزشی، دانشگاه علوم پزشکی شاهرود، شاهرود، ایران.

۵. کمیته تحقیقات دانشجویی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شاهرود، شاهرود، ایران.

۶. گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی شاهرود، شاهرود، ایران.

چکیده

اطلاعات مقاله

سابقه و هدف: هر آزمونی به‌عنوان ابزار سنجش و اندازه‌گیری باید دارای روایی و پایایی کافی برای سنجش صفت موردنظر باشد. آزمون‌های چندگزینه‌ای رایج‌ترین انواع آزمون‌ها در آموزش پزشکی می‌باشند که از درجه پایایی بالایی برخوردارند. این مطالعه باهدف ارزیابی کمی و کیفی سؤالات چهارگزینه‌ای در دانشگاه علوم پزشکی شاهرود در سال‌های تحصیلی ۱۴۰۰-۱۳۹۹ انجام شد.

نوع مقاله:

مقاله پژوهشی

مواد و روش‌ها: این مطالعه به‌صورت توصیفی - تحلیلی بر روی آزمون‌های پایان ترم چهارگزینه‌ای اختصاصی و عمومی در دو سال تحصیلی ۱۴۰۰-۱۳۹۹ در دانشگاه علوم پزشکی شاهرود انجام شد که برای ارزیابی کمی سؤالات (درجه تمیزی و سختی) از سامانه هم آوا و ارزیابی کیفی از چک‌لیست ۱۴ گانه میلن استفاده شد. داده‌ها بعد از جمع‌آوری وارد SPSS ۱۸ شده و با کمک آمار توصیفی و تحلیلی تجزیه و تحلیل گردید.

دریافت:

۱۴۰۲/۱/۲۳

اصلاح:

۱۴۰۲/۲/۱۰

پذیرش:

۱۴۰۲/۴/۲۰

انتشار:

۱۴۰۲/۵/۲۵

یافته‌ها: نتایج نشان داد که از ۵۰۰۰ سؤال مورد بررسی، سؤال ۲۵۶۹ (۵۱/۴٪) ضریب دشواری ساده و ۲۵۰۰ (۵۰٪) ضریب تمیز مناسب داشتند و میانگین انحراف معیار درجه سختی $۰/۷۰۰ \pm ۰/۳۱۴$ و درجه تمیزی $۰/۶۲۵ \pm ۰/۲۴۷$ بود. همچنین ۳۳۰ سؤال (۴۸/۵٪) تاکسونومی دو و ۱۹۰ سؤال (۲۷/۹٪) بدون نقص بود. بیشترین نوع نقص به ترتیب عدم وجود بخش اعظم اطلاعات در ساقه سؤال (۱۹۶ مورد)، وجود لغات منفی در ساقه سؤال (۱۳۶ مورد) و وجود گزینه‌های همه موارد (۱۲۱ سؤال)، هیچ کدام (۴۵ مورد)، گزینه‌های ترکیبی (۳۴ سؤال) بود و در ساختار سؤال ۳۹۸ مورد (۵۸/۱٪) بیشترین نوع نقص بود. **نتیجه‌گیری:** برای بهبود کیفیت آزمون‌ها و استفاده از نتایج تجزیه و تحلیل کمی و کیفی سؤالات باید راهکارهای عملی اندیشیده شود و اساتید در زمینه طراحی و استفاده از سؤالات مناسب و تهیه بانک سؤالات چهارگزینه‌ای آموزش داده شوند.

واژه‌های کلیدی: ارزیابی کمی، ارزیابی کیفی، سؤالات چهارگزینه‌ای

استناد: سحر پاریاب، ربابه زروج حسینی، سپهر زمانی، معصومه لشکری، مریم یوسفی، کیمیا زروج حسینی، امید گرکز. ارزیابی کمی و کیفی سؤالات چهارگزینه‌ای در دانشگاه علوم پزشکی شاهرود (۱۳۹۹-۱۴۰۰). مجله آموزش پزشکی. ۱۴۰۲؛ ۱۲: ۵۲.



© The Author(s)

Publisher: Babol University of Medical Sciences

این مقاله مستخرج از طرح تحقیقاتی به شماره ۱۴۰۰۵۶ و کد اخلاق IR.SHMU.REC.1400.219 دانشگاه علوم پزشکی شاهرود می باشد.

* مسئول مقاله: امید گرکز
آدرس: ایران، شاهرود، دانشگاه علوم پزشکی شاهرود.

رایانامه: omid.garkaz23@gmail.com

تلفن: ۰۲۳۳۲۳۹۵۰۵۴

سابقه و هدف

ارزیابی یا ارزشیابی فرآیند نظام‌مند گردآوری تحلیل و تفسیر داده‌ها به منظور ارزش‌گذاری و قضاوت می‌باشد. اگر این فرآیند در مورد افراد، گروه‌ها و سازمان انجام گردد، معمولاً به عنوان ارزیابی و اگر در مورد برنامه انجام شود، به آن ارزشیابی اطلاق می‌گردد. آزمون‌ها، از انواع رایج ارزیابی هستند که در سیستم آموزشی صورت می‌گیرند. آزمون‌ها از نظر محتوای آزمون (دانش، مهارت و ...)، هدف آزمون (پیشرفت تحصیلی، گزینش داوطلبان و ...)، معیار قبولی و جنبه‌های گوناگون دیگر قابل طبقه‌بندی می‌باشند. برای آن‌که هر آزمون نتایج قابل استناد داشته باشد، باید از روایی، پایایی و قابلیت اجرا برخوردار باشد که هر یک جنبه‌های متفاوتی از آزمون را در برمی‌گیرند (۱ و ۲). آزمون‌های کتبی هسته مرکزی ارزشیابی دانشجویان را تشکیل می‌دهد. از این میان به خصوص آزمون‌های چندگزینه‌ای (MCQs) به خاطر سهولت و عینیت در نمره دهی، بیشتر مورد استفاده قرار می‌گیرند. اغلب اوقات حتی قبل از اجرای آزمون‌های بالینی، یک آزمون چندگزینه‌ای به منظور اطمینان از دانش پایه‌ای دانشجویان برگزار می‌شود (۳-۶). از طرف دیگر پایاترین روش مورد استفاده برای اندازه‌گیری دانش فراگیران آزمون چندگزینه‌ای است. سؤالات چندگزینه‌ای خوب طراحی شده، فراتر از یک یادآوری ساده اطلاعات هستند. چنین سؤالاتی آزمون شونده را مجبور می‌کنند که برای یک مشکل کلینیکی، به تحلیل و نتیجه‌گیری از مطالب تئوری ذهن خود پردازد (۷). این سؤالات می‌تواند برای سنجش اغلب سطوح یادگیری به کار برده شود. همچنین توانایی اندازه‌گیری بیشتر اهداف آموزشی را دارا می‌باشد (۸ و ۹)، از طرف دیگر آزمون‌های چهارگزینه‌ای توانایی ارزیابی تعداد زیادی دانشجو با به کارگیری حداقل نیروی انسانی را دارد (۱۰). هر سؤال چندگزینه‌ای شامل چند قسمت است: ۱- قسمت اصلی یا تنه یا ساقه سؤال: این بخش از سؤال، متن اصلی سؤال را تشکیل می‌دهد و دربرگیرنده موضوعی است که سؤال باید آن را اندازه‌گیری کند. ۲- گزینه درست یا پاسخ سؤال: یکی از گزینه‌های پیشنهادی پاسخ درست سؤال است که آزمون شونده باید آن را برگزیند که به آن (گزینه کلید) هم گفته می‌شود. ۳- گزینه‌های انحرافی: به غیر از پاسخ درست سؤال، تعدادی گزینه دیگر نیز برای هر سؤال تهیه می‌شود که به آن گزینه‌های انحرافی می‌گویند. نقش گزینه‌های انحرافی، منحرف کردن آزمون شدگانی است که پاسخ درست سؤال را نمی‌دانند. به‌طور کلی، متن اصلی یا تنه سؤال را یک جمله استفهامی یا یک جمله ناتمام تشکیل می‌دهد و گزینه‌های پیشنهادی، جواب جمله استفهامی یا تکمیل‌کننده جمله ناقص هستند، توانایی در نوشتن سؤالات چندگزینه‌ای خوب، نوعی هنر است (۱۱، ۱۲).

مطالعات داخلی مختلفی در علوم پزشکی به تجزیه و تحلیل آزمون‌های چندگزینه‌ای پرداخته‌اند. مطالعات حاکی از آن است که با انجام مداخلات آموزشی وضعیت طراحی سؤالات، بهبودی قابل توجهی می‌یابد. به‌طور مثال، مطالعه معیاری و همکاران، تفاوت آماری معنی‌داری را در ارتقای کیفیت طراحی سؤالات، از دو منظر تاکسونومی و ساختار پس از برگزاری کارگاه آموزشی جهت استادان جوان طراح سؤالات در آزمون‌های ارتقای دستیاری دندان پزشکی نشان داد. وحید شاهی و همکاران نیز نشان دادند که با انجام مداخله آموزشی از نوع ارائه بازخورد و متون تحصیلی و آموزشی به دبیران و اعضای هیات‌های ممتحنه، می‌توان تأثیر مثبتی بر طراحی سؤالات آزمون‌های ارتقای دستیاری گذاشت (۱۳ و ۱۴). با توجه به آزمون‌های تستی استاندارد مختلف و مزایای آن‌ها که در کتاب‌های درسی آموزشی آمده است، گاهی اصول طراحی این نوع سؤالات رعایت نمی‌شود. سؤالاتی که در امتحانات بورد آمریکا استفاده می‌شود عمدتاً با روش چندگزینه‌ای است و متخصصین مربوطه پس از بازنگری و بحث‌های موشکافانه طولانی، سؤالات را انتخاب کرده و بقیه سؤالات را حذف می‌کنند که خود نشان‌دهنده مشکل بودن طراحی این گونه سؤالات است و از طرف دیگر با منسوخ شدن تخصیص نمره منفی به این نوع سؤالات، ارتباط قبولی در امتحان با دانش فراگیران ضعیف به نظر می‌رسد. نکته حائز اهمیت آن است که گذراندن امتحانات، کلاس‌های درس را قاعدتاً محل انتقال اندیشه، تحلیل و تفسیر اطلاعات و حتی آموزش و تحقیق قرار داده است (۱۵-۱۷). از طرف دیگر پایاترین روش مورد استفاده برای اندازه‌گیری دانش فراگیران، آزمون‌های چندگزینه‌ای است. سؤالات چندگزینه‌ای خوب طراحی شده، فراتر از یک یادآوری ساده اطلاعات هستند. چنین سؤالاتی آزمون شونده را مجبور می‌کند که برای یک شکل کلینیکی

سؤال به تحلیل و نتیجه‌گیری از مطالب تئوری ذهن خود پردازد. این سؤالات می‌تواند برای سنجش اغلب سطوح یادگیری به‌کاربرده شود. در اغلب دانشگاه‌های علوم پزشکی ایران هنوز ارزشیابی دانشجویان حتی در اغلب بخش‌های بالینی هم توسط آزمون‌های کتبی و به خصوص سؤالات چندگزینه‌ای انجام می‌شود (۱۸-۲۱). طی سال‌های متوالی کیفیت پایین این آزمون‌ها همواره به چشم می‌خورد، آگاهی کم اساتید از اصول طراحی سؤال خوب و عدم بازخورد مناسب به اساتید و فرصت کم ایشان برای بررسی سؤالات، از علل این چالش به نظر می‌رسد. لذا از آنجایی که تاکنون بررسی سؤالات چهارگزینه‌ای چه از نظر کیفی چه از نظر کمی در دانشگاه علوم پزشکی شاهرود صورت نگرفته است، محققین را بر آن داشت که به بررسی کیفی و کمی سؤالات چهارگزینه‌ای در این دانشگاه پردازند.

مواد و روش‌ها

این مطالعه به صورت توصیفی - تحلیلی بر روی ۵۰ درس اختصاصی و عمومی از تمام رشته‌ها شامل (پزشکی، پرستاری، مامایی، هوشبری، اتاق عمل، رادیولوژی، فناوری اطلاعات سلامت، فوریت پزشکی، بهداشت حرفه‌ای، بهداشت عمومی، بهداشت محیط، اپیدمیولوژی، نانوفناوری، بیوتکنولوژی) در دو سال تحصیلی ۱۳۹۹-۱۴۰۰ در دانشگاه علوم پزشکی شاهرود انجام شد. جهت دسترسی به اطلاعات موردنظر، هماهنگی‌های لازم از طریق معاونت تحقیقات و فناوری دانشگاه با معاونت آموزشی دانشگاه صورت گرفت. معیار ورود مطالعه شامل موارد زیر است: سؤالات چهارگزینه‌ای باشد / سؤالات یک جواب درست داشته باشد / مربوط به آزمون‌های کتبی باشد / سؤالات مربوط به آزمون دروس اصلی و تخصصی باشد و معیار خروج مطالعه سؤالات تشریحی بود. برای تعیین نمره کمی سؤالات یعنی دشواری (فوق‌العاده دشوار نمره کمتر از ۰/۲، خیلی دشوار نمره ۰/۲-۰/۴، دشوار نمره ۰/۴-۰/۶، متوسط نمره ۰/۶-۰/۸، ساده نمره ۰/۸-۱) و تمیزی سؤالات (مناسب، کم و نامناسب) که سال‌های موردنظر از سامانه هم‌آوا استفاده شد که در این حالت اطلاعات ۵۰۰۰ هزار سؤال شامل (نوع دروس، سال، دانشکده، درجه دشواری، نمره دشواری، درجه تمیزی و نمره تمیزی، مقطع تحصیلی) از سامانه هم‌آوا کسب شد و برای تعیین نمره کیفی سؤالات نیز از چک‌لیست مبتنی بر اصول ۱۴ گانه میلن برای طراحی ساقه، گزینه‌های سؤال و تاکسونومی بلوم که شامل سه سطح (تاکسونومی سطح یک: سؤالی که تنها بر مبنای حافظه باشد؛ شایع‌ترین تظاهر بیماری موردنظر کدام است / تاکسونومی سطح دو: وقتی ذهن خود با انجام یک فرایند به پاسخ می‌رسد؛ تظاهرات بیماری (علائم و نشانه‌ها) داده شود و تشخیص خواسته شود. البته اگر مورد فوق قبلاً در کلاس درس مطرح شده باشد و دانشجو بتواند با حافظه به آن پاسخ دهد، سطح آن یک می‌شود. تاکسونومی سطح سه: وقتی ذهن با انجام دو مرحله فرایند به پاسخ می‌رسد) که در این حالت نیز اطلاعات ۳۰ درس شامل ۶۸۰ سؤال مورد ارزیابی قرار گرفت. در محله بعد، داده‌ها وارد نرم‌افزار SPSS18 گردید و با کمک آزمون‌های آمار توصیفی (فراوانی، درصد، میانگین و انحراف معیار) و آمار تحلیلی (کای دو) تجزیه و تحلیل گردید. در این مطالعه، نکات اخلاقی بیانیه هلسینکی از جمله محرمانه ماندن اطلاعات، رعایت صداقت در انتخاب واحدهای مورد پژوهش و جمع‌آوری و تجزیه و تحلیل داده‌ها، ارائه نتایج پژوهش به واحدهای مورد پژوهش و مسئولین محیط پژوهش در صورت درخواست و قدردانی از کلیه واحدهای مورد پژوهش و کلیه مسئولین ذی‌ربط که همکاری داشته‌اند، رعایت شد. مطالعه حاضر با کد اخلاق IR.SHMU.REC.1400.219 در دانشگاه علوم پزشکی شاهرود تصویب گردید و تمام اصول اخلاقی در انتشار نتایج در نظر گرفته شد.

یافته‌ها

در این مطالعه نتایج نشان می‌دهد که از ۵۰۰۰ سؤال مورد بررسی، ۴۲۷۵ سؤال تخصصی (۸۵/۵٪)، ۳۰۹۳ سؤال (۶۱/۹٪) مربوط به سال ۱۳۹۹، ۲۰۴۶ سؤال (۴۰/۹٪) مربوط به دانشکده پزشکی و ۱۹۳۹ سؤال (۳۸/۸٪) درصد) مربوط به مقطع دکترای عمومی بود (جدول ۱). همچنین ۲۵۶۹ سؤال (۵۱/۴٪) ضریب دشواری ساده و ۲۵۰۰ سؤال (۵۰٪) درصد) ضریب تمیز مناسب داشتند و میانگین انحراف معیار درجه سختی $0/700 \pm 0/314$ و درجه تمیزی $0/625 \pm 0/247$ بود (جدول ۲). همچنین ۳۳۰ سؤال (۴۸/۵٪) در سطح تاکسونومی دو، ۱۹۰ سؤال (۲۷/۹٪) بدون نقص بود. بیش‌ترین نوع نقص به ترتیب، عدم وجود بخش اعظم اطلاعات در ساقه سؤال (۲۸/۱٪/۱۹۶)، وجود لغات منفی در ساقه سؤال (۲۰٪/۱۳۶) و وجود گزینه‌های همه موارد، هیچ‌کدام، گزینه‌های ترکیبی (۱۷/۸٪/۱۲۱) مورد بود و در ساختار سؤال (۵۸/۱٪/۳۹۸) بیش‌ترین نوع نقص بود. (جدول ۳). بین سطح شناختی سؤالات و نقص در ساختار سؤالات رابطه معنی‌داری وجود داشت ($P=0.001$).

جدول ۱. اطلاعات کلی سؤالات

متغیر	زیر گروه	فراوانی (درصد)
نوع دروس	عمومی	۱۴/۵)۷۲۵
	تخصصی	۸۵/۵)۴۲۷۵
سال	۱۳۹۹	۶۱/۹)۳۰۹۳
	۱۴۰۰	۳۸/۱)۱۹۰۷
دانشکده	پزشکی	۴۰/۹)۲۰۴۶
	پرستاری و مامایی	۲۷/۶)۱۳۸۲
	پیراپزشکی	۱۷/۸)۸۸۸
	بهداشت	۷/۲)۳۶۰
مقطع	کارشناسی	۴۲/۴)۲۱۲۱
	ارشد	۱۰/۹)۵۴۶
	دکترای تخصصی	۱/۴)۷۰
	عمومی	۳۸/۸)۱۹۳۹

جدول ۲. ضریب دشواری و تمیز سؤالات

نوع متغیر	زیر گروه	فراوانی (درصد)
ضریب دشواری سؤالات	فوق العاده دشوار	۱۱/۳)۵۶۷
	خیلی دشوار	۷/۹)۳۹۶
	دشوار	۱۱/۸)۵۹۲
	متوسط	۱۷/۵)۸۷۶
	ساده	۵۱/۴)۲۵۶۹

(۵۰/۲۵۰۰)	مناسب	ضریب تمیز سؤالات
(۲۹/۷۴)۱۴۸۷	کم	
(۲۰/۲۶)۱۰۱۳	نامناسب	

جدول ۳. نتایج بررسی سؤالات از نظر چکلیست میلمن

نوع متغیر	زیر گروه	فراوانی (درصد)
تاکسونومی	تاکسونومی یک	(۱۳/۱)۲۶۰
	تاکسونومی دو	(۴۸/۵)۳۳۰
	تاکسونومی سه	(۳۸/۴)۹۱
تعداد نقص‌ها	بدون نقص	(۲۷/۹)۱۹۰
	یک نقص	(۲۶/۳)۱۷۹
	دو نقص	(۱۹/۸)۱۳۵
	سه نقص	(۱۵/۵)۱۰۶
نوع نقص	چهار نقص	(۱۰/۵)۷۰
	عدم وجود بخش اعظم اطلاعات در ساقه سؤال	(۲۸/۸)۱۹۶
	وجود گزینه منفی برای ساقه منفی	(۶/۶)۴۵
	وجود لغات منفی در ساقه سؤال	(۲۰)۱۳۶
	وجود لغات منفی در گزینه‌ها	(۵)۳۴
	وجود گزینه‌های همه موارد، هیچ کدام، گزینه‌های ترکیبی	(۱۷/۸)۱۲۱
	وجود گزینه‌های متضاد	(۱/۷)۱۲
	عدم همبستگی گزینه‌ها از نظر طول و ساختار بندی	(۳/۲)۲۲
	کاربرد عبارات تکراری در گزینه‌ها	(۲/۳)۱۶
	عدم مطابقت سؤال با اهداف دوره	(۸/۸)۶۰
بخش دارای نقص	عدم وجود فقط یک گزینه صحیح	(۵/۵)۳۸
	ساختار سؤال	(۵۸/۱)۳۹۸
	تنه سؤال	(۳۴/۱)۲۳۲
	ساقه سؤال	(۷/۸)۵۰

بحث و نتیجه گیری

ارزشیابی از مهم‌ترین بخش‌های هر فرآیند آموزشی است. انجام صحیح ارزشیابی می‌تواند باعث ایجاد انگیزه یادگیری در دانشجویان و بازخورد روند آموزش به مدرس شود. از آنجاکه آزمون‌های چندگزینه‌ای از سایر آزمون‌های عینی انعطاف‌پذیرتر می‌باشند و همچنین علاوه بر دانش، توانایی استدلال، قضاوت و بسیاری از بازده‌های مهم دیگر را می‌سنجند. از طرفی با استفاده از سؤال‌های چندگزینه‌ای می‌توان در یک‌زمان محدود، تعداد زیادی از هدف‌های آموزشی و بخش مهمی از محتوای درس را اندازه‌گیری نمود. اگر گزینه‌های انحرافی این آزمون‌ها با توجه به اشتباهات و کج فهمی‌های متداول دانشجویان تهیه شود، منبع بسیار مناسبی برای تشخیص مشکلات

دانشجویان خواهد بود (۲۲). نتایج آزمون‌های چندگزینه‌ای قابل تجزیه و تحلیل کردن است و منجر به استانداردسازی سؤالات، رعایت عدالت در نمره دهی آزمون و افزایش انگیزه دانشجویان به مطالعه و یادگیری می‌شود. همه این محاسن باعث گردیده است که تمایل اساتید در استفاده از این شیوه ارزیابی در آزمون‌های پیشرفت تحصیلی افزایش یابد (۲۳). نتایج مطالعه حاضر نشان داد که بیشتر سؤالات ۲۵۶۹ مورد ضریب دشواری ساده (۵۱/۴٪) و ۲۵۰۰ مورد ضریب تمیز مناسب (۵۰٪) داشتند و میانگین انحراف معیار درجه سختی سؤالات ۰/۷۰۰±۰/۳۱۴ و درجه تمیز ۰/۶۲۵±۰/۲۴۷ بود که با مطالعه بقایی و همکاران در سال ۱۳۹۵ (۲۴)، خواجه علی و همکاران در سال ۲۰۲۰ (۲۵) و وفا مهر و همکاران در سال ۲۰۱۱ (۲۶) و احمد و همکاران در سال ۲۰۱۲ هم‌خوانی دارد که بیانگر اهمیت توجه اساتید به این مسائل بوده و از طرف دیگر ضریب دشواری ساده سؤالات، شاید به دلیل شرایط تدریس و برگزاری امتحانات به صورت مجازی در سال تحصیلی ۱۴۰۰-۱۳۹۹ بود. همچنین بیشتر سؤالات در سطح تاکسونومی دو و بدون نقص بود. در ساختار سؤال بیش‌ترین نوع نقص بود؛ که با مطالعه وفا مهر و همکاران در سال ۲۰۱۱ (۲۶) و مطالعه الماسی و همکاران در سال ۱۳۹۵ (۲۷) هم‌خوانی دارد که این مسئله می‌تواند نشان‌دهنده دقت اساتید در طراحی سؤالات و همچنین تأثیر برگزاری کارگاه‌های مختلف برگزار شده توسط مرکز مطالعات دانشگاه و همچنین بررسی سؤالات هر دانشکده به صورت تصادفی قبل از برگزاری آزمون‌های پایان‌ترم باشد. بیش‌ترین نوع نقص به ترتیب، عدم وجود بخش اعظم اطلاعات در ساقه سؤال، وجود لغات منفی در ساقه سؤال ۱۳۶ مورد (۲۰٪) و وجود گزینه‌های همه موارد، هیچ‌کدام و گزینه‌های ترکیبی بود که با مطالعه اکرامی و همکاران در سال ۱۳۹۴ (۲۸) و راش و همکاران در سال ۲۰۱۶ (۲۹) هم‌خوانی دارد. در مطالعه تارانت و همکاران (۳۰) بر روی سؤالات پرستاری، فراوانی نقص در سؤالات (۴۶/۲٪) بود، در مطالعه راش و همکاران (۲۹) نیز نتایج مشابهی بیان شد. در مطالعه رامراجی و همکاران (۳۱) میزان نقص در سؤالات ۶۰٪ درصد بود، که هر سه مطالعه با مطالعه حاضر هم‌خوانی دارد. به‌طور کلی وجود لغات منفی در ساقه سؤال باعث گیج شدن دانشجو می‌شود، از طرفی در مورد سؤالاتی که به صورت (کدام مورد صحیح است) یا (کدام مورد غلط است) این سؤالات نیز خیلی ساده و راحت طراحی می‌شوند؛ و در نهایت بین سطح شناختی سؤالات (تاکسونومی) و نقص در ساختار سؤالات رابطه معنی‌داری وجود داشت که هرچه سؤال سخت‌تر باشد، احتمال اشتباه در طراحی سؤال ممکن است بیشتر رخ بدهد. استفاده روز افزون از سؤالات چند گزینه ای در ارزشیابی‌ها ضرورت آنالیز دقیق سؤالات توسط اساتید قبل و بعد از آزمون را بیشتر می‌کند (۳۲). برای این کار لازم است بر تهیه بانک سؤال و آنالیز سؤالات موجود در بانک سؤالات تأکید گردد. دانه و همکاران در مطالعه خود نشان داده‌اند (۳۳) که حتی بیشتر سؤالات انتخاب شده از بانک سؤال دانشکده پرستاری، دارای اشکالات جزئی و برخی از آنها دارای اشکالات اساسی بوده است. آنها اساتید را تشویق می‌کنند که قبل از استفاده از سؤالات بانک سؤال، بازهم سؤال مورد نظر را به دقت مورد بررسی قرار دهند. تحلیل موردی سؤالات، اطلاعات مفیدی را برای بهبود کیفیت و صحت سؤالات چند گزینه ای در اختیار آنان قرار می‌دهد (۳۴ و ۳۵). از نقاط قوت این مطالعه عدم انجام چنین مطالعه‌ای در سطح دانشگاه آن هم بر روی مسئله‌ای مهم مثل تحلیل کمی و کیفی سؤالات چهارگزینه‌ای بر روی سؤالات دو سال تحصیلی بود و از نقاط ضعف این مطالعه تعداد کم سؤالات در بررسی به‌وسیله چک‌لیست میلمن بود که پیشنهاد می‌شود علاوه بر تجزیه و تحلیل مرتب سؤالات پایان‌ترم، برای استفاده از نتایج آن برنامه‌ریزی لازم صورت گیرد. ایجاد بانک سؤالات برای اساتید و یا اصلاح آن، افزایش آگاهی آنان به طور مستمر و مناسب، به‌کارگیری کارشناسان جهت ارزیابی سؤالات و ایجاد انگیزه برای اعضای هیات علمی می‌تواند باعث شود تا قضاوتی مناسب از یادگیری دانشجویان به عمل آید و در جهت بهبود روش تدریس و فرایند یادگیری یاددهی اقدام مناسب صورت گیرد. همچنین پیشنهاد می‌گردد مطالعات مشابه بر روی سایر آزمون‌ها صورت پذیرد.

تقدیر و تشکر

بدین وسیله نویسندگان مقاله مراتب تشکر را از کارکنان انفورماتیک معاونت آموزشی (سرکار خانم میترا حسینی، سرکار خانم مطهره صادقی) و معاونت تحقیقات و فناوری دانشگاه علوم پزشکی شاهرود و کلیه عزیزانی که ما را یاری نمودند ابراز می‌دارند.

References:

1. Brookhart SM. Educational assessment knowledge and skills for teachers. *Educational Measurement: issues and practice*. 2011;30(1):3-12.
2. Oguguo BC, Nannim FA, Agah JJ, Ugwuanyi CS, Ene CU, Nzeadibe AC. Effect of learning management system on Student's performance in educational measurement and evaluation. *Education and Information Technologies*. 2021;26(2):1471-83.
3. Wakeford R, Denney M, Ludka-Stempien K, Dacre J, McManus I. Cross-comparison of MRCGP & MRCP (UK) in a database linkage study of 2,284 candidates taking both examinations: assessment of validity and differential performance by ethnicity. *BMC Medical education*. 2015;15(1):1-12.
4. Dehnad A, Nasser H, Hosseini AF. A comparison between three-and four-option multiple choice questions. *Procedia-Social and Behavioral Sciences*. 2014;98:398-403.
5. Colbert MA. Statistical analysis of multiple choice testing. Air Command And Staff Coll Maxwell Afb AL; 2001.
6. Thompson AR, Husmann PR. Developing multiple-choice questions for anatomy examinations . *Teaching Anatomy: Springer*; 2020. p. 405-16.
7. Stevens NT, Holmes K, Grainger RJ, Connolly R, Prior A-R, Fitzpatrick F, et al. Can e-learning improve the performance of undergraduate medical students in Clinical Microbiology examinations? *BMC medical education*. 2019;19(1):1-8.
8. Amin MM, Shayan S, Hashemi H, Poursafa P, Ebrahimi A. Analysis of Multiple Choice Questions Based On Classical Test Theory. *Iranian Journal of Medical Education*. 2011;10(5):719-725.
9. Bailey PH, Mossey S, Moroso S, Cloutier JD, Love A .Implications of multiple-choice testing in nursing education. *Nurse Educ Today*. 2012;32(6):e40-e4.
10. Boland RJ, Lester NA, Williams E. Writing multiple-choice questions. *Acad Psychiatry*. 2010;34(4):310-6.
11. Mardapi D. Educational measurement, assessment, and evaluation. Parama Publishing; 2017.
12. Dent J, Harden RM, Hunt D. A Practical Guide for Medical Teachers, E-Book: Elsevier health sciences; 2021.
13. Meyari A, Beiglarkhani M, Zandi M, Vahedi M, Miresmaeili Af. The Effect Of Education On Improvement Of Multiple Choice Questions'designing In Annual Residency Exams Of Dental School. 2012.9(1).36-45.
14. Ramani S, Ring BN, Lowe R, Hunter D. A pilot study assessing knowledge of clinical signs and physical examination skills in incoming medicine residents. *J Grad Med Educ*. 2010;2(2):232-5.
15. Walsh J, Harris B, Smith P. Single best answer question-writing tips for clinicians. *Postgraduate medical journal*. 2017;93(1096):76-81.
16. Miller-Matero LR, Martinez S, MacLean L, Yaremchuk K, Ko AB. Grit: A predictor of medical student performance. *Education for Health*. 2018;31(2):109-113.
17. Belfi LM, Bartolotta RJ, Giambone AE, Davi C, Min RJ. "Flipping" the introductory clerkship in radiology: impact on medical student performance and perceptions. *Acad Radiol*. 2015;22(6):794-801.
18. Landers RN. Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & gaming*. 2014;45(6):752-68.
19. Baig M, Ali SK, Ali S, Huda N. Evaluation of multiple choice and short essay question items in basic medical sciences. *Pakistan journal of medical sciences*. 2014;30(1):3-6.
20. Coughlin P, Featherstone C. How to write a high quality multiple choice question (MCQ): A guide for clinicians. *Eur J Vasc Endovasc Surg*. 2017;54(5):654-8.
21. Yonker JE. The relationship of deep and surface study approaches with factual and applied test-bank multiple-choice question performance. *Assessment & Evaluation in Higher Education*. 2011;36(6):673-86.
22. Secolsky C, Denison DB. Handbook of measurement, assessment, and evaluation in higher education: Routledge New York; 2012.

- 23.Sanagoo A, Jouybari L, Ghanbari Gorji M. Quantitative and qualitative analysis of academic achievement tests in Golestan University of Medical Sciences. *Research in Medical Education*. 2010;2(2):24-32.
- 24.Baghaei R, Shams S, Feizi A. Evaluation of the nursing students final exam multiple-choice questions in uremia university of medical sciences. *Nursing And Midwifery Journal*. 2016;14(4):291-9.
- 25.Khajeali N, Aslami M, Araban M. Analysis of medical and dentistry basic sciences examinations: A case study. *Payees (Health Monitor)*. 2020;19(4):383-9.
- 26.Vafamehr V, Dadgostarnia M. Reviewing the results of qualitative and quantitative analysis of MCQs in Introduction to clinical medicine course. *Iranian Journal of Medical Education*. 2011;10(5):1146-1152.
- 27.Almasi K, Bavani SM, Mohammadpour Y. Examining the preferred learning styles (PLSs) of nursing and midwifery students of Urmia University of Medical Sciences. *Middle East Journal Of Family Medicine* . 2018;7(10):244-253.
- 28.M Ekrami SR, F Afzali Nasab, F Vijehzadeh, A Farajpour. Analysis of the four-choice end-of-semester gymnastics course 2 exam at Payame Noor University. *Journal of Organizational Behavior Management Studies in Sport*. 2016;2(6):65-72.
- 29.Rush BR ,Rankin DC, White BJ. The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC medical education*. 2016;16(1):1-10.
- 30.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 Today*. 2006;26(8):662-71.
- 31.Ramraje S, Sable P. Comparison of the effect of post-instruction multiple-choice and short-answer tests on delayed retention learning. *The Australasian medical journal*. 2011;4(6):332-339.
- 32.Thompson AR, Husmann PR. Developing multiple-choice questions for anatomy examinations. *Teaching Anatomy: A Practical Guide*. 2020:405-16.
- 33.Danh T, Desiderio T, Herrmann V, Lyons HM, Patrick F, Wantuch GA ,et al. Evaluating the quality of multiple-choice questions in a NAPLEX preparation book. *Currents in Pharmacy Teaching and Learning*. 2020;12(10):1188-93.
- 34.Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians . *International journal of endocrinology and metabolism*. 2012;10(2):486-489.
- 35.Surry LT, Torre D, Trowbridge RL, Durning SJ. A mixed-methods exploration of cognitive dispositions to respond and clinical reasoning errors with multiple choice questions. *BMC medical education*. 2018;18:1-11.