

Mashhad University of
Medical Sciences

Navid No

کمیته تحقیقات دانشجویی
معاونت پژوهش و فناوری
دانشگاه علوم پزشکی مشهدJournal homepage: <https://nmj.mums.ac.ir/>*Original Article*

Analysis of the questions of the Postgraduate Fellowship Admission Exam of the Cardiology Department of Tehran University of Medical Sciences in 2021 and 2022

Ali Hosseinsabet^{1*} , Talat Diba-Vajari² , Tahereh Davarpassand³ , Hasan Aghajani³

1. Professor, Tehran Heart Center, Faculty of Medicine, Tehran University of Medical Sciences, Tehran, Iran
2. Assistant Professor, Educational Sciences Department, Payame Noor University, Tehran, Iran
3. Associate Professor, Tehran Heart Center, Faculty of Medicine, Tehran University of Medical Sciences, Tehran, Iran

Corresponding author: ali_hosseinsabet@yahoo.com

Received: 28 December 2023; Revised: 13 May 2024 ; Accepted: 30 June 2024

Abstract

Background and Aims: Fellowship courses have been designed to improve medical clinicians' knowledge, skills, and performance. The admission exam consisted of four-choice questions, but developing these questions is not easy, so continuous question monitoring is vital to maintain and improve the quality of questions. In this study, the questions of the admission fellowship exam of the cardiology department of Tehran University of Medical Sciences in 2021 and 2022 were analyzed.

Materials and Methods: This is a cross-sectional study. The structure of 130 four-choice questions was evaluated by using a checklist, and then the taxonomies of each of the questions were determined. Difficulty coefficient, discrimination coefficient, efficiency of distractors, and reliability of the exam were determined and the relationship between them was calculated.

Results: 61% of questions had bugs in the stem and 62% had bugs in the choices part. 66% were in Taxonomy I. 67% of the questions had appropriate difficulty, 58% had appropriate discrimination coefficient and questions with three functioning distractors made up 47%. The reliability of these exams was 0.788 and 0.826. There was a significant correlation between some variables mentioned above.

Conclusion: The reliability of these exams was good, but questions with appropriate difficulty, with appropriate discrimination, or questions with high-performance distractors accounted for approximately only one-half of the questions and there were structural problems in many questions. In addition, most of the questions were in category I.

Keywords: Test; Cardiovascular; Fellowship; Discrimination coefficient; Difficulty coefficient; Distraction options.

Cite this article as: Hosseinsabet A, Diba-Vajari T, Davarpassand T, Aghajani H. Analysis of the Questions of the Postgraduate Fellowship Admission Exam of the Cardiology Department of Tehran University of Medical Sciences in 2021 and 2022. Navid No, 2024; 27(89): 11-23. <https://doi.org/10.22038/nmj.2024.77164.1429>.

E-ISSN: 2645-5927 / P-ISSN: 2645-5919

Copyright: © 2024 by the author.

Open Access: This is an open access article under the CC BY license[\(http://creativecommons.org/licenses/by/4.0/\)](http://creativecommons.org/licenses/by/4.0/).

Publisher's Note: Mashhad University of Medical Sciences remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Mashhad University of
Medical Sciences

نوید نو

Navid No

Journal homepage: <https://mnj.mums.ac.ir/>



کمیته تحقیقات دانشجویی
معاونت پژوهش و فناوری
دانشگاه علوم پزشکی مشهد

نوع مقاله (پژوهشی)

تحلیل سؤالات آزمون کتبی پذیرش دستیار دوره تکمیلی تخصص (فلوشیپ)

گروه قلب و عروق دانشگاه علوم پزشکی تهران در سال‌های ۱۴۰۰ و ۱۴۰۱

علی حسین ثابت^{#۱} ID، طلعت دیبا واجاری^۲ ID، طاهره داورپسند^۳ ID، حسن آقاجانی^۳ ID

۱. استاد، مرکز قلب تهران، دانشکده پزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران
۲. استادیار، دانشگاه پیام نور، تهران، ایران
۳. دانشیار، مرکز قلب تهران، دانشکده پزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران
پست الکترونیک نویسنده مسئول: ali_hosseinsabet@yahoo.com
تاریخ دریافت: ۱۴۰۲/۱۰/۰۷، تاریخ بازنگری: ۱۴۰۳/۰۲/۲۴، تاریخ پذیرش: ۱۴۰۳/۰۴/۱۰

چکیده

مقدمه و هدف: دوره‌های فلوشیپ برای ارتقا سطح دانش، مهارت و عملکرد متخصصان بالینی پزشکی طراحی شده است. آزمون ورودی این دوره شامل سؤالات چهارگزینه‌ای است. طراحی این گونه سؤالات آسان نمی‌باشد، لذا پایش مداوم سؤالات برای حفظ و ارتقای کیفیت آن‌ها حیاتی است. در این پژوهش هدف تحلیل سؤالات آزمون ورودی فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران در سال‌های ۱۴۰۰ و ۱۴۰۱ می‌باشد.

مواد و روش‌ها: این مطالعه مقطعی می‌باشد. ساختار ۱۳۰ سوال چهارگزینه‌ای با استفاده از چک‌لیست طراحی شده ارزیابی و تاکسونومی آنها مشخص شد. ضریب دشواری و تمیز و کارایی گزینه‌های انحرافی و پایایی سؤالات آزمون تعیین و ارتباط آنها محاسبه شد.

یافته‌ها: ۶۱٪ سؤالات اشکال در ساقه و ۶۲٪ سؤالات اشکال در گزینه‌ها داشتند. ۶۶٪ سؤالات در تاکسونومی I قرار داشتند، ۶۷٪ سؤالات دارای دشواری مناسب بودند، ۵۸٪ سؤالات دارای تمیز مناسب بودند و سؤالات با سه گزینه کارا، ۴۷٪ سؤالات را تشکیل می‌دادند. میزان پایایی این آزمون‌ها ۰/۷۸۸ و ۰/۸۲۶ بود. بین تعداد اشکال در ساقه با تعداد اشکال در گزینه‌های سؤال، بین طبقه دشواری و تعداد گزینه‌های کارا همبستگی معنی‌دار و مثبت وجود داشت. بین تعداد اشکال در ساقه سؤال با تاکسونومی سؤالات، بین طبقه دشواری و طبقه تمیز، طبقه تمیز و تعداد گزینه‌های انحرافی کارا همبستگی معنی‌دار و منفی وجود داشت.

نتیجه‌گیری: پایایی این آزمون‌های خوب بود؛ اما سؤالات با دشواری یا تمیز مناسب یا سؤالات با تعداد گزینه‌های انحرافی با عملکرد بالا حدود نیمی از سؤالات را تشکیل می‌دادند و اشکالات ساختاری در بسیاری از سؤالات وجود داشت. تاکسونومی بیشتر سؤالات در طبقه I قرار داشت.

کلمات کلیدی: آزمون، قلب و عروق، فلوشیپ، ضریب تمیز، ضریب دشواری، گزینه‌های انحرافی.

مقدمه

ضعف و کاستی‌های آزمون را شناخت تا اقدام درست برای رفع کاستی‌ها را انجام داد (۹). هم‌چنین، بررسی عملکرد آزمون می‌تواند به شناسایی اختلاف اهداف آموزشی و میزان یادگیری فراگیران کمک نماید؛ پس ارزیابی آزمون فرایندی است که با هدف بهبودی در فرایند طراحی سؤال و دوره آموزشی انجام می‌گیرد (۷).

تحلیل سؤالات آزمون به دو روش کیفی و کمی انجام می‌گیرد. در ارزیابی کمی سؤالات شاخص‌هایی مثل ضریب دشواری، تمیز و گزینه‌های انحرافی و در ارزیابی کیفی سؤالات به تاکسونومی و بررسی ساختار سؤالات پرداخته می‌شود (۱۰-۱۲).

باتوجه به اینکه نظام آموزش پزشکی هر کشوری بر اساس مقتضیات و شرایط بومی آن متفاوت است، هم‌چنین مراحل طی مدارج و نظام ارزیابی هر کشوری متفاوت است لازم است هر کشور باتوجه به نظام آموزش پزشکی خاص آن کشور به ارزیابی کیفیت سؤالات مطرح شده در مقاطع مختلف تحصیلی بپردازد. از این رو نمی‌توان نتایج تحقیقات در سایر کشورها را به کشور خودمان تعمیم دهیم. علی‌رغم این که موضوع بررسی و تحلیل سؤالات چندگزینه‌ای مسئله‌ای شناخته شده است؛ اما پایش کیفیت سؤالات به طور مستمر به‌عنوان یکی از کارکردهای نظام آموزش پزشکی هر کشور باید مورد توجه باشد.

دوره‌های تکمیلی تخصص (فلوشیپ) که مرحله‌ای از ارتقای دانش و مهارت است، برای ارتقا سطح دانش و به‌خصوص مهارت و عملکرد متخصصان بالینی پزشکی که دارای مدرک دانشنامه تخصصی مرتبط است طراحی شده است. اما بیشتر مطالعات انجام شده در مورد ارزیابی سؤالات آزمون‌های پزشکی، مربوط به دوره پزشکی عمومی بوده و بسیار کمتر به حوزه‌های آموزش تخصصی و فوق‌تخصص پرداخته‌اند. کیفیت سؤالات آزمون پذیرش دستیار فلوشیپ در کشور، تا کنون مورد بررسی قرار نگرفته است، لذا مهم‌ترین نقص در دانش موجود ما عدم آگاهی از کیفیت سؤالات طراحی شده این آزمون‌ها است که از طریق مطالعات توصیفی برطرف خواهد شد. از این رو در این پژوهش سؤالات آزمون پذیرش دستیار فلوشیپ گروه

ارزیابی یک جز مهم از نظام آموزش پزشکی است و در مراحل مختلف آموزش دارای نقش است. یکی از این نقش‌ها، ارزیابی میزان یادگیری فراگیران در طی دوره آموزشی است. یکی از راه‌های ارزیابی استفاده از آزمون‌های چندگزینه‌ای است. آزمون‌های چندگزینه‌ای به‌صورت معمول در ارزیابی فراگیران رشته‌های پزشکی مورد استفاده قرار می‌گیرد. این آزمون‌ها عینی بوده و تصحیح آن ساده است هر چند که عامل حدس می‌تواند نتایج آن را مخدوش نماید (۱،۲).

در سؤالات چندگزینه‌ای علاوه بر یک گزینه صحیح، گزینه‌های انحرافی وجود دارد تا با منحرف کردن فراگیر ضعیف از گزینه درست، بر اعتبار آزمون اضافه نماید و آن را دارای قابلیت اعتماد بیشتری نماید (۳). تهیه گزینه‌های انحرافی نیز کار آسانی نمی‌باشد؛ زیرا این گزینه‌ها نباید به طور واضح نادرست باشد. نامتناسب بودن گزینه‌های انحرافی منجر به خیلی آسان شدن یا خیلی مشکل شدن سؤال می‌شود که تغییر گزینه‌ای انحرافی می‌تواند منجر به بهبودی سؤال شود (۴).

پایش مداوم سؤالات آزمون برای حفظ و ارتقای کیفیت سؤالات آزمون امری حیاتی است. سؤالات آزمون‌های چندگزینه‌ای هم باید قبل از اجرا و هم بعد از اجرا مورد ارزیابی قرار گیرد (۵). قبل از اجرا باید از نظر رعایت قواعد نگارشی، اهداف و ساختار مورد ارزیابی قرار گیرد و بعد از آزمون نیز از نظر شاخص‌هایی چون ضریب دشواری و تمیز مورد ارزیابی قرار گیرد (۶،۵). ارزیابی سؤالات آزمون به روش علمی کمک می‌کند که کیفیت آزمون بر پایه شواهد و مستندات ارتقا پیدا کند تا کیفیت آموزش و ارزیابی فراگیران رشد و تحول پیدا کند (۸،۷). اگر یک آزمون دارای شاخص‌های مطلوب در ارزیابی کمی و کیفی سؤالات آزمون نباشد نه فقط کارکرد اصلی آزمون از بین می‌رود؛ بلکه بر انگیزه فراگیران اثر نامطلوب داشته و تلاش آموزشگران را بی‌ثمر می‌سازد. برای ارتقای کیفیت آزمون لازم است نقاط

۳۰ سوال دیگر برای هر رشته متفاوت است. در این پژوهش ۷۰ سوال مشترک در هر سال (در مجموع ۱۴۰ سوال) مورد تجزیه و تحلیل قرار گرفت. باتوجه به اینکه در سال ۱۴۰۰، ۴ سوال و در سال ۱۴۰۱، ۶ سوال پس از رسیدگی به اعتراض ها حذف شدند، لذا ۶۶ سوال سال ۱۴۰۰ و ۶۴ سوال (در مجموع ۱۳۰ سوال) مورد بررسی نهایی قرار گرفت.

ساختار هر سوال با استفاده چک لیست ارائه شده در جدول ۱ مورد بررسی قرار گرفت. موارد ارائه شده بیانگر ساختار صحیح یک سوال چهار گزینه‌ای می باشد. این موارد چک لیست براساس مطالعه متون و پژوهش‌های گذشته در این زمینه انتخاب شدند. این چک‌لیست دارای ۱۷ مورد است که موارد ۱ تا ۸ آن در مورد ساقه سؤال و موارد ۹ تا ۱۷ در مورد گزینه‌های سؤال است. جدول ۱ (۱۳) تاکسونومی سوالات براساس طبقه‌بندی بلوم انجام شد و سوالات در سه گروه I,II,III طبقه‌بندی شدند (۱۴).

آموزشی قلب و عروق دانشگاه علوم پزشکی تهران در سال ۱۴۰۰ و ۱۴۰۱ از نظر ضریب دشواری، ضریب تمیز، پایایی، کارکرد گزینه‌های انحرافی، تاکسونومی سوالات و ساختار سوالات مورد بررسی قرار گرفت.

روش کار

این مطالعه مقطعی می باشد. در این مطالعه فایل اکسل پاسخ‌نامه‌ها و سؤال‌های آزمون پذیرش دستیار فلوشیپ گروه قلب و عروق از مرکز آزمون‌های دانشگاه علوم پزشکی تهران دانشگاه علوم پزشکی تهران در سال ۱۴۰۰ و ۱۴۰۱ دریافت شد. گروه قلب و عروق، در چهار رشته اکوکاردیوگرافی، اینترونشنال کاردیولوژی، الکتروفیزیولوژی و نارسایی قلب، دستیار فلوشیپ می پذیرد و برای هر رشته آزمون کتبی با ۱۰۰ سؤال چهار گزینه‌ای برگزار می‌شود. ۷۰ سوال این آزمون بین رشته‌های مختلف یکسان است و

جدول ۱: چک لیست ارزیابی ساختار سوال های آزمون

ردیف	محتوای چک لیست	گروه
۱	محور مناسب برای سوال انتخاب شده است (بیشتر از یک مساله یا مطلب در سوال قرار ندارد و یک مورد یا موقعیت بالینی مبنای طرح سوال می باشد مثل تشخیص، درمان یا ...)	ساقه سوال
۲	ایده اصلی (بخش اعظم اطلاعات) به طور کامل در ساقه سوال گنجانده شده است (ساقه سوال حاوی اطلاعاتی است که داوطلب توانمند برای پاسخ دهی نیاز دارد)	
۳	بدنه اصلی سوال واضح، بدون ابهام و روان بوده و از نظر لغوی و دستوری پیچیده نمی باشد و قواعد دستوری رعایت شده است	
۴	از به کارگیری عبارت های مبهم مثل همه موارد، هیچکدام در ساقه سوال پرهیز شده است	
۵	از به کارگیری قیود خاص مثل هیچوقت، همیشه، کاملاً یا مطلقاً در ساقه سوال پرهیز شده است	
۶	سوال هدایت کننده به روشنی مشخص کرده است که چگونه باید به سوال پاسخ داد	
۷	می توان بدون نگاه کردن به گزینه ها به سوال پاسخ داد	
۸	سوال به صورت جمله مثبت طراحی شده است، در صورت منفی بودن سوال هدایت کننده، لغات منفی مشخص شده است	
۹	گزینه ها با هم همپوشانی ندارند و مستقل از هم طراحی شده اند	گزینه ها
۱۰	گزینه ها از نظر محتوی و دستوری همگن طراحی شده اند و به موضوع واحدی مربوط می باشند	
۱۱	طول جمله گزینه ها تقریباً یکسان می باشد	
۱۲	از دادن سرنخ که نشان می دهد گزینه صحیح کدام گزینه است خودداری شده است مانند: - کاربرد قیود خاص نظر همیشه، هیچ وقت، کاملاً، مطلقاً - گزینه هایی که از نظر لغوی دارای کلمات مشابه پایه سوال می باشند - تفاوت دستوری گزینه ها که داوطلب را به سوی گزینه صحیح هدایت می کند - گزینه ای که به طور کاملاً واضح صحیح می باشد	
۱۳	از عبارت هایی نظیر همه موارد فوق، همه موارد، بجز، هیچکدام از موارد فوق، در گزینه ها استفاده نشده باشد	
۱۴	از عبارت هایی مبهمی نظیر غالباً، معمولاً، گاهی، هرگز، استفاده نشده باشد	
۱۵	تا حد امکان از کاربرد عبارت های تکراری در گزینه ها پرهیز شده است	
۱۶	از کاربرد گزینه های منفی برای ساقه منفی (منفی مضاعف) خود داری شده است	
۱۷	از کاربرد گزینه های متضاد یکدیگر خودداری شده است	

دشوار نشان داده شد. توزیع فراوانی میزان تمیز سؤالات (بر اساس ضریب تمیز Rpbis) براساس طبقه‌بندی خوب، قابل قبول و ضعیف نیز انجام شد توزیع فراوانی کارائی گزینه‌های انحرافی برای هر آزمون - سال به تصویر کشیده شد. از آزمون خوبی برازندگی کای اسکوتر برای ارزیابی توزیع تصادفی گزینه های صحیح استفاده شد. همبستگی بین متغیرهای پیش‌گفت با استفاده از ضریب همبستگی اسپیرمن- براون مورد ارزیابی قرار گرفت. برای تجزیه و تحلیل داده‌ها از نرم‌افزار SPSS برای ویندوز، ویرایش ۲۴ و اکسل ۲۰۲۱ استفاده شد. نتایج آزمون‌های آماری اگر دارای $P\text{-value} < 0.05$ بودند، از نظر آماری معنی‌دار تشخیص داده شدند.

یافته‌ها

در سال ۱۴۰۰، ۳۱ نفر و در سال ۱۴۰۱، ۲۳ نفر در آزمون شرکت کردند. میزان فراوانی اشکالات ساختاری برحسب سال و مکان آن در جدول ۲ نشان داده شده است. ۵۰ سوال (۳۹٪) بدون اشکال در ساقه سوال و ۳۸ سوال (۲۹٪) بدون اشکال در گزینه‌ها بودند و در مجموع ۱۶ سوال (۱۲٪) بدون اشکال ساختاری در ساقه و گزینه‌ها بودند.

منظور از دشواری سؤالات، نسبت آزمودنی‌هایی است که به سؤال موردنظر پاسخ درست داده‌اند. سؤالات داری ضریب دشواری کمتر از ۰/۳ دشوار، بین ۰/۷ - ۰/۳ مطلوب و بیشتر از ۰/۷ آسان در نظر گرفته شدند (۱۳). منظور از تمیز سؤالات توانایی سوال برای افتراق آزمودنی‌های قوی از ضعیف گفته می‌شود. ضریب تمیز براساس ضریب همبستگی دورشته ای نقطه ای (Rpbis) که همبستگی نمره کل آزمودنی با نمره هر سوال را می‌سنجد محاسبه شد. ضریب تمیز ۰/۲۵ و بیشتر را عالی، ۰/۲۴ - ۰/۱۵ خوب و کمتر از ۰/۱۵ را ضعیف در نظر گرفته شد (۱۵). پایایی آزمون با استفاده از آلفای کرونباخ مورد ارزیابی قرار گرفت. گزینه‌های سؤال به جز گزینه صحیح سؤال گزینه انحرافی نامیده می‌شود و گزینه انحرافی فاقد عملکرد گزینه‌ای در نظر گرفته شد که کمتر از یک آزمودنی آن را انتخاب نمایند (۱۳). ضریب کارائی گزینه‌های انحرافی، نسبت درصدی تعداد گزینه‌های انحرافی مناسب به تعداد کل گزینه‌های انحرافی است (۱۶).

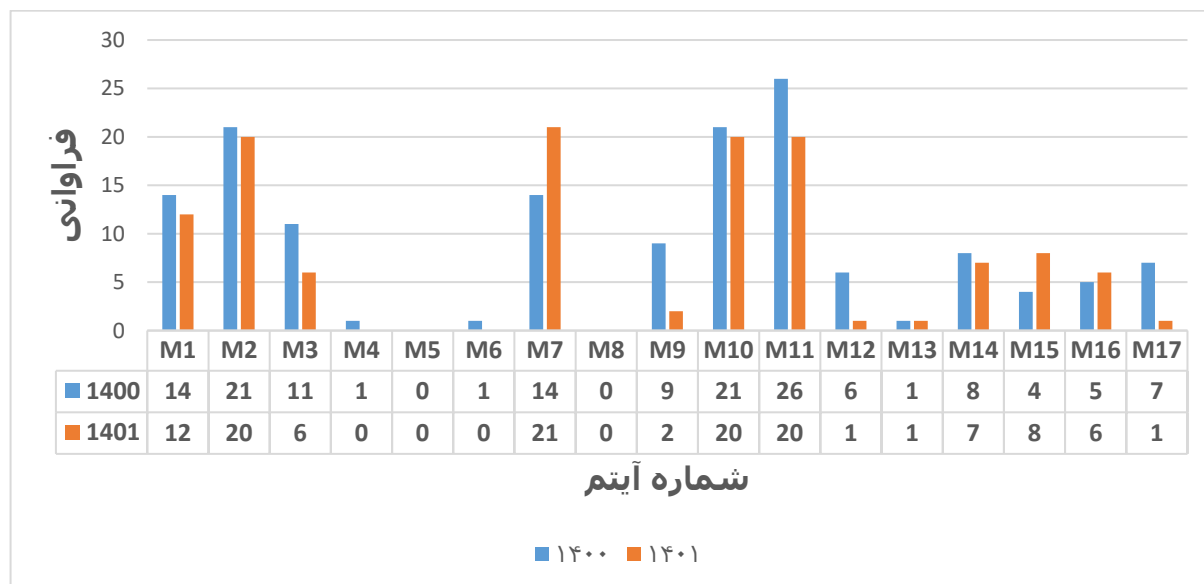
میزان رعایت هریک از موارد چک لیست پیش‌گفت در کل سؤالات به صورت فراوانی ارائه شد. تاکسونومی سؤالات نیز به صورت فراوانی بیان شد. توزیع فراوانی میزان دشواری سؤالات هر آزمون بر اساس طبقه‌بندی آسان، مناسب و

جدول ۲: توزیع فراوانی تعداد اشکالات سؤالات در آزمون پذیرش دستیار فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران برحسب مکان اشکال در سال‌های ۱۴۰۱ و ۱۴۰۰

تعداد اشکال		ساقه سوال						گزینه‌های سوال						ساقه و گزینه‌های سوال					
		۱۴۰۰		۱۴۰۱		مجموع		۱۴۰۰		۱۴۰۱		مجموع		۱۴۰۰		۱۴۰۱		مجموع	
تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد
۰		۲۵	۳۸	۲۵	۳۹	۲۴	۱۶	۲۴	۲۲	۳۴	۲۲	۲۴	۱۶	۲۴	۲۲	۲۲	۳۳	۲۲	۲۴
۱		۲۴	۳۶	۲۱	۳۳	۴۵	۳۵	۲۲	۳۳	۲۲	۲۲	۳۳	۲۲	۲۲	۲۲	۲۲	۲۲	۲۲	۲۲
۲		۱۴	۲۱	۱۷	۲۶	۳۱	۲۴	۱۹	۲۹	۱۶	۲۵	۱۶	۲۹	۱۹	۲۴	۳۱	۲۶	۱۷	۲۱
۳		۲	۳	۰	۰	۲	۱	۹	۱۴	۴	۷	۱۳	۱۰	۱۳	۷	۱۴	۱۰	۱۳	۷
۴		۱	۲	۱	۲	۲	۱	۲	۱	۱	۲	۱	۲	۱	۲	۱	۲	۱	۲
۵		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
۶		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
۷		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
جمع		۶۶	۱۰۰	۶۴	۱۰۰	۱۳۰	۱۰۰	۶۴	۱۰۰	۶۴	۱۰۰	۶۶	۱۰۰	۱۳۰	۱۰۰	۶۴	۱۰۰	۶۶	۱۰۰

سوالات آزمون سال ۱۴۰۱، در آیتم های مربوط به ساقه سوال، در آیتم های ۷، ۲، ۱ دارای بیشترین اشکال و در آیتم های ۴، ۵، ۶ و ۸ دارای کمترین اشکال بودند. در آیتم های مربوط به گزینه سوال، آیتم های ۱۱، ۱۰ دارای بیشترین اشکال و در آیتم های ۱۲، ۱۳ و ۱۷ دارای کمترین اشکال بودند.

با توجه به نمودار ۱، سوالات آزمون سال ۱۴۰۰ براساس آیتم های جدول ۱، در آیتم های مربوط به ساقه سوال، موارد ۲، ۱، ۷ و ۳ دارای بیشترین اشکال و در آیتم های ۴، ۵، ۶ و ۸ دارای کمترین اشکال بودند. در آیتم های مربوط به گزینه سوال آیتم های ۱۰ و ۱۱ دارای بیشترین اشکال بودند و در آیتم ۱۳ دارای کمترین اشکال بود. هم چنین،



نمودار ۱: فراوانی میزان اشکالات سوالات آزمون پذیرش دستیار فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران در سال ۱۴۰۰ و ۱۴۰۱ برحسب موارد چک لیست ارائه شده در جدول ۱.

بوده و هر سه گزینه انحرافی کارا بودند و سوال ۴۲ (۳۲٪) دارای میزان دشواری مناسب بوده و هر سه گزینه انحرافی کارا بودند. سوال ۲۸ (۲۲٪) دارای میزان دشواری مناسب، تمیز خوب و هر سه گزینه انحرافی کارا بودند. گزینه صحیح در ۲۱ سوال (۱۶٪) گزینه ۱، در ۲۷ سوال (۲۱٪) گزینه ۲، در ۴۲ سوال (۳۲٪) گزینه ۳ و در ۴۰ سوال (گزینه ۴) بود که انحراف آن توزیع تصادفی معنا دار بود ($P = 0/023$ ، $\text{chi-square} = 9/51$ ، $df = 3$ ، value).

از نظر تاکسونومی سوالات، ۸۶ سوال (۶۶٪) در سطح تاکسونومی I و بقیه سوالات در سطح تاکسونومی بالاتر قرار داشتند. از نظر میزان دشواری ۵۲ سوال (۴۰٪) آسان، ۶۷ سوال (۵۲٪) مناسب و بقیه سخت بودند. از نظر میزان ضریب تمیز براساس شاخص Rpbis، ۷۶ سوال (۵۸٪) دارای ضریب تمیز خوب بودند و ۳۶ سوال (۲۸٪) دارای ضریب تمیز ضعیف بودند. ۳ سوال (۲٪) دارای ضریب تمیز صفر و ۱۳ سوال (۱۰٪) دارای ضریب تمیز منفی بودند (جدول ۳). از نظر میزان کارایی گزینه‌های انحرافی، در ۶۱ سوال (۴۷٪) هر ۳ گزینه انحرافی کارا بودند و در ۴ سوال (۳٪) هیچ گزینه انحرافی کارا نبود (جدول ۴). پایایی (آلفای کرونباخ) آزمون در سال ۱۴۰۰ و ۱۴۰۱ به ترتیب ۰/۷۸۸ و ۰/۸۲۶ بود.

هم چنین ۴۱ سوال (۳۲٪) دارای میزان دشواری مناسب و ضریب تمیز خوب بودند، ۴۳ سوال (۳۳٪) دارای تمیز خوب

جدول ۳: توزیع فراوانی سؤالات آزمون پذیرش دستیار فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران سال ۱۴۰۰ و ۱۴۰۱، بر حسب تاکسونومی سؤالات، طبقه دشواری و طبقه تمیز (Rpbis)

تاکسونومی	۱۴۰۰		۱۴۰۱		طبقه تمیز	مجموع		۱۴۰۰		۱۴۰۱		طبقه دشواری	مجموع		۱۴۰۰		۱۴۰۱	
	تعداد	درصد	تعداد	درصد		تعداد	درصد	تعداد	درصد	تعداد	درصد		تعداد	درصد	تعداد	درصد	تعداد	درصد
I	۴۸	۷۳	۳۸	۵۹	خوب	۴۰	۵۲	۳۶	۲۳	۴۴	۲۹	آسان	۶۶	۸۶	۵۹	۳۸	۷۳	۴۸
II	۱۱	۱۷	۱۶	۲۵	قابل قبول	۵۲	۶۷	۵۸	۳۷	۴۵	۳۰	مناسب	۲۱	۲۷	۲۵	۱۶	۱۷	۱۱
III	۷	۱۰	۱۰	۱۶	ضعیف	۹	۱۱	۶	۴	۱۱	۷	سخت	۱۳	۱۷	۱۶	۱۰	۱۰	۷
جمع	۶۶	۱۰۰	۶۴	۱۰۰	جمع	۱۰۰	۱۳۰	۱۰۰	۶۴	۱۰۰	۶۶	جمع	۱۰۰	۱۳۰	۱۰۰	۶۴	۱۰۰	۶۶

جدول ۴: توزیع فراوانی سؤالات آزمون پذیرش دستیار فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران سال ۱۴۰۰ و ۱۴۰۱ بر حسب میزان کارکرد گزینه‌های انحرافی

تعداد گزینه‌های انحرافی کارا	میزان کارائی گزینه‌های انحرافی	۱۴۰۰		۱۴۰۱		مجموع	
		تعداد	درصد	تعداد	درصد	تعداد	درصد
۳	۱۰۰٪	۳۰	۴۶	۳۱	۴۸	۶۱	۴۷
۲	۶۶٪	۲۶	۳۹	۲۶	۴۱	۵۲	۴۰
۱	۳۳٪	۸	۱۲	۵	۸	۱۳	۱۰
۰	۰	۲	۳	۲	۳	۴	۳
جمع	-	۶۶	۱۰۰	۶۴	۱۰۰	۱۳۰	۱۰۰

بحث

در سؤالاتی که دارای اشکال مورد ۱ چک لیست بودند، سوال به نحوی طراحی شده بود که تشخیص، درمان یا پیش آگهی در سوال مشخص نشده بود. سؤالاتی که دارای اشکال مورد ۲ چک لیست بودند، دارای اطلاعات اضافی بودند. به عبارتی در بسیاری از موارد سؤال تاکسونومی I به صورت یک مورد بالینی مطرح شده بود. به عبارت دیگر به جای طراحی سناریوی بالینی، سرگذشت یک مورد بالینی روایت شده بود. بیشتر سؤالاتی که اشکال مورد ۷ چک لیست را داشتند، سؤالاتی بودند که ساقه سؤال به صورت "کدام یک از موارد زیر صحیح است" یا "..... نادرست است" طراحی شده بودند. مورد ۱۰ چک لیست به همگنی گزینه‌ها اشاره می‌کند. بیشتر سؤالاتی که دارای این اشکال بودند در حیطه‌های مختلف طراحی شده بودند به عبارت دیگر به عنوان مثال یک گزینه به روش تشخیصی، گزینه دیگر به پیش‌آگهی و دیگری به روش درمانی اشاره می‌کرده است. مورد ۱۱ چک لیست که به طول گزینه‌ها اشاره

بررسی همبستگی متغیرهای پیش‌گفت در آزمون سال ۱۴۰۰، همبستگی بین طبقه دشواری سؤالات و تعداد گزینه‌های انحرافی کارا را نشان داد ($P \text{ value} < 0/001$)، $0/446$ ، $\rho =$ بررسی همبستگی متغیرهای پیش‌گفت در آزمون سال ۱۴۰۱، همبستگی بین طبقه دشواری سوال و تعداد گزینه‌های انحرافی کارا ($P \text{ value} < 0/001$)، $0/444$ ، $\rho =$ همبستگی بین تعداد اشکالات ساقه سوال با تعدد اشکالات در گزینه‌های سوال ($P \text{ value} < 0/001$)، $0/425$ ، $\rho =$ همبستگی بین تعداد اشکالات در ساقه سوال و تاکسونومی سؤالات ($P \text{ value} = 0/001$)، $0/418$ ، $\rho = -$ همبستگی بین طبقه تمیز سؤالات و تعداد گزینه‌های انحرافی کارا ($P \text{ value} = 0/006$)، $0/342$ ، $\rho = -$ و همبستگی بین طبقه دشواری سؤالات و تمیز سؤالات ($P \text{ value} = 0/032$)، $0/268$ ، $\rho = -$ را نشان داد.

قرار گرفته است و کمتر سطوح ترکیب اطلاعات و تجزیه و تحلیل آن مورد توجه قرار گرفته است. به نظر می‌رسد راحت بودن طراحی سؤالات در سطح تاکسونومی I منجر به گرایش بیشتر طراحان به طراحی سؤال در حد تاکسونومی I شده است. از سوی دیگر باتوجه به تعیین مبحث خاصی برای طراحی سؤال، امکان طراحی سؤالات چندگزینه‌ای باتوجه به محتوای مبحث وجود ندارد و لازم است که فراگیری دانش موجود در آن مبحث با شیوه‌ای دیگر به‌غیر از آزمون‌های چندگزینه‌ای مورد ارزیابی قرار گیرد. هم چنین این آزمون برای افرادی برگزار می‌شود که در دارای تحصیلات عالی پزشکی هستند و هدف از آن انتخاب افراد شایسته برای ادامه تحصیل و فراگیری روش‌های تشخیصی و درمانی پیشرفته و پیشتاز در رشته تخصصی خود هستند؛ لذا باید نسبت بیشتری از سؤالات این آزمون‌ها در سطح تاکسونومی III باشد. هر چند که طراحی سؤال در سطح تاکسونومی III همیشه مقدور نمی‌باشد و در بعضی اوقات می‌تواند منجر به تکراری شدن سؤال و تنزل سؤال از تاکسونومی III به I شود. عدم تفاوت توزیع فراوانی سؤالات برحسب تاکسونومی سؤالات در دو سال متوالی شاید نشانگر عدم وجود برنامه سازماندهی شده، عدم تمایل و یا عدم توانایی طراحان سؤال و یا عدم امکان طرح سؤال در سطح تاکسونومی III باشد.

سؤالات با دشواری مناسب در آزمون‌های ورودی فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران در سال‌های ۱۴۰۰ و ۱۴۰۱، نیمی از سؤالات آزمون را تشکیل می‌دادند. در پژوهش‌های انجام شده در آزمون‌های مشابه، ۶۵-۴۵٪ سؤالات دارای میزان دشواری مناسب بودند (۲۹، ۲۸، ۲۵، ۲۳، ۱۷). لذا فراوانی سؤالات با دشواری مناسب، هماهنگ با سایر نتایج بدست آمده از بررسی آزمون‌های مشابه می‌باشد. صرف‌نظر از طراحی مناسب یا نامناسب سؤال از منظر طبقه دشواری، تفاوت بین شرکت‌کنندگان در سال‌های پیش گفت و کم‌بودن آزمون‌ها در این رشته‌ها می‌تواند نتایج به‌دست‌آمده را تحت‌تأثیر قرار دهد؛ لذا این موضوع در تفسیر این نتایج باید مدنظر قرار گیرد. پایین بودن نسبت سؤالات با دشواری مناسب را می‌توان به جدی‌نگرفتن آزمون توسط طراحان سؤال به دلیل

می‌کند در سؤالات زیادی رعایت نشده بود، یک گزینه به‌صورت یک جمله کامل و دیگری به‌صورت یک کلمه و مشابه آن طراحی شده بودند. این موارد بیانگر نقاط ضعف سؤالات طرح شده است که به‌راحتی می‌توان از آن پیشگیری کرد. به نظر می‌رسد که عدم صرف وقت کافی برای طراحی سؤالات منشأ بسیاری از اشکالات است. ضمن آنکه به نظر می‌رسد در زمان طراحی چک‌لیست میلمن مورد توجه طراحان سؤالات قرار نگرفته است. از سوی دیگر موارد ۴، ۵، ۶، ۸ و ۱۳ مورد توجه طراحان سوال بوده و در طراحی سؤالات مورد دقت قرار گرفته است. این موارد بسیار عینی بوده است و به راحتی در ذهن طراح جای داشته است. در پژوهش‌های داخلی حدود یک‌سوم سؤالات دارای اشکال ساختاری بودند. (۱۷-۲۰). در پژوهش‌های خارجی تا ۷۰٪ سؤالات دارای اشکالات ساختاری بودند. (۲۱) در پژوهش‌های داخلی بیشتر از چک لیست‌هایی که تعداد آیتم‌های مورد سنجش کمتر بود استفاده شده بود و افرادی که سؤالات را مورد ارزیابی قرار داده بودند، در رشته ای که سؤالات آن را مورد ارزیابی قرار داده بودند، تخصص نداشتند، لذا به نظر می‌رسد این ارزیابان به نتوانسته‌اند به خوبی به اشکالات سوال پی ببرند. این در حالی است که در پژوهش ما ارزیابی توسط فردی انجام شد که متخصص رشته مورد نظر بود و بر ارزیابی سؤالات احاطه داشت.

بررسی آزمون‌های ورودی فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران در سال‌های ۱۴۰۰ و ۱۴۰۱ نشان داد که بیشتر سؤالات (۶۶٪) این آزمون‌ها در حد تاکسونومی I است. در پژوهش‌های مشابه در داخل کشور که در مورد آزمون پیش ارتقا، ارتقا و دانشنامه تخصصی پزشکی و یا دندانپزشکی انجام شده است بیانگر آن است که سؤالات با تاکسونومی I، ۹۱-۳۳٪ سؤالات این آزمون‌ها را تشکیل می‌دادند، هر چند که در بیشتر این آزمون‌ها این میزان از ۵۰٪ کمتر بود. (۱۷-۲۰ و ۲۲-۲۶) در مطالعات خارج کشور نیز سؤالات با تاکسونومی I را تشکیل‌دهنده ۹۴-۷۴٪ سؤالات آزمون گزارش کردند (۲۷).

اینکه بیشتر سؤالات این آزمون‌ها در حد تاکسونومی I است، نشانگر آن است که پایین‌ترین سطح یادگیری مورد توجه

جمله میزان پایای آزمون‌های مورد بررسی خود را ۰/۷۱ تا ۰/۸۸ گزارش کردند (۲۲،۳۱،۳۵،۳۸). آزمون‌های مورد بررسی ما نیز در همین محدوده از ضریب پایایی قرار داشتند. لذا با توجه به پایایی مناسب در صورت تکرار آزمون، نتایج تفاوت قابل توجهی پیدا نمی کرد.

در آزمون‌های ورودی فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران در سال‌های ۱۴۰۰ و ۱۴۰۱ بین تعداد گزینه‌های انحرافی کارا و طبقه دشواری سؤالات همبستگی متوسط، مثبت و معنی‌دار وجود داشت. به عبارت دیگر هر چه میزان گزینه‌های انحرافی کارا افزایش پیدا می کردند، طبقه دشواری سؤالات افزایش پیدا می کرد. به عبارت دیگر هر چه میزان گزینه‌های انحرافی کارا کاهش پیدا می کردند، سؤالات آسان تر می شدند. نتایج پژوهش ما با نتایج پژوهش شکورنیا (۱۳۹۸) و Kheyami و همکاران (۲۰۱۸) تطابق داشت (۱۲،۳۱). از این رو به نظر می رسد که هر چه طراح سؤال در طراحی گزینه‌های انحرافی دقت بیشتری به خرج دهد و گزینه‌ای انحرافی بهتر طراحی نماید میزان طبقه دشواری سؤال افزایش پیدا می کند و سؤالات دیگر آسان نخواهد بود.

بین اشکالات ساقه سؤال با اشکالات گزینه‌های سؤال همبستگی متوسط، معنی‌دار و مثبتی وجود داشت. به عبارت دیگر هر چه میزان اشکالات در ساقه سؤالات بیشتر بود، میزان اشکالات در گزینه‌های سؤال بیشتر بودند. به عبارت دیگر اگر ساختار سؤال در یک قسمت دارای اشکال باشد، با اشکال در قسمت دیگر همراهی کند. بین اشکالات ساقه سؤال و تاکسونومی سؤالات همبستگی متوسط، معنی‌دار و منفی وجود داشت. به عبارت دیگر هر چه تاکسونومی سؤالات کاهش پیدا می کرد، میزان اشکال در ساقه سؤال افزایش پیدا می کرد. به این مفهوم که سؤالات دارای تاکسونومی پایین تر دارای اشکالات بیشتری نسبت به سؤالات با تاکسونومی بالاتر در ساقه سؤال بودند. به عبارت دیگر طراحانی که سؤالات با تاکسونومی پایین طراحی می کردند فاقد توانایی لازم و یا وقت کافی برای طراحی سوال بوده، لذا سؤالات آنان دارای اشکالات بیشتری بود. در عوض، طراحان خبره که وقت کافی برای طراحی

شرکت کننده کمتر آزمون، طراحی سریع بدون وقت گذاشتن کافی، عدم طراحی گروهی و عدم مرور توسط همتایان نسبت داد.

سؤالات با طبقه تمیز مناسب در آزمون‌های ورودی فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران در سال‌های ۱۴۰۰ و ۱۴۰۱، ۵۸٪ از سؤالات آزمون را تشکیل می دادند. در آزمون‌های مشابه برگزار شده میزان سؤالاتی که ضریب تمیز مناسب داشتند بین ۱۹-۵۵٪ سؤالات آزمون را تشکیل می دادند (۱۷،۲۳،۲۵،۲۹). برای آزمون‌هایی که برای پذیرش در دوره‌های آموزشی خاصی با ظرفیت اندک انجام می شود هدف از برگزاری آزمون تمیز حداکثری بین افراد است تا افرادی که صلاحیت و شایستگی دارند را به خوبی گزینش نماید، لذا از این دیدگاه هر چه تمیز سؤالات و یا سؤالات با تمیز خوب، بیشتر باشد آزمون بهتری برگزار شده است و افراد بهتری انتخاب می شوند. در آزمون‌های مورد بررسی مطالعه ما نیز میزان سؤالاتی که دارای ضریب تمیز مناسب بودند در این محدوده قرار دارند هر چند که میزان‌های گزارش شده وابسته به نوع ضریب تمیز به کاررفته در مطالعه نیز است.

در آزمون‌های ورودی فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران در سال‌های ۱۴۰۰ و ۱۴۰۱، حدود ۴۷٪ سؤالات دارای سه گزینه انحرافی کار بودند. در پژوهش‌های مشابه انجام میزان سؤالاتی که هر سه گزینه انحرافی کارا بودند بین ۱۳-۱۰۰٪ سؤالات آزمون را تشکیل می دادند (۳۰-۳۸، ۳۰، ۱۲، ۱۵، ۶). آزمون‌های مورد بررسی ما نیز در نیمه این محدوده قرار داشت. طراحی گزینه‌های انحرافی کارا نیازمند صرف زمان توسط طراح است به گونه‌ای که طراحی یک سؤال چهارگزینه‌ای به طور متوسط نیازمند یک ساعت زمان است. هر گونه کاهش در صرف زمان در طراحی سؤالات چندگزینه‌ای منجر به طراحی گزینه‌های انحرافی فاقد کارایی می شود. هر چه تعداد گزینه‌های انحرافی کارا بیشتر باشد، سؤال چندگزینه‌ای بهتری طراحی شده است. میزان پایایی آزمون‌های ورودی فلوشیپ گروه قلب و عروق دانشگاه علوم پزشکی تهران در سال‌های ۱۴۰۰ و ۱۴۰۱ در حد خوبی بود. پژوهش‌های مشابه از

این پژوهش دارای محدودیت‌هایی بود که می‌تواند نتایج آن را تحت تاثیر قرار دهد. تعداد شرکت‌کنندگان در این آزمون‌ها اندک می‌بود و تعداد کم شرکت‌کنندگان مانع از بررسی و تجزیه و تحلیل سؤالات آزمون بر اساس تئوری سؤال - پاسخ بود. هم‌چنین طراح هر سؤال نامشخص بود که مانع از بررسی عملکرد هر طراح در آزمون‌های برگزار شده می‌شد.

نتیجه‌گیری

آزمون‌های ورودی دوره‌های فلوشیپ، برگزار شده توسط گروه قلب و عروق دانشگاه علوم پزشکی تهران، در سال‌های ۱۴۰۰ و ۱۴۰۱ دارای نقاط قوت و نقاط ضعف هست. از جمله نقاط قوت این آزمون‌ها پایایی خوب آزمون‌های برگزار شده است. از سوی دیگر وجود اشکالات ساختاری در بسیاری از سؤالات که قابل پیشگیری است از جمله نقاط ضعف این آزمون‌ها است. سؤالات با دشواری مناسب یا سؤالات با تمیز مناسب و سؤالات با گزینه‌های انحرافی با کارا بالا، تنها نیمی از سؤالات این آزمون‌ها را تشکیل می‌دادند. تاکسونومی بیشتر سؤالات در طبقه I قرار داشت و بیشتر سؤالات دارای اشکال در قسمت ساقه یا گزینه‌ها بودند. لذا پیشنهاد می‌شود:

علاوه بر آموزش مداوم و مکرر طراحی سؤالات چند گزینه‌ای، با طراحان سؤالاتی که دارای اشکال ساختاری هستند و یا دارای ضریب دشواری پایین یا بالا و یا ضریب تمیز ضعیف بوده و یا دارای تاکسونومی پایین بوده و یا همه گزینه‌های انحرافی آن فاقد کارایی است در مورد علل طرح این دست سؤالات مصاحبه شود. هم‌چنین با گروهی از طراحان که سؤالات مناسبی طراحی کرده بودند نیز در مورد علل طرح این دست سؤالات مصاحبه شود. از طرفی با شرکت‌کنندگان در مورد ویژگی‌های سؤالات آزمون مصاحبه شود و اثربخشی ارائه پس‌خوراند به طراحان سؤال در طراحی بهتر سؤالات در سال‌های آینده تعیین شوند. هم‌چنین مسئولان امور آموزش و برگزاری آزمون‌ها نظارت بیشتری بر طراحی این آزمون‌ها داشته باشند، به نحوی که طراحان سؤالات الزام بر کاربرد چک لیست طراحی سؤالات مناسب داشته باشند و تاکسونومی سؤالات را تعیین

سؤال صرف کرده بودند و سؤالات با تاکسونومی بالا طراحی می‌کردند، سؤالات کم‌نقص تری طراحی کرده بودند.

بین طبقه تمیز با طبقه دشواری همبستگی ضعیف، معنی‌دار و منفی وجود داشت. به عبارت دیگر هر چه سؤالات سخت می‌شدند، میزان طبقه تمیز سؤالات کاهش پیدا می‌کرد؛ لذا به نظر می‌رسد که نقطه‌ای وجود داشته باشد که هم میزان دشواری و هم میزان تمیز در آن نقطه بهینه باشد تا سؤال کلیه عملکردهای خود را به خوبی ایفا نماید. نتایج پژوهش ما با نتایج پور میرزا کلهری و همکاران (۱۳۹۴)، Soraya و همکاران (۲۰۲۱) تطابق داشت (۲۳،۲۹) درحالی‌که Rao و همکاران (۲۰۱۶) و Kumar و همکاران (۲۰۲۱) همبستگی معنی‌دار و مثبتی را بین ضریب دشواری و ضریب تمیز پیدا کردند (۳۸،۳۹). پژوهش ما و دیگر پژوهش‌های انجام شده در این زمینه با تعداد کمی آزمودنی انجام شده است و سؤالات آزمون‌ها از درجات متفاوتی از اشکالات ساختاری، میزان دشواری و میزان تمیز و گزینه‌های انحرافی کارا برخوردار بودند که نتیجه‌گیری در مورد این ارتباط را به طور عملی دشوار می‌سازد.

بین تعداد گزینه‌های کارا و طبقه تمیز، همبستگی متوسط، معنی‌دار و منفی وجود داشت. به عبارت دیگر هر چه تعداد گزینه‌های کارا افزایش می‌یافت طبقه تمیز کاهش پیدا می‌کرد. شکورنیا (۱۳۹۸) بین کارایی گزینه‌های انحرافی و میزان تمیز سؤالات رابطه‌ای را پیدا نکرد، (۱۲) درحالی‌که Kheyami و همکاران (۲۰۱۸) همبستگی معنی‌دار و مثبتی را بین ضریب تمیز سؤالات و کارایی گزینه‌های انحرافی پیدا کردند. (۳۱) با توجه به اینکه در این آزمون‌ها بین طبقه دشواری و طبقه تمیز نیز همبستگی معنی‌دار و منفی وجود داشت به نظر می‌رسد که این یافته در پرتو همبستگی بین طبقه دشواری و طبقه تمیز باید مورد تعبیر و تفسیر قرار گیرد؛ لذا سؤالاتی که طبقه دشواری بالاتری داشتند، تعداد گزینه‌های انحرافی کارای بیشتر و طبقه تمیز کمتری داشتند. به نظر می‌رسد باید نقطه‌ای وجود داشته باشد که میزان طبقه دشواری و طبقه تمیز در آن بهینه باشد و گزینه‌های انحرافی کارا نیز بیشترین میزان باشد.

ملاحظات اخلاقی

این مقاله حاصل از پایان نامه کارشناسی ارشد مولف اول در دانشگاه پیام نور واحد کرج می باشد که با کد اخلاق IR.TUMS.MEDICINE.REC.1402.361 مورد حمایت معاونت پژوهشی دانشکده پزشکی دانشگاه علوم پزشکی تهران می باشد.

حمایت مالی

این پژوهش فاقد حمایت بوده است

تضاد منافع

تضاد منافی وجود ندارد.

شود و یا از افراد آموزش دیده در این زمینه استفاده شود تا ساختار و تاکسونومی سوال مورد ارزیابی قرار گیرد و سوالاتی که دارای ایراد ساختاری هستند به جمع سوالات نهایی برای آزمون راه پیدا نکند. با برگزاری جلسه های طراحی جمعی سوالات آزمون و مرور سوالات آزمون توسط همتایان سوالات دارای اشکالات ساختاری سوالات شناسایی شود و با بحث گروهی سوالات بدون ایراد ساختاری طرح شود پس خوراند در مورد میزان دشواری و تمیز سوالات و کارائی گزینه های انحرافی طرح شده به طراحان سوالات آزمون پس از برگزاری آزمون توسط مسئولان آموزشی ارائه شود. هم چنین با برگزاری جلسات مرور مجدد سوالات طرح شده بعد از برگزاری آزمون، به شناخت علل طرح سوالات آسان یا بسیار مشکل یا فاقد تمیز لازم و یا دارای گزینه های انحرافی بدون کارکرد پی برده شود تا از طرح این گونه سوالات در آینده پیشگیری شود.

تشکر و قدردانی

از مدیریت محترم وقت گروه قلب و عروق دانشگاه علوم پزشکی تهران جناب آقای دکتر سالاریفر، معاونت محترم وقت آموزش تخصصی و فوق تخصصی دانشکده پزشکی دانشگاه علوم پزشکی تهران جناب آقای دکتر علی لباف، مدیریت وقت امور آموزشی دانشگاه علوم پزشکی تهران سرکار خانم آزاده اشتیاقی و هم چنین سرکار خانم مهندس قریشی که برای انجام این پژوهش نویسندگان را یاری کردند تشکر می شود.

مراجع

- [1] Dellinges MA, Curtis DA. Will a Short Training Session Improve Multiple-Choice Item-Writing Quality by Dental School Faculty? A Pilot Study. J Dent Educ. 2017;81(8):948-955.
- [2] Gierl MJ, Bulut O, Guo Q, Zhang X. Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. Review of Educational Research. 2017;87(6):1082-116.
- [3] Sajjad M, Iltaf S, Khan RA. Nonfunctional distractor analysis: An indicator for quality of Multiple choice questions. Pak J Med Sci. 2020;36(5):982-986.
- [4] Abdulghani HM, Ahmad F, Irshad M, Khalil MS, Al-Shaikh GK, Syed S, Aldrees AA, Alrowais N, Haque S. Faculty development programs improve the quality of Multiple Choice Questions items' writing. Sci Rep. 2015;5:9556.

- [5] Dory V, Allan K, Birnbaum L, Lubarsky S, Pickering J, Young M. Ensuring the Quality of Multiple-Choice Tests: An Algorithm to Facilitate Decision Making for Difficult Questions. *Acad Med.* 2019;94(5):740.
- [6] Bhat SK, Prasad KHL. Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: A cross-sectional study. *Indian J Ophthalmol.* 2021;69(2):343-346.
- [7] Young M, Cummings BA, St-Onge C. Ensuring the quality of multiple-choice exams administered to small cohorts: A cautionary tale. *Perspect Med Educ.* 2017;6(1):21-28.
- [8] Tavakol M, Dennick R. The foundations of measurement and assessment in medical education. *Med Teach.* 2017;39(10):1010-1015.
- [9] Baharvand, M., Hoseinzadeh, M., Jaberiansari, Z., Abbaszadeh, E., Mortazavi, H. Evaluation of 30 Structural Validity and Content Quality Indices of Theoretical Dental Exams. *Journal of Mashhad Dental School.* 2014;38(4):291-302. [Persian]
- [10] Rostami-Moe M, Biglarkhani M, Meyari A. Quality of design and analysis of pre-internship exams for medical students of Hamadan University of Medical Sciences in the years 2018 to 2020. *JMED* 2022;14 (44):36-42.
- [11] Khajeali N, Aslami M, Araban M. Analysis of medical and dentistry basic sciences examinations: A case study. *Payesh* 2020;19(4):383-389. [Persian]
- [12] Shakurnia A. A survey on distractors in multiple-choice questions and its relationship on difficulty and discriminative indices. *Iranian Journal of Medical Education* 2019;19(20):180-188. [Persian]
- [13] Jalili M, Khabaz-Mafinejad M, Gandomkar R, Mourtaz-Hejri S. Principles and methods of student assessment in health professions. *The Academy of Medical Sciences Islamic Republic of Iran.* 2016, pp.154,693,697. [Persian]
- [14] Salam A, Yousuf R, Bakar SM. Multiple choice questions in medical education: how to construct high quality questions. *Int J Hum Health Sci.* 2020;4(2):79.
- [15] Gajjar S, Sharma R, Kumar P, Rana M. Item and Test Analysis to Identify Quality Multiple Choice Questions (MCQs) from an Assessment of Medical Students of Ahmedabad, Gujarat. *Indian J Community Med.* 2014;39(1):17-20.
- [16] Adeosun SO. Differences in Multiple-Choice Questions of Opposite Stem Orientations Based on a Novel Item Quality Measure. *Am J Pharm Educ.* 2023;87(2):ajpe8934.
- [17] Anbari Z, Jadidi R. Assessment of the resident's promotion exam: One step to validity of competency measurement in Arak University of Medical Sciences. *JMED.* 2013;7(4):52-62. [Persian]
- [18] Meyari A, Biglarkhani M, Zandi M, Vahedi M, Miresmaeili A F. The Effect of Education on Improvement of Multiple Choice Questions' Designing in Annual Residency Exams of Dental School. *Iranian Journal of Medical Education* 2012;12 (1):36-45. [Persian]
- [19] Taheri M, Khoshrang H, Asadi Louyeh A, Hidarzadeh A. Quality of Residents' Promotion Exams before & after Educational Intervention in 2010-2011. *Iranian Journal of Medical Education* 2013;13 (7):551-560. [Persian]
- [20] Sayyadmelli M, Barzegar M, Bilan N, Aslanabadi S, Khoshbaten M, Ghasem Zadeh A, Karegar-Maher, M. Comparison of multiple-choice questions quality parameters of pediatric, general surgery, internal medicine and gynecology and obstetrics residency tests between preboard examination of Tabriz University of Medical Sciences and national board examination in 2010 and 2011. *JMED.* 2015; 8(18):43-53. [Persian]
- [21] Rodríguez-Díez MC, Alegre M, Díez N, Arbea L, Ferrer M. Technical flaws in multiple-choice questions in the access exam to medical specialties ("examen MIR") in Spain (2009-2013). *BMC Med Educ.* 2016;16:47.
- [22] Pourmirza Kalhori R, Rezeai M, Karami Matin B, Roshan Pour F. A survey of quality and quantity indexes of multiple choice question (MCQ) exams of medical residents at Kermanshah University of Medical Sciences: 2008-2012. *JMED.* 2014; 8(4):64-75. [Persian]
- [23] Pourmirza Kalhori R, Rezaie M, Shojee Moghadam AR, Sepahi V, Memar Eftekhari L . Correlation of quality and quantity index of multiple choice questions exams of residency promotion in Kermanshah University of Medical Sciences, 2013. *J Clin Res Paramed Sci.* 2015;4(1):e81958. [Persian]

- [24] Sabouri M, Arbabi F, Dehghani Poudeh M. The investigation into the board examinations in majors related to adult cancer in Iran. *Iranian Journal of Medical Education* 2020; 20:272-278. [Persian]
- [25] Derakhshan F, Allami A, Ahmady S. Quantitative and qualitative indicators evaluation of residency exams in Qazvin University of Medical Sciences (2012-13). *JMED*. 2016;8 (20):29-38. [Persian]
- [26] Meyari, A., Beiglarkhani, M. Improvement of design of multiple choice questions in annual residency exams by giving feedback. *Strides in Development of Medical Education*. 2013;10(1):109-118. [Persian]
- [27] Capan Melsner M, Steiner-Hofbauer V, Lilaj B, Agis H, Knaus A, Holzinger A. Knowledge, application and how about competence? Qualitative assessment of multiple-choice questions for dental students. *Med Educ Online*. 2020;25(1):1714199.
- [28] Shigli K, Nayak SS, Gali S, Sankeshwari B, Fulari D, Shyam Kishore K, Upadhya PN, Jirge V. Are multiple choice questions for post graduate dental entrance examinations spot on? Item analysis of MCQs in prosthodontics in India. *J Natl Med Assoc*. 2018;110(5):455-458.
- [29] Soraya S, Shabani A, Kamalzadeh L, Kashaninasab F, Rashedi V, Saeid MI, Seddigh R, Asadi S. Predictability of discrimination coefficient and difficulty index of psychiatry multiple-choice questions. *Journal of Iranian Medical Council*. 2021;4(3):165-72.
- [30] Shakurnia A, Ghafourian Boroujerdnia M, Khodadadi A, Ghadiri A, Amari A. Analytical study of quantitative indices of multiple-choice questions of immunology department in Ahvaz Jundishapur University of Medical Sciences. *Educational Development of Judishapur*, 2018;9(2):72-83. [Persian]
- [31] Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item analysis of multiple choice questions at the department of paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos Univ Med J*. 2018;18(1):e68-e74.
- [32] Mehmood B, Sultan S, Azhar M, Shakoor A, Khan NB, Niazi Z. Quality analysis of multiple choice questions. *Pakistan Oral & Dental Journal*. 2020;40(4):236-9.
- [33] Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. *Int J Appl Basic Med Res*. 2016;6(3):170-3.
- [34] Khare AS, Kadam AA, Verma A, Gandhi KL. Evaluation of Distractor Effectiveness of Single Best Response type of Multiple Choice Questions by Item Analysis. *European Journal of Molecular & Clinical Medicine (EJMCM)*.;9(07):2022.
- [35] Salih KEMA, Jibo A, Ishaq M, Khan S, Mohammed OA, Al-Shahrani AM, Abbas M. Psychometric analysis of multiple-choice questions in an innovative curriculum in Kingdom of Saudi Arabia. *J Family Med Prim Care*. 2020 30;9(7):3663-3668.
- [36] Shaikh S, Kannan SK, Naqvi ZA, Pasha Z, Ahamad M. The Role of Faculty Development in Improving the Quality of Multiple-Choice Questions in Dental Education. *J Dent Educ*. 2020; 84(3):316-322.
- [37] Bhattacharjee S, Mukherjee A, Bhandari K, Rout AJ. Evaluation of Multiple-Choice Questions by Item Analysis, from an Online Internal Assessment of 6th Semester Medical Students in a Rural Medical College, West Bengal. *Indian J Community Med*. 2022; 47(1):92-95.
- [38] Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivas V. Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Med J Armed Forces India*. 2021;77(Suppl 1):S85-S89.
- [39] Rao C, Prasad HK, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. *International Journal of Educational and Psychological Researches*. 2016;2(4):201.