

Evaluation of multiple-choice and essay questions of three consecutive courses of Saez health educational center

Gelavij Ghafouri MSc student in Community Based Education of Health sciences, School of Medical Education and Learning Technologies, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
Peigham Heidarpoor* Faculty member, Community Based Education of Health sciences, School of Medical Education and Learning Technologies, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Abstract

Introduction Behvarz training centres are recognized as primary training centres for community health workers. The aim of this study was to assess the quality of descriptive multiple-choice and essay questions across three cohorts at the Saez Behvarz Training Center.

Methods This descriptive study analyzed a total of 2077 multiple-choice questions, 75 essay questions, and 360 multiple-choice questions related to the final exams of three recent courses. The questions were evaluated based on their structure, discrimination (differentiation) and difficulty indices, and taxonomy levels. A checklist was developed and utilized following a comprehensive review of relevant literature and consensus among medical education experts. The checklist demonstrated an interrater reliability of 90%, based on agreement calculations between two independent evaluators.

Results Analysis of the multiple-choice questions' structure revealed that 82.5% of the test questions possessed an appropriate structure. The primary structural flaw identified was the vertical listing of options. Regarding taxonomy, 65.48% of the multiple-choice questions were categorized at Taxonomy Level I, 23.92% at Level II, and 10.6% at Level III.

Conclusion While the questions demonstrated an appropriate difficulty index, their discrimination index was inadequate. The majority of questions belonged to the lower cognitive domain. Although multiple-choice questions were generally better designed than essay questions, it is recommended that all questions undergo review and revision. Furthermore, providing training on standard design principles for various types of student evaluation questions is essential to empower trainers.

Keywords: *differentiation index, difficulty index, multiple choice questions*

* Corresponding Author

E-mail: peigham.heidarpoor@yahoo.com

Extended Abstract

Introduction

Behvarz Training Centre (BTC) is one of the most sensitive training centres of medical universities, where students receive theoretical and practical training and start working immediately after completing their training. To ensure the best possible evaluation of desired learning outcomes, it is important to use the most efficient and appropriate assessment methods. This will improve students' learning, which is the main goal of the evaluation program. There is a scarcity of research on the assessment of BTC students in Iran. The aim of this study is to assess the quality of multiple-choice and essay questions at Saqez Behvarz Training Center from 2016 to 2021. The results of this study can improve the quality of descriptive and multiple-choice questions and the final exam of health education centres based on existing standards. Additionally, each instructor can benefit from the results obtained in designing the questions of their respective courses so that they can reduce the weak points to the minimum possible level.

Methods

The current study is descriptive and has the approval of the ethics committee with the ID IR.SBMU.SME.REC.1400.077 on December 14, 2021. The data of this study were prepared from the available documents in the Health Education Center of Saqez City. In this research, 2,077 multiple-choice questions, 75 descriptive exam papers, 27 practical work checklists, and final exam questions were examined during three courses (the 22nd, 23rd, and 24th courses). The tools used were valid and reliable checklists. The validity of the checklists was verified by Medical Education Experts of Saqez City, and the reliability of the checklists was also measured by calculating the agreement between two expert evaluators, with the percentage of agreement for all checklists above 90%. The differentiation index and difficulty of the final tests were also calculated based on the formula. Data analysis was done using SPSS version 25 software.

Results

Multiple-choice questions extracted from three training courses were evaluated with the 12-question Millman checklist. In examining multiple-choice questions, the most correct answer rate (98.84%) and the least correct answer rate (16.18%) were related to the transparency and directness of the wording and the vertical listing of the options, respectively. Analysis of multiple-choice questions in various

studies shows that multiple-choice questions are designed for information recall and memorization; for this reason, a variety of competency assessments are used nowadays, but this method remains as a primary assessment method. In this study, the instructors were more careful in designing some multiple-choice questions than others, and most of the questions were designed appropriately. One of the criteria was the correct spelling of the words used in the stem and options and whether the question evaluates a specific learning goal or not. Also, the questions of some courses, such as workplace health, teenagers and schools, and environmental health, were very suitable and correct more than 80%, and principled question design was observed. It is suggested that trainers identify ideal multiple-choice questions and add some of them to a future question bank. Taxonomically, 65.48% of multiple-choice questions were designed in Taxonomy Level I, 23.92% in Taxonomy Level II, and 10.6% in Taxonomy Level III.

The high percentage of questions in Taxonomy Level I indicates that the instructors of the health training center have designed a higher percentage of multiple-choice questions to measure students' recall. In similar studies, question designers also placed a high percentage of questions at the awareness and recall levels.

Learners who have knowledge and skills in analysis, interpretation, organization, and reporting tend to get higher scores in descriptive questions. In this study, the scores of three courses for descriptive tests were distributed as very good 38.5%, good 19.8%, average 7.9%, and poor 33.8%, respectively. Improving the evaluation and scoring of descriptive tests requires using different and higher-level methods, so educators should consider these in training evaluation. The descriptive tests taken at this Health Education Center during the research period were limited; therefore, it is suggested that instructors use descriptive questions more often to encourage students to have a deeper approach to learning.

To evaluate the difficulty index of the final exam, 57.5% of the questions were optimal, and 27.5% were easy. Also, regarding the differentiation index, 62.8% of the questions were standard, and 8.9% were non-standard. In this study, the differentiation index and difficulty across the three courses were not significantly different; however, a similar study conducted over a 10-year period showed that in the last years of that period, compared to earlier years, the differentiation index increased—in other words, questions became more difficult. This study examined courses held over five years. To better

study the difficulty and differentiation indices and obtain clearer results about educational evaluation, it is suggested to examine courses over a longer period of at least 10 years.

Conclusion

In general, multiple-choice questions were designed appropriately by competent instructors, and the majority of these questions were aimed at measuring students' recall and knowledge. The difficulty index was acceptable, and based on the discrimination index, the percentage of standard questions was high. It is advisable for educators with a high level of knowledge and expertise to pay attention to the relevant principles in designing descriptive questions so that the quality of evaluation through this method can be improved as much as possible. Additionally, weaknesses in test design affecting learner evaluation can be addressed through the training of trainers and standardization of tests under expert supervision.

Keywords: *differentiation index, difficulty index, multiple choice questions*

ارزشیابی سؤالات چندگزینه‌ای و تشریحی سه دوره متوالی مرکز آموزش سلامت شهرستان سقز

گلاویژ غفوری

پیغام حیدرپور*

دانشجو کارشناسی ارشد رشته آموزش جامعه‌نگر در نظام سلامت، دانشکده آموزش پزشکی و فناوری‌های یادگیری، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

عضو هیأت علمی، بخش آموزش جامعه‌نگر در نظام سلامت، دانشکده آموزش پزشکی و فناوری‌های یادگیری، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

چکیده

اهداف مراکز آموزش بهورزی به‌عنوان یکی از مراکز اصلی آموزش کارکنان سلامت جامعه محسوب می‌شود. هدف این مطالعه ارزشیابی کیفیت سؤالات چندگزینه‌ای، تشریحی سه دوره آموزشی در مرکز آموزش بهورزی شهرستان سقز است.

روش در این مطالعه توصیفی، کلیه سؤالات چندگزینه‌ای (۲۰۷۷ سؤال) و تشریحی (۷۵ آزمون) و ۳۶۰ سؤال چندگزینه‌ای مربوط به آزمون نهایی سه دوره اخیر از نظر ساختار سؤالات، ضریب تمیز و ضریب دشواری و سطح تاکسونومی سؤالات بررسی شد. چک لیست استاندارد بر اساس مرور متون و اجماع متخصصین حوزه آموزش پزشکی تهیه و مورد استفاده قرار گرفت. پایایی چک لیست‌ها نیز از طریق محاسبه توافق بین دو ارزیاب بالای ۹۰ درصد بود.

یافته‌ها در بررسی ساختار سؤالات چندگزینه‌ای ۸۲/۵ درصد سؤالات تستی ساختار مناسب داشته، بیشترین مشکل لیست کردن عمودی سؤالات بود. ۶۵/۴۸ درصد از سؤالات چندگزینه‌ای در تاکسونومی I، ۲۳/۹۲ درصد تاکسونومی II و ۱۰/۶ درصد در تاکسونومی III طراحی شده بودند. در سؤالات تشریحی ساختار خیلی خوب، خوب، متوسط و ضعیف به ترتیب ۴۷/۰۳، ۱۱/۴، ۷/۷۷ و ۳۳/۸ درصد از کل سؤالات را تشکیل می‌دادند. با اندازه‌گیری شاخص دشواری آزمون نهایی، ۵۷/۵ درصد سؤالات دشواری قابل قبول داشته و ضریب تمیز ۲۲/۵ درصد سؤالات قابل قبول و ۲۵/۳ درصد آن‌ها ضریب تمایز خوب داشتند.

نتیجه‌گیری سؤالات ضریب دشواری مناسبی و ضریب تمیز نامناسب داشتند. سؤالات بیشتر در حیطة پایین شناختی قرار داشت. گرچه طراحی سؤالات چندگزینه‌ای از سؤالات تشریحی مناسب‌تر بود. لذا پیشنهاد می‌شود سؤالات مرور و اصلاح شده و آموزش‌هایی برای طراحی استاندارد سؤالات مختلف و ارزیابی فراگیران برای توانمندسازی مربیان برنامه‌ریزی شود.

کلیدواژه‌ها: شاخص تمایز، شاخص دشواری، سؤالات چندگزینه‌ای

* نویسنده مسئول

پست الکترونیکی: peigham.heidarpoor@yahoo.com

مناطق روستایی داشته و دارند. با پیشرفت‌های اخیر، در سطح تحصیلات و صلاحیت‌های بهروزان کمتر از انتظار مردم است و این باعث می‌شود مردم مراجعه کمتری به مراکز برای دریافت آموزش سلامت داشته باشند. یکی از چالش‌های عملکردی بهروزی در ایران تعداد و صلاحیت ناکافی آنها می‌باشد. بیشتر آنها دانش، نگرش و عملکرد مناسبی ندارند که به علت سطح تحصیلات پایین و آموزش ناکافی آنهاست (۷). نیازهای بهداشتی جامعه، اهمیت آموزش‌های علمی و عملی بهروزان را که خط اول ارائه خدمات به جامعه را دارند، بیش از پیش مشخص می‌کند. یکی دیگر از چالش‌های مهم آموزشی این افراد استفاده از مربیانی است که از روش‌های آموزش سنتی استفاده می‌کنند (۸). طرح سؤالات به صورت استاندارد و تعیین کیفیت آزمون پس از ارزیابی یکی از موارد اصلی مورد توجه در ارزیابی فراگیران محسوب می‌شود و باید مورد توجه تیم ارزیابی باشد (۹). مهارت طرح سؤال استاندارد امروزه به عنوان یک مهارت مهم مطرح است. از طرف دیگر ضریب تمیز که به وسیله آن فراگیر قوی و ضعیف را متمایز می‌کند و ضریب دشواری که میزان درصد پاسخ درست فراگیران را مشخص می‌کند، از مواردی است که اعتبار و روایی آزمون را افزایش می‌دهد (۱۰).

مطالعات مختلف به بررسی کمی و کیفی سؤالات مختلف دانشکده‌های مختلف علوم پزشکی پرداخته‌اند ولی طبق بررسی ما، تاکنون مطالعه‌ای با این عنوان در مراکز آموزش بهروزی که منحصر به کشور ایران است انجام نشده است. لذا هدف از این مطالعه ارزشیابی کیفیت ساختار انواع سؤالات چندگزینه‌ای، تشریحی، انواع تاکسونومی سؤالات چند گزینه‌ای و ضریب دشواری و تمیز سؤالات نهایی سه دوره آموزش بهروزی در مرکز آموزش سلامت شهرستان سقز بین سال‌های ۱۳۹۶ تا ۱۴۰۰ می‌باشد. نتایج حاصل از این مطالعه می‌تواند سبب ارتقای کیفیت سؤالات تشریحی، چندگزینه‌ای و آزمون نهایی مراکز آموزش سلامت بر اساس استانداردهای موجود شود و هر کدام از مربیان از نتایج به دست آمده در طراحی سؤالات دروس مربوط به خود بهره برده به نحوی که نقاط ضعف را به حداقل ممکن کاهش دهند.

روش

مطالعه حاضر یک مطالعه توصیفی مقطعی می‌باشد. در این پژوهش سؤالات کلیه تستی و تشریحی سه دوره آموزشی مرکز آموزش بهروزی شهرستان سقز مورد بررسی قرار گرفته است. این مرکز تا کنون ۲۵ دوره آموزشی را پذیرش نموده است و آموزش‌دیدگان در شهرستان‌های سقز، بانه، بیجار و دیواندره مشغول به خدمت می‌باشند. با بررسی مستندات

ارزیابی فراگیر پایه و اساس فرایند یادگیری محسوب می‌شود که فراتر از اندازه‌گیری دانش و صلاحیت به تنهایی است. ارزیابی در هدایت و تحریک فرایند یادگیری نقش اساسی دارد. علاوه بر این بازخوردهایی را برای مدرسان و فراگیران مهیا می‌کند. سه حیطة یادگیری شامل دانش و درک، مهارت و ارزش قابل ارزشیابی می‌باشند. روش‌های متعددی برای ارزیابی مهارت وجود دارد که هر کدام منافع و مشکلاتی دارد. ترکیب این روش‌ها برای اطمینان از روایی و پایایی آنها ضروری است (۱). ارزیابی فراگیران، میزان و کیفیت مطالعه فراگیران و تلاش آنها را تحت تاثیر قرار می‌دهد. ارزیابی برای یادگیری امروزه از ارزیابی یادگیری بیشتر مورد توجه قرار می‌گیرد (۲). روش‌های مورد استفاده برای ارزیابی دانش و کاربردهای آن شامل سؤالات چندگزینه‌ای، سؤالات پاسخ کوتاه، سؤالات تشریحی پاسخ بلند و امتحان شفاهی است (۳). اینکه دانشجویان چه چیزی و چگونه یاد می‌گیرند تا حد زیادی به نحوه ارزیابی آنها بستگی دارد. درک دانشجویان از سطح دشواری ارزیابی‌های مختلف منعکس کننده عملکرد آنها در امتحانات می‌باشد. بازخورد برای بهبود و جستجوی بازخورد به موقع و با کیفیت بهتر و شخصی مهم است (۴). روش ارزشیابی آموزش باید بسته به استفاده از کارآمدترین و مناسب‌ترین روش برای ارزیابی نتایج یادگیری مورد نظر باشد و بهبود یادگیری دانشجویان هدف اصلی برنامه ارزشیابی در نظر گرفته می‌شود. گزارش نتایج ارزیابی برای ذی‌نفعان باید واضح، جامع و قابل درک باشد تا آنها بتوانند دوره آموزشی را به طور مؤثر ارزشیابی و برای بهبود بیشتر بازبینی کنند (۵). صلاحیت‌های مدرس، فراوانی و طول اجرای آموزش، پیشینه نظری و ابزارهای آموزشی مرتبط با برنامه آموزشی برای طراحی، اجرا و ارزشیابی برنامه‌های آموزشی ضروری است (۶).

مراکز آموزش سلامت (آموزش بهروزی) یکی از حساس‌ترین مراکز آموزشی دانشگاه‌های علوم پزشکی است که بهروزان در آن آموزش دیده و بعد از به اتمام رساندن دوره آموزشی مستقیم در خانه‌های بهداشت مشغول به کار خواهند شد. مراکز آموزش بهروزی بر اساس دستورالعمل وزارتی فراگیران خود را بر اساس دوره‌های آموزشی دو ساله پذیرش می‌کند که فراگیران در طی این دو سال باید ۵۴ واحد درس عملی - نظری و ۲۷ واحد درس کارآموزی را بگذرانند. مواد درسی، تعداد ساعات و تقویم زمانی برنامه آموزش بهروزی بر اساس نظام واحدی به مقدار ۳۶۱۵ ساعت در مدت ۲۴ ماه معادل ۸۱ واحد با احتساب دروس نظری عملی و کارآموزی می‌باشد. بهروزان به عنوان کارکنان سلامت جامعه در ایران نقش غیر قابل انکاری در بهبود سلامت مردم به خصوص در

تواند فراگیران قوی را از ضعیف متمایز سازد. ضریب تمایز صفر یعنی سؤال قادر به تمایز فراگیر قوی از ضعیف نمی‌باشد ضریب تمایز منفی مناسب نیست و نشان‌دهنده این است که فراگیران ضعیف بهتر به سؤال جواب داده‌اند.

ضریب دشواری از طریق فرمول $P = H + L/N \times 100$ محاسبه می‌شود. اگر این شاخص کمتر از ۳۰ درصد باشد سؤال دشوار، و اگر بین ۳۰ تا ۷۰ درصد باشد مناسب، و اگر بیش از ۷۰ درصد باشد، سؤال آسان است. داده‌ها با استفاده از نرم‌افزار اکسل ۲۰۱۰ و SPSS نسخه ۲۳ تحلیل شدند. جهت شروع مطالعه مصوبه کمیته اخلاق در پژوهش دانشگاه علوم پزشکی شهید بهشتی با شناسه IR.SBMU.SME.REC.1400.077 دریافت شد. مطابق دستورالعمل راهنمای اخلاق در پژوهش با رعایت ملاحظات اخلاقی در تحلیل داده‌ها نامی از مربی و دانشجو برده نشده و پژوهش بدون هرگونه کپی‌برداری از مطالعات مشابه انجام شده است.

یافته‌ها

ارزیابی سؤالات چندگزینه‌ای از نظر اصول ساختاری و مطابقت با اصول میلن: سؤالات چندگزینه مستخرج از سه دوره آموزشی با چک لیست میلن ۱۲ سؤالی ارزیابی شد. نتیجه آن در جدول ۱ به تفکیک دروس و سؤالات چک لیست میلن مشاهده می‌شود. بر اساس این چک لیست دروسی که بیشترین درصد پاسخ مناسب را به خود اختصاص داده‌اند، دروس سلامت محیط کار (۸۷/۶ درصد)، نوجوانان و مدارس (۸۶/۲ درصد) و بهداشت محیط (۸۶/۱ درصد) می‌باشند. دروسی که بیشترین جواب نامناسب را از چک لیست گرفته‌اند عبارتند از: سلامت باروری (۲۳/۸ درصد)، تعاملات سازمانی (۲۲/۸ درصد) و درمان‌های ساده علامتی (۲۰/۷ درصد). تطابق با چک لیست بقیه دروس در حد فاصل ۷۹ تا ۸۶ درصد بود. بیشترین تعداد سؤالات مربوط به دروس بیماری‌ها و واگیر (۳۲۲ سؤال، ۱۵/۵ درصد) و غیرواگیر (۲۳۳ سؤال، ۱۱/۳ درصد) و کمترین مربوط به دروس تغذیه (۵۲ سؤال، ۲/۵ درصد) و شکایت‌های شایع (۵۳ سؤال، ۲/۶ درصد) بود.

شفافیت و مستقیم بودن لغات در ساقه و گزینه در ۹۸/۸۴ درصد (۲۰۵۳ مورد) بیشترین و عمودی لیست شدن گزینه‌ها (۳۳۶ مورد، ۱۶/۱۸ درصد) کمترین مواردی بودند که رعایت شده بود. ۷۴ درصد سؤال‌ها طبق چک لیست میلن مناسب بودند. در ۹۸/۷۵ درصد سؤالات و گزینه‌ها مشکل املائی وجود نداشت. در ۷۴/۱۰ درصد سؤالات به‌طور مشخص یک گزینه صحیح وجود داشت. در کل سؤالات طبق چک

موجود برگه‌های امتحانی و سؤالات تشریحی سه دوره آموزشی شامل دوره ۲۲ (سال‌های ۱۳۹۵ تا ۱۳۹۷)، دوره ۲۳ (سال‌های ۱۳۹۷ تا ۱۳۹۹) و دوره ۲۴ (سال‌های ۱۳۹۹ تا ۱۴۰۱) به تعداد ۷۵ برگه امتحانی در این مطالعه ارزیابی شد. درس‌های بررسی شده شامل کمک‌های اولیه، شکایت‌های شایع، بیماری‌های غیرواگیر، سالمندان، کودکان، مادران، تعاملات سازمانی، آموزش بهداشت، بیماری‌های واگیر، میانسالان، درمان‌های ساده علامتی، سلامت باروری، بهداشت محیط، بهداشت دهان و دندان، ایمن‌سازی، جوانان، نوجوانان و مدارس می‌باشد. تعداد ۲۰۷۷ سؤال چندگزینه‌ای و سؤالات آزمون نهایی سه دوره آموزشی (هر دوره ۱۲۰ سؤال) نیز ارزیابی و آنالیز شد. تعداد برگه‌های سؤالات تشریحی دوره‌های آموزشی ۲۲، ۲۳ و ۲۴ نیز به ترتیب ۲۶، ۲۰ و ۲۹ بود. روش نمونه‌گیری به صورت در دسترس بود و تمام آزمون‌های تستی و تشریحی این مرکز بررسی شد. معیار ورود به مطالعه، نمونه سؤالات مطرح شده در دوره بیست و یکم به بعد و معیار خروج برگه‌های امتحانی با مصرف بیشتر از یک بار و تکراری بود. کیفیت ساختار سؤالات چندگزینه‌ای و تشریحی و ضریب تمیز و دشواری آزمون فینال سه دوره با ابزارهای مناسب بررسی شد. برای بررسی ساختار سؤالات چندگزینه‌ای از چک لیست ۱۲ سؤال میلن استفاده شد که در مطالعات مختلف ایرانی و خارجی استفاده می‌شود. چک لیست سؤالات تشریحی براساس استانداردهای حاصل از مرور متون استخراج و روایی آن توسط متخصصین آموزش پزشکی مورد تأیید قرار گرفت. جهت بررسی پایایی این چک لیست نیز چند مورد از انواع سؤالات ارزشیابی توسط دو ارزیاب به صورت مستقل بررسی و در همه این موارد درصد توافق بین دو ارزیاب بالاتر از ۹۰ درصد برآورد شد. سؤالات گزینه‌ای از نظر سه سطح شناختی بلوم نیز بررسی شدند. سطح یک سؤالات دانش و یادآوری اطلاعات، سطح دوم سؤالات مربوط به فهم و تفسیر داده‌ها و سطح سوم کاربرد دانش به‌منظور حل مشکلات را شامل می‌شود. برای بررسی اعتبار آزمون‌های نهایی هر دوره از ضریب تمیز و ضریب دشواری استفاده شد. ضریب تمیز افتراق دانشجویان قوی از ضعیف را امکان‌پذیر می‌سازد در حالی که ضریب دشواری برای اندازه‌گیری نسبتی از آزمون‌شوندگان که به سؤال جواب درست می‌دهند، استفاده می‌شود. در شاخص تمیز، از فرمول $d = H - L \times 2/N$ استفاده می‌شود که تعداد افرادی که در گروه بالا به سؤال پاسخ صحیح داده‌اند (H) و تعداد افرادی که در گروه پایین به سؤال پاسخ صحیح داده‌اند (L) و N تعداد دانشجویان دوگروه می‌باشد. این شاخص عددی بین -۱ تا ۱ تغییر می‌کند. هرچه ضریب تمیز سؤالی به عدد یک نزدیک‌تر باشد، بهتر می

لیست میلمن در ۸۲/۵ درصد موارد ساختار مناسب و ۱۸/۵ درصد ساختار نامناسب داشتند.

جدول ۱. مطابقت سؤالات چندگزینه‌ای مرکز آموزش بهوزری سقز با اصول میلمن

Table 1. Compliance with each of Millman's principles according to indices and divided by all courses of Saqez Behvarz Training center

ردیف	نام درس	سؤال		میزان مجموع تمامی سؤالات هر درس با چک لیست میلمن	
		تعداد	درصد	تعداد	درصد
۱	کمک‌های اولیه	۱۱۴	۵/۵	۱۱۵۳	۸۴/۳
۲	شکایت‌های شایع	۵۳	۲/۶	۵۱۲	۸۰/۵
۳	بیماری‌های غیرواگیر	۲۳۳	۱۱/۲	۲۲۹۹	۸۲/۲
۴	سالمدان	۹۲	۴/۴	۹۱۸	۸۳/۱
۵	کودکان	۱۰۲	۴/۹	۱۰۱۷	۸۳/۱
۶	مادران	۷۴	۳/۶	۷۳۰	۸۲/۲
۷	تعاملات سازمانی	۶۰	۲/۹	۵۵۶	۷۷/۲
۸	آموزش بهداشت	۱۳۷	۶/۶	۱۳۴۷	۸۱/۹
۹	بیماری‌های واگیر	۳۲۲	۱۵/۵	۳۰۸۵	۷۹/۸
۱۰	میانسالان	۱۰۳	۵	۱۰۰۵	۸۱/۳
۱۱	درمان‌های ساده علامتی	۵۸	۲/۸	۵۵۲	۷۹/۳
۱۲	سلامت باروری	۶۱	۲/۹	۵۵۸	۷۶/۲
۱۳	بهداشت محیط	۱۳۰	۶/۳	۱۳۴۳	۸۶/۱
۱۴	دارونامه	۵۵	۲/۶	۵۲۸	۸۰
۱۵	تغذیه	۵۲	۲/۵	۵۳۱	۸۵/۱
۱۶	سلامت محیط کار	۱۴۴	۶/۹	۱۵۱۹	۸۷/۹
۱۷	بهداشت دهان و دندان	۶۹	۳/۳	۶۸۶	۸۲/۸
۱۸	ایمن‌سازی	۷۸	۳/۸	۷۸۸	۸۴/۲
۱۹	جوانان	۶۵	۳/۱	۶۵۱	۸۳/۵
۲۰	نوجوانان و مدارس	۷۵	۳/۶	۷۷۶	۸۶/۲
	جمع	۲۰۷۷	۱۰۰	۲۰۵۵۴	۸۲/۵

در سطح تاکسونومی II طراحی شده‌اند. درس تغذیه که بعد از درس تعاملات سازمانی در تاکسونومی I بیشترین درصد (۹۲/۳ درصد) را به خود اختصاص داده بود، در تاکسونومی III هیچ سؤالی نداشته و درس کودکان با ۲۶/۵ درصد بیشترین سؤالات را در تاکسونومی III داشته است. فقط ۱۰/۶ درصد کل سؤالات در سطح III تاکسونومی طراحی شده بود.

ارزیابی تاکسونومی سؤالات چندگزینه‌ای: مطابق نتایج جدول ۲ بیشترین درصد سؤالات (۶۵/۴۸ درصد) در تاکسونومی I طراحی شده بودند. درس تعاملات سازمانی با ۹۳/۳ درصد بیشترین و درس کودکان با حدود ۴۵ درصد کمترین سؤالات را در سطح تاکسونومی I دارند. دروس درمان‌های ساده، شکایت‌های شایع، دارونامه و مادران در تاکسونومی II هیچ سؤالی نداشته در عوض درس باروری با ۳۶ درصد بیشترین سؤال را در این سطح داشته است. ۲۳/۹۲ درصد کل سؤالات

جدول ۲. توزیع فراوانی انواع تاکسونومی سؤالات چند گزینه‌ای سه دوره بهورزی شهرستان سقز

Table 2. Frequency distribution of taxonomy types of multiple-choice questions in three cohort of Saez Behvarz Training center

نام درس	تاکسونومی I		تاکسونومی II		تاکسونومی III	
	تعداد	درصد	تعداد	درصد	تعداد	درصد
کمک‌های اولیه	۵۸	۵۰/۸۸	۳۲	۲۸/۰۷	۲۴	۲۱/۰۵
شکایت‌های شایع	۴۰	۷۵/۴۷	۰	۰	۱۳	۲۴/۵
بیماری‌های غیر واگیر	۱۴۰	۶۰/۱	۷۶	۳۲/۶	۱۷	۷/۳
سالمندان	۵۴	۵۸/۷	۳۱	۳۳/۷	۷	۷/۶
کودکان	۴۵	۴۴/۱	۳۰	۲۹/۴	۲۷	۲۶/۵
مادران	۵۷	۷۷	۰	۰	۱۷	۲۳
تعاملات سازمانی	۵۶	۹۳/۳۳	۱	۱/۶۷	۳	۵
آموزش بهداشت	۹۱	۶۶/۴۲	۴۰	۲۹/۲	۶	۴/۳۸
بیماری‌های واگیر	۲۱۰	۶۵/۲	۹۶	۲۹/۸	۱۶	۵
میانسالان	۵۵	۵۳/۴	۳۰	۲۹	۱۸	۱۶
درمان‌های ساده علامتی	۴۷	۸۱/۰۳	۰	۰	۱۱	۱۸/۹۷
سلامت باروری	۳۴	۵۵/۸	۲۲	۳۶	۵	۸/۲
بهداشت محیط	۸۷	۶۶/۹۲	۴۱	۳۱/۵۴	۲	۱/۵۴
دارونامه	۴۸	۸۷/۲۷	۰	۰	۷	۱۲/۷۳
تغذیه	۴۸	۹۲/۳	۴	۷/۷	۰	۰
سلامت محیط کار	۱۰۴	۷۲/۲	۳۵	۲۴/۳	۵	۱/۷۴
بهداشت دهان و دندان	۵۴	۷۸/۲۶	۱۴	۲	۱	۱/۴۵
ایمنسازی	۴۰	۵۱/۳	۲۳	۲۹/۵	۱۵	۱۹/۲
جوانان	۴۶	۷۰/۸	۵	۷/۷	۱۴	۲۱/۵
نوجوانان و مدارس	۴۶	۶۱/۳	۱۷	۲۲/۷	۱۲	۱۶
جمع	۱۳۶۰	۶۵/۴۸	۴۹۷	۲۳/۹۲	۲۲۰	۱۰/۶

ماکزیمم امتیاز برای دوره‌های ۲۲، ۲۳ و ۲۴ به ترتیب ۲۸۶، ۲۲۰ و ۳۱۹ بود. ۴۴/۲ درصد امتیازها خیلی خوب، ۱۱/۶ درصد امتیازها خوب، ۷/۸ درصد متوسط و ۳۳/۴ درصد امتیازها ضعیف بودند.

نتایج آزمون تشریحی بر اساس دوره‌های آموزشی آنالیز و در جدول شماره ۳ نشان داده شده است. به ازای هر سؤال آزمون ۱۱ امتیاز در نظر گرفته شده است. تمام سؤالات در سه دوره وزن یکسانی داشتند.

جدول ۳. توزیع فراوانی تطابق سؤالات تشریحی با چک لیست استاندارد سه دوره به‌پوزی شهرستان سقز

Table 3. Compliance with each of standard essay question checklist in three cohort of Saqez Behvarz Training center

صورت سؤال	دوره اول		دوره دوم		دوره سوم		جمع کل	
	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد
۱- سؤالات تاپی شده، خوانا و روان می‌باشد	۲۵	۹۶/۱	۲۰	۱۰۰	۲۷	۹۳/۱	۷۲	۹۶
	۰	۰	۰	۰	۲	۶/۹	۲	۲/۷
	۱	۳/۹	۰	۰	۰	۰	۱	۱/۳
	۰	۰	۰	۰	۰	۰	۰	۰
۲- از به‌کاربردن حروف اختصاری غیر رایج خودداری شده است	۲۶	۱۰۰	۱۷	۸۵	۲۵	۸۶/۲	۶۸	۹۰/۷
	۰	۰	۲	۱۰	۴	۱۳/۸	۶	۸
	۰	۰	۱	۵	۰	۰	۱	۱/۳
	۰	۰	۰	۰	۰	۰	۰	۰
۳- تعداد موارد مورد نظر در پاسخ به سؤالات ذکر شده است (مثلاً چهار مورد ذکر شود)	۱۴	۵۳/۷	۸	۴۰	۲۲	۷۵/۹	۴۴	۵۸/۶
	۴	۱۵/۳	۴	۲۰	۴	۱۳/۸	۱۲	۱۶
	۷	۲۶/۹	۷	۳۵	۲	۶/۹	۱۶	۲۱/۳
	۱	۳/۹	۱	۵	۱	۳/۴	۳	۴/۱
۴- بارم‌بندی سؤالات مشخص شده و جلوی هر سؤال نوشته شده است	۱۱	۴۲/۳	۳	۱۵	۱۶	۵۵/۲	۳۰	۴۰
	۰	۰	۱	۵	۰	۰	۱	۱/۳
	۰	۰	۱	۵	۰	۰	۱	۱/۳
	۱۵	۵۷/۷	۱۵	۷۵	۱۳	۴۴/۹	۴۳	۵۳/۴
۵- فاصله مناسب برای پاسخ هر سؤال در نظر گرفته شده است	۹	۳۴/۶	۴	۲۰	۱۶	۵۵/۲	۲۹	۳۸/۷
	۰	۰	۰	۰	۲	۶/۹	۲	۲/۶
	۰	۰	۳	۱۵	۲	۶/۹	۵	۶/۷
	۱۷	۶۵/۴	۱۳	۶۵	۹	۳۱/۱	۳۹	۵۲
۶- صورت سؤالات با عبارات و کلمات واضح و روشن نوشته شده است (به‌جای کلمات و عبارات "بحث کنید" و "بررسی کنید" از کلمات واضح‌تری مانند "تعریف کنید"، "حل کنید"، "خلاصه کنید" و "طبقه‌بندی کنید" استفاده شود)	۱۱	۴۲/۳	۴	۲۰	۲۲	۷۵/۹	۳۷	۴۹/۳
	۹	۳۴/۶	۱۱	۵۵	۲	۶/۹	۲۲	۲۹/۳
	۴	۱۵/۴	۴	۲۰	۵	۱۷/۲	۱۳	۱۷/۳
	۲	۷/۷	۱	۵	۰	۰	۳	۴/۱
۷- سؤالات تشویقی، اضافی و انتخابی طرح نشده است.	۲۵	۹۶/۱	۲۰	۱۰۰	۲۹	۱۰۰	۷۴	۹۸/۷
	۰	۰	۰	۰	۰	۰	۰	۰
	۰	۰	۰	۰	۰	۰	۰	۰
	۱	۳/۹	۰	۰	۰	۰	۱	۱/۳
۸- تعداد سؤالات و محتوای آزمون منطبق با زمان آزمون می‌باشد.	۳	۱۱/۴	۱	۵	۶	۲۰/۷	۱۰	۱۳/۳
	۱	۳/۹	۰	۰	۱	۳/۴	۲	۲/۷
	۱	۳/۹	۱	۵	۰	۰	۲	۲/۷
	۲۱	۸۰/۸	۱۸	۹۰	۲۲	۷۵/۹	۶۱	۸۱/۳
۹- پاسخنامه نمونه یا الگو برای هر سؤال بعنوان کلید طراحی شده و هم‌زمان با ارائه سؤالات به اداره آموزش ارائه شده است (بارم هر بخش به تفکیک مشخص شده باشد).	۱	۳/۹	۱	۵	۴	۱۳/۸	۶	۷/۸
	۲	۷/۷	۰	۰	۱	۳/۴	۳	۴/۱
	۱	۳/۹	۱	۵	۱	۳/۴	۳	۴/۱
	۲۲	۸۴/۵	۱۸	۹۰	۲۳	۷۹/۴	۶۳	۸۴
۱۰- هر سؤال یک هدف اختصاصی را ارزیابی می‌کند و پاسخ‌ها ترجیحاً و حتی‌الامکان کوتاه هستند	۲	۷/۷	۲	۱۰	۱۰	۳۴/۴	۱۴	۱۸/۷
	۱۳	۵۰	۱۱	۵۵	۱۶	۵۵/۲	۴۰	۵۳/۳
	۹	۳۴/۶	۵	۲۵	۳	۱۰/۴	۱۷	۲۲/۷
	۲	۷/۷	۲	۱۰	۰	۰	۴	۵/۳
۱۱- راهنمای سؤال به‌صورت کلی یا مختص هر پرسش ذکر شده است	۱	۳/۹	۲	۱۰	۱	۳/۴	۴	۵/۳
	۱	۳/۹	۲	۱۰	۲	۶/۹	۵	۶/۷
	۲۳	۸۸/۳	۱۶	۸۰	۲۳	۷۹/۳	۶۲	۸۲/۷
	۱۲۸	۴۴/۷۳	۸۰	۳۶/۳۶	۱۸۰	۵۶/۴۴	۳۸۸	۴۷/۲۱
جمع	۳۰	۱۰۰/۴۹	۳۱	۱۴۰/۹	۳۳	۱۰۰/۳۴	۹۴	۱۱/۵۴
	۲۴	۸۰/۴۱	۲۵	۱۱۰/۳۶	۱۵	۴/۷	۶۴	۷/۷۸
	۱۰۴	۳۶/۳۵	۸۴	۳۸/۱۸	۹۱	۲۸/۵۵	۲۷۹	۳۳/۴۷

سه دوره نشان داد که توزیع داده‌ها در سه دوره آموزشی با چهار گروه منفی، ضعیف، قابل قبول و خوب با میانگین درصدهای ۲۸/۳، ۲۳/۹، ۲۲/۵ و ۲۵/۳ به ترتیب با اختلاف جزئی در یک سطح می‌باشند و اختلاف معنی‌داری در بین دوره‌ها وجود ندارد. ۵۲/۲ درصد سؤالات نقشی در تمایز دانشجویان قوی و ضعیف ندارد. ۲۲/۵ درصد سؤالات قابل قبول و ۲۵/۳ درصد سؤالات در تمایز دانشجویان خوب و کاربردی هستند.

جدول ۴ تقسیم‌بندی سؤالات در سه دوره مرکز آموزش بهورزی شهرستان سقز در سه شاخص دشوار، مطلوب و آسان نشان می‌دهد. با بررسی شاخص دشواری سؤالات آزمون نهایی نتایج نشان داد، سه دوره توزیع اعدادشان مشابه و تفاوت معنی‌داری با هم ندارد. با بررسی سه دوره، ۱۵ درصد سؤالات دشوار بود. ۵۷/۵ درصد سؤالات مطلوب و به‌طور متوسط ۲۷/۵ درصد سؤالات آسان بود. همان‌طور که در جدول ۴ ملاحظه می‌شود، بررسی شاخص تمایز سؤالات آزمون نهایی این

جدول ۴. شاخص دشواری و تمایز سؤالات آزمون نهایی سه دوره مرکز آموزش بهورزی شهرستان سقز

Table 4. Difficulty and differentiation index of the final questions of the three cohort of the Saqez Behvarz Training center

P Value	درصد	میانگین \pm انحراف معیار شاخص دشواری	تعداد	دوره آموزشی	شاخص دشواری
۰/۰۶	۱۵/۸	۱۳/۳ \pm ۹/۱	۱۹	دوره اول	کمتر از ۳۰ دشوار
	۱۳/۳	۱۰/۶ \pm ۶/۸	۱۶	دوره دوم	
	۱۵/۸	۱۳/۹ \pm ۷/۸	۱۹	دوره سوم	
۰/۴۸	۶۰	۴۵ \pm ۱۰/۴	۷۲	دوره اول	۳۰ تا ۷۰ قابل قبول (مطلوب)
	۴۸/۴	۴۴/۷ \pm ۱۱	۵۸	دوره دوم	
	۶۴/۲	۴۵/۴ \pm ۱۱/۵	۷۷	دوره سوم	
۰/۴۰	۲۴/۲	۸۶/۳ \pm ۱۰/۵	۲۹	دوره اول	بیشتر از ۷۰ ساده (آسان)
	۳۸/۳	۷۹/۶ \pm ۹/۶	۴۶	دوره دوم	
	۲۰	۸۰/۳ \pm ۹/۲	۲۴	دوره سوم	
۰/۴۵	۲۵/۸۳	-۰/۳۱ \pm ۰/۱۳	۳۱	دوره اول	کمتر از صفر منفی
	۲۵	-۰/۳۹ \pm ۰/۰۶	۳۰	دوره دوم	
	۳۴/۱۷	-۰/۳۳ \pm ۰/۱۷	۴۱	دوره سوم	
۰/۵	۲۰/۸۳	صفر	۲۵	دوره اول	۰-۰/۱ ضعیف
	۲۵	صفر	۳۰	دوره دوم	
	۲۵/۸۳	صفر	۳۰	دوره سوم	
۰/۵	۱۶/۶۷	۰/۲۰ \pm ۰/۰۱	۲۰	دوره اول	۰/۱-۰/۳ قابل قبول
	۳۲/۵	۰/۲۰ \pm ۰/۰۳	۳۹	دوره دوم	
	۱۸/۳۳	۰/۲۰ \pm ۰/۰۲	۲۲	دوره سوم	
۰/۳۴	۳۵/۸۳	۰/۵۳ \pm ۰/۱۷	۴۳	دوره اول	بیشتر از ۰/۳ خوب
	۱۷/۵	۰/۴۶ \pm ۰/۰۹	۲۱	دوره دوم	
	۲۲/۵	۰/۴۰ \pm ۰/۲۰	۲۷	دوره سوم	

درصد به ترتیب، بیشترین و کمترین امتیاز را از چک لیست گرفتند که مقایسه با مطالعه مشابهی در دانشگاه علوم پزشکی ایران در دانشجویان پزشکی، بر اساس چک لیست میلن از ۱۲۰۲ سؤال، خودداری از به‌کار بردن گزینه‌های منفی (۹۳/۸ درصد) بیشترین و چیدمان گزینه‌ها به‌صورت عمودی (۳۱/۳ درصد) کمترین امتیاز را داشتند (۱۲). در مطالعه دیگری که با هدف ارزیابی سؤالات چند گزینه‌ای دانشجویان پرستاری در دانشگاه علوم پزشکی ارومیه انجام گرفته بود، بیشترین اشکالات ساختاری مانند مطالعه ما عمودی لیست شدن گزینه‌ها بود (۱۳). در مطالعه مشابه دیگر که با هدف ارزیابی سؤالات چندگزینه‌ای

بحث

سؤالات چندگزینه‌ای در ارزشیابی دانشجویان گروه پزشکی بسیار پرطرفدار است. این سؤالات توانایی بیشتری برای آزمایش دانش و یادآوری واقعیات دارند، اما در ارزیابی مهارت‌های حل مسئله فراگیران، قدرت کمتری دارند. ساخت سؤالات چندگزینه‌ای خوب، دشوار و نیاز به تخصص دارد. اگر خوب طراحی شود مهارت‌های سطح بالای شناختی را هم می‌تواند ارزیابی کند (۱۱). نتایج ارزیابی سؤالات چندگزینه‌ای در این مطالعه نشان داد شفافیت لغات در ساقه و گزینه‌ها با ۹۸/۷۵ درصد و لیست شدن گزینه‌ها به‌صورت عمودی با ۱۶/۱۸

سؤالات مناسب بر اساس چک لیست‌های مربوطه، در سؤالات چندگزینه‌ای ۸۲/۵ درصد و تشریحی ۵۸/۳ درصد می‌باشد. این مورد مؤید این است که مهارت و دقت طراحان سؤال در سؤالات چند گزینه‌ای بیشتر از سؤالات تشریحی است. در این مرکز آموزش طی دوره‌های بررسی شده، آزمون‌های تشریحی برگزار شده محدود بود که پیشنهاد می‌شود مریبان با استفاده از سؤالات تشریحی مناسب‌تری دانشجویان را به سمت رویکرد عمقی یادگیری سوق دهند (۱۶).

در مطالعه‌ای که به تحلیل آزمون‌های علوم پایه پزشکی و دندانپزشکی دانشگاه علوم پزشکی جندی شاپور اهواز در سال ۹۹ پرداخته بود، در هر دو رشته فراوانی سؤالات با ضریب دشواری مناسب نسبتاً خوب و فراوانی سؤالات با ضریب تمیز مناسب بسیار کم بود. سؤالات علوم پایه پزشکی و علوم پایه دندانپزشکی به ترتیب دارای ۵۰/۱ و ۵۰/۰۷ درصد ضریب دشواری مطلوب و ۲۵/۸ و ۲۳/۰۷ درصد ضریب تمیز مطلوبی داشته در حالی که در مطالعه ما ۵۷/۵۳ و ۲۵/۳ درصد سؤالات ضریب دشواری و ضریب تمیز مطلوبی داشتند. در مقایسه ضریب دشواری در مطالعه ما مناسب‌تر و ضریب تمیز نامطلوب‌تر بود (۹). در مطالعه مشابه در دانشگاه علوم پزشکی ایران که با هدف تحلیل آزمون سؤالات چند گزینه‌ای دانشجویان پزشکی انجام شده بود، ۵۵/۳۹ درصد سؤالات شاخص دشواری قابل قبول و ۴۴/۰۹ درصد ضریب تمیز متوسطی داشتند که در مقایسه با مطالعه حاضر شاخص دشواری نامناسب‌تر و ضریب تمیز مناسب‌تری حاصل شده بود (۱۲).

در مطالعه‌ای که در دانشگاه علوم پزشکی ارومیه با هدف ارزیابی سؤالات چند گزینه‌ای دانشجویان پرستاری انجام شده بود، ۲۵/۹۵ و ۲۸/۷۹ درصد سؤالات به ترتیب ضریب دشواری مناسب و ضریب تمیز مناسبی داشتند که نسبت به مطالعه حاضر ضریب دشواری نامطلوب‌تر و ضریب تمیز بهتری داشتند (۱۳). در مطالعه‌ای که به تحلیل کمی و کیفی سؤالات چهار گزینه‌ای در دانشگاه علوم پزشکی شاهرود پرداخته بود، ۵۰ درصد سؤالات ضریب تمیز مناسب داشتند و ۵۱/۴ درصد سؤالات ساده بودند که در مقایسه با مطالعه حاضر ضریب تمیز مشابه و سؤالات ساده بیشتری داشتند. البته قابل ذکر است که تعریف این ضرایب با مطالعه ما کمی فرق دارد. در مطالعه شاهرود میزان زیاد سؤالات آسان را به برگزاری امتحانات به صورت مجازی نسبت داده است (۱۵). در مطالعه‌ای که با هدف تحلیل آزمون سؤالات تستی و تشریحی در دانشجویان پزشکی در آمریکای شمالی انجام شد، شاخص دشواری و تمیز چهار دوره از دانشجویان پزشکی محاسبه و مقایسه

دانشجویان دندانپزشکی دانشگاه علوم پزشکی تهران انجام شده بود، کمترین امتیاز نیز مربوط به عمودی لیست شدن گزینه‌ها و بهترین مربوط به املای درست سؤالات و گزینه‌ها، مثبت بودن ساقه سؤال و عدم استفاده از هیچ کدام در گزینه‌ها بود (۱۴). در مطالعه مشابهی که با هدف تحلیل کمی و کیفی سؤالات چهار گزینه‌ای دانشگاه علوم پزشکی شاهرود پرداخته بود، بیشترین نوع نقص به ترتیب عدم وجود بخش اعظم اطلاعات در ساقه سؤال، وجود لغات منفی در ساقه سؤال و وجود گزینه‌های همه موارد و هیچ کدام و گزینه‌های ترکیبی بود که با مطالعه ما متفاوت بود (۱۵).

قبلاً افراد معتقد بودند که سؤالات چند گزینه‌ای برای اهداف سطح پایین شناختی و یادآوری اطلاعات مناسب است ولی اگر این سؤالات خوب طراحی شود برای اهداف سطح بالاتر مانند درک، آنالیز و کاربرد هم قابل استفاده است (۱۶). در مطالعه حاضر درصد بالایی از سؤالات دارای تاکسونومی یک بودند. در مطالعه‌ای مشابه در دانشکده پزشکی دانشگاه علوم پزشکی ایران، ۵۰/۷ درصد سؤالات تاکسونومی یک و ۱۴/۵ درصد تاکسونومی دو و ۳۴/۸٪ تاکسونومی سه داشتند (۱۱). که در مقایسه با مطالعه حاضر سؤالات تاکسونومی سه بیشتر و سؤالات با تاکسونومی یک کمتری داشتند. ولی درصد سؤالات با تاکسونومی دو در مطالعه ما بیشتر بود. این امر توجه بیشتر به آموزش مریبان به‌روزی در مورد طراحی سؤال را مشخص می‌سازد. در مطالعات مشابه دیگر نیز درصد بالایی از سؤالات در سطح آگاهی و یادآوری قرار می‌گرفتند (۱۷). در مطالعه مشابه در ارومیه که به تحلیل سؤالات چند گزینه‌ای دانشجویان پرستاری پرداخته بود، اکثر سؤالات بررسی شده در تاکسونومی یک قرار داشتند (۱۳). در مطالعه‌ای که به تحلیل کمی و کیفی سؤالات چهار گزینه‌ای در دانشگاه علوم پزشکی شاهرود پرداخته بود سؤالات تاکسونومی دو بیشترین سؤال را به خود اختصاص دادند در حالی که در مطالعه حاضر سؤالات با تاکسونومی یک بیشترین سؤالات را به خود اختصاص داده بود (۱۵). ارتقای ارزیابی و نمره‌دهی به آزمون‌های تشریحی نیاز به استفاده از روش‌های مختلف و ترکیب روش‌های چندگانه و سطح بالا دارد که مریبان در ارزشیابی آموزش باید مدنظر قرار دهند. در جستجوی به‌عمل آمده کمتر به بررسی کیفیت سؤالات تشریحی از نظر ساختاری پرداخته شده بود بنابراین مقایسه با دیگر مطالعات قابل انجام نبود. در این مطالعه امتیاز سه دوره آزمون تشریحی با توزیع همسان در دوره‌ها به ترتیب ۳۸/۵ درصد خیلی خوب، ۱۹/۸ درصد خوب، ۷/۹ درصد متوسط و ۳۳/۸ درصد ضعیف ارزیابی شد. با مقایسه نتایج سؤالات تشریحی و چندگزینه‌ای مشاهده شد که

فارماکولوژی دانشجویان پزشکی انجام شده بود، ۴۷/۵ درصد سؤالات ضریب دشواری مناسب، ۴۲/۵ درصد سؤالات آسان و ۱۰ درصد سؤالات مشکل بودند که مطالعه ما سؤالات ضرایب دشواری مناسب‌تری داشتند در حالی که در این مطالعه ضریب تمایز ۲۲/۵ درصد سؤالات عالی (بیشتر از ۰/۳۵)، ۳۲/۵ درصد خوب (۰/۳۴-۰/۲۵) ۲۰ درصد مرزی (۰/۲۴-۰/۱۵) و ۲۵ درصد ضعیف (کمتر از ۰/۱۵) بودند که نسبت به مطالعه حاضر کمی بهتر بود (۲۲).

تجزیه و تحلیل سؤالات چندگزینه‌ای در مطالعات مختلف نشان می‌دهد که چندگزینه‌ای‌ها برای یادآوری اطلاعات و به خاطر سپردن طراحی شده‌اند (۲۳-۲۵). نقطه قوت مطالعه حاضر ارزیابی کیفی عمده سؤالات سه دوره اخیر مرکز آموزش بهورزی می‌باشد که تاکنون در ایران انجام نشده است که می‌تواند منجر به ارتقا سؤالات آزمون و اصلاح آنها در جهت استفاده در بانک سؤال می‌باشد. البته قابل ذکر است نتایج این مطالعه قابل تعمیم به مراکز دیگر در کشور نمی‌باشد. نقطه ضعف دیگر این مطالعه این بود که نتایج نهایی سؤالات تشریحی در دسترس نبود لذا اندازه‌گیری ضریب دشواری و ضریب تمیز این سؤالات مقدور نبود. پیشنهاد می‌شود ارزشیابی کارآموزی دوره کاردانی بهورزی این مرکز در مطالعات بعدی انجام شود.

نتیجه‌گیری

به‌طور کلی سؤالات چندگزینه‌ای توسط مربیان مناسب طراحی شده بود ولی درصد بیشتر این سؤالات برای سنجش یادآوری دانش دانشجویان بود. شاخص دشواری قابل قبول و بر اساس شاخص تمایز به نسبت نامناسب بود. بهتر است مربیان در طرح سؤالات تشریحی با مطالعه، اصول مربوطه را رعایت فرمایند تا کیفیت ارزشیابی با این روش تا حد ممکن ارتقاء یابد و ضعف طراحی آزمون‌ها برای ارزشیابی فراگیران از طریق آموزش مربیان و استانداردسازی آزمون زیر نظر متخصصان برطرف شود.

شد. درصد ضریب دشواری مطلوب در این چهار دوره بین ۴۳ تا ۶۲/۵ متغیر بود. که بین چهار دوره، تفاوت معنی‌داری در دوره اول وجود داشت در حالی که در مطالعه حاضر تفاوتی در شاخص دشواری بین سه دوره وجود نداشت. در مطالعه حاضر درصد ضریب دشواری در سه دوره بین ۴۸/۴ تا ۶۴/۲ متغیر بود. در این مطالعه حدود ۵۰ تا ۷۰ درصد سؤالات در چهار دوره ضریب تمیز مناسب (بیشتر از ۰/۲) داشتند که بین چهار دوره تفاوت معنی‌داری نداشت. ولی در مطالعه حاضر ضریب دشواری بیشتر از ۰/۳ و ۰/۱-۰/۳ به ترتیب برابر با ۲۵/۳ و ۲۲/۵ بود. با توجه که تعریف ضریب تمیز مناسب این مطالعه با مطالعه ما متفاوت بود، مقایسه ضریب تمیز در این مطالعه دقیق نمی‌باشد ولی مانند مطالعه ما تفاوت ضریب تمیز بین دوره‌ها معنی‌دار نبود (۱۹). در مطالعه‌ای که در سودان به‌منظور تحلیل سؤالات چند گزینه‌ای درس فیزیولوژی در دانشکده پزشکی خارطوم انجام شده بود، ۶۳/۳ درصد سؤالات ضریب دشواری قابل قبول، ۲۵/۵ درصد سؤالات آسان و ۱۱/۱ درصد آنها دشوار بودند در صورتی که در مطالعه حاضر این اعداد به‌ترتیب ۵۷/۵، ۲۷/۵ و ۱۵ درصد بودند و سؤالات مناسب درصد کمتری داشتند. در این مطالعه ضریب تمیز ۴۳/۵ درصد سؤالات عالی (بیشتر از ۰/۴)، ۳۰ درصد خوب (۰/۳-۰/۳۹) و ۱۶/۶ درصد قابل قبول (۰/۲۹-۰/۲) و ۹/۶ درصد ضعیف (کمتر از ۰/۲) بودند که نسبت به مطالعه ما به نسبت بهتر بودند. گرچه قابل ذکر است که تعریف ضریب تمیز در این مطالعه نیز متفاوت بود ولی با توجه به اینکه در مطالعه حاضر ۵۲/۲ درصد سؤالات ضریب تمایز کمتر از ۰/۱ داشتند، وضعیت سؤالات مطالعه حاضر از نظر ضریب تمیز مناسب نمی‌باشد (۲۰). در مطالعه‌ای که در دانشگاه بیسا عربستان به ارزیابی کیفی ۳۰۰ سؤال چند گزینه‌ای درس جراحی پرداخته بود، ۷۵/۳ درصد سؤالات شاخص دشواری مناسب داشته و ۱۸ درصد سؤالات آسان بود که نسبت به مطالعه حاضر بهتر بود. در این مطالعه ۶۳/۳ درصد ضریب تمیز خوبی داشتند که نسبت به مطالعه ما بهتر بود (۲۱) در مطالعه‌ای در هند که با هدف اندازه‌گیری روایی و پایایی سؤالات چند گزینه‌ای درس

References

- Eldakhkhny B, Elsamanoudy AZ. Discrimination Power of Short Essay Questions Versus Multiple Choice Questions as an Assessment Tool in Clinical Biochemistry. *Cureus*. 2023; 24;15(2):e35427. doi:10.7759/cureus.35427.
- Kordestani Moghaddam A, Khankeh HR, Shariati M, et al. Educational impact of assessment on medical students' learning at Tehran University of Medical Sciences: a qualitative study. *BMJ Open*. 2019; 29;9(7):e031014. doi:10.1136/bmjopen-2019-031014.
- Kipkulei J, Kangeeth S, Boibanda F, et al. Assessment Methods Used during Clinical Years of Undergraduate Medical Education at Moi University School of Medicine. *Kenya. Health*. 2022;14 (3). doi:10.4236/health.2022.143023.

4. Preston R, Gratani M, Owens K, et al. Exploring the Impact of Assessment on Medical Students' Learning. *Assessment & Evaluation in Higher Education*. 2019; 45(1), 109–124.
doi.org/10.1080/02602938.2019.1614145.
5. Mehranfard S, Pelarak F, Mashalchi H, et al. Efficacy of logbook as a clinical assessment: Using DOPS evaluation method. *Journal of Multidisciplinary Care (JMDC)*. 2022;11(4):184-189.
[doi:10.34172/jmdc.2022.62](https://doi.org/10.34172/jmdc.2022.62).
6. Lee CN, Matthew RA, Orpinas P. Design, implementation, and evaluation of community health worker training programs in Latinx communities: A scoping review. *J Community Psychol*. 2023; 51(1):382-405.
[doi:10.1002/jcop.22910](https://doi.org/10.1002/jcop.22910).
7. Gharibi F, As' habi A, Nazari M, et al. Performance challenges faced by "Behvarzes" as community health workers of Iran: a systematic review of literature. *Shiraz E-Medical Journal*. 2023; 24(1):e131792.
<https://doi.org/10.5812/semj-131792>
8. Shayan, S, Rafieian, M, Kazemi, M. Design and Develop Logbook Student's Nationwide Health Providers Training Courses of the Entire Country. *Medicine and Spiritual Cultivation*. 1397; 27(2): 124-132. [In Persian]
https://www.tebvatazkiye.ir/article_69288.html
9. Naidoo M. The pearls and pitfalls of setting high-quality multiple choice questions for clinical medicine. *S Afr Fam Pract*. 2023; 65 (1), a5726.
doi.org/10.4102/safp.v65i1.5726.
10. Khajaeali N, Aslami M, Araban M. Analysis of medical and dentistry basic sciences examinations: A case study. *Payesh*. 2020; 19 (4) :383-389. [In Persian]
<https://payeshjournal.ir/article-1-1409-en.html>
11. Yousuf ASR, Bakar SMA. Multiple Choice Questions in Medical Education: How to Construct High Quality Questions. *International Journal of Human and Health Sciences (IJHHS)*. 2020; 4(2).
<http://dx.doi.org/10.31344/ijhhs.v4i2.180>.
12. Nojomi M, Mahmoudi M. 2022. Assessment of multiple-choice questions by item analysis for medical students' examinations. *Res Dev Med Educ*. 2022; 11 (24):1-6.
[doi: 10.34172/rdme.2022.024](https://doi.org/10.34172/rdme.2022.024).
13. Baghaei R, Shams S, Feizi A, et al. Evaluation of the nursing students final exam multiple-choice question in urmia university of medical sciences. *Nursing and Midwifery Journal* 2016; 14 (4) :291-299. [In Persian]
<https://unmf.umsu.ac.ir/article-1-2645-en.html>
14. Rezai M, Farrokhnia T, Vatanpour M, et al. Evaluation of Multiple Choice Questions of Oral and Maxillofacial Medicine Courses 1, 2, and 3 in the First Semester of Academic Year 2014-2015. *J Res Dentomaxillofac Sci*. 2018;3 (2):24-30.
[doi:10.29252/jrdms.3.2.24](https://doi.org/10.29252/jrdms.3.2.24).
15. Paryab S, Zarouj Hosseini R, Zamani S, et al. Quantitative and Qualitative Evaluation of Four Choice Questions In Shahroud University of Medical Sciences During 2021-2022. *Medical Education Journal*. 2023; 12.
<https://mededj.ir/article-1-455-fa.html>.
16. Allanson P, Notar C. Writing multiple choice items that are reliable and valid. *American International Journal of Humanities and Social Science*, 2019;5 (3), pp.1-9.
<https://ijehd.cgrd.org/images/vol5no3/1.pdf>.
17. Baig M, Ali SK, Ali S, et al. Evaluation of Multiple Choice and Short Essay Question items in Basic Medical Sciences. *Pakistan Journal of Medical Sciences*. 2014; 30(1):3-6.
[doi: 10.12669/pjms.301.4458](https://doi.org/10.12669/pjms.301.4458).
18. Shakurnia A, Alijani H, Najjar S, et al. The Effect of Two Assessment Methods on Exam Preparation and Study Strategies: Multiple Choice and Essay Questions. *Iranian Journal of Medical Education*. 2013; 13 (4) :306-318. [In Persian]
<http://ijme.mui.ac.ir/article-1-2414-en.html>
19. Kumar A, George C, Campbel MH, et al. Item analysis of multiple choice and extended matching questions in the final MBBS medicine and therapeutics examination. *Journal of Medical Education*, 2022; 21 (1): e129450.
doi.org/10.5812/jme-129450.
20. Musa A, Shaheen S, Elmardi A, et al. Item difficulty & item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine Khartoum University. *Khartoum Medical Journal*. 2018.11 (2).
[doi:10.53332/kmj.v11i2.610](https://doi.org/10.53332/kmj.v11i2.610).
21. Al Ameer AY. Assessment of the Quality of Multiple-Choice Questions in the Surgery Course for an Integrated Curriculum, University of Bisha College of Medicine, Saudi Arabia. *Cureus*. 2023 Dec 13;15(12):e50441.
[doi:10.7759/cureus.50441](https://doi.org/10.7759/cureus.50441).
22. Patil RP, Bahekar SE, Kulkarni MD, et al. Evaluation of validity and reliability of multiple-choice questions in second MBBS competency-based medical education-based pharmacology examination of medical institute of India. *International Journal of Research in Medical Sciences*. 2022; 10:2878-83.
doi.org/10.18203/2320-6012.IJRMS20223091.
23. Bone E, Prosser M. Multiple Choice Questions: An Introductory Guide. The University of Melbourne. Educational resource. 2021.
doi.org/10.26188/14384831.v2.
24. Ghazy M, Ata A. MCQ Examinations for Large Number of Students: Design and Grading. *International Journal of Education*. 2020; 12 (1), pp. 118-125.
[doi:10.5296/ije.v12i1.16319](https://doi.org/10.5296/ije.v12i1.16319).
25. Pepple DJ, Young LE, Carroll RG. A comparison of student performance in multiple-choice and long essay questions in the MBBS stage I physiology examination at the University of the West Indies (Mona Campus). *Adv Physiol Educ*. 2010;34(2):86-9.
[doi: 10.1152/advan.00087.2009](https://doi.org/10.1152/advan.00087.2009).