

## تأثیر تجزیه و تحلیل سؤالات آزمون‌های پایان ترم بر کیفیت طراحی سؤال

مرضیه شبان\* فرهاد رمضانی‌بدر\*\*

### چکیده

زمینه و هدف: تجزیه و تحلیل سؤالات فرآیندی است که کیفیت و کمیت سؤالات و کل آزمون با آن ارزیابی می‌شود. این مطالعه با هدف تعیین میزان تأثیر تجزیه و تحلیل سؤالات آزمون‌های پایان ترم بر کیفیت طراحی سؤال توسط اعضای هیأت علمی دانشکده پرستاری و مامایی دانشگاه علوم پزشکی تهران انجام گرفته است. روش بررسی: بررسی حاضر یک مطالعه نیمه تجربی است. در این مطالعه سؤالات آزمون یکی از دروس تدریس شده توسط هر یک از اعضای هیأت علمی دانشکده پرستاری و مامایی دانشگاه علوم پزشکی تهران در نیمسال دوم ۸۴-۸۳ مورد تجزیه و تحلیل قرار گرفت و نتایج همراه توصیه‌های آموزشی به صورت یک جزوه به هر یک از آنان ارایه شد. سپس در نیمسال اول ۸۵-۸۴ سؤالات طرح شده توسط همان اعضای هیأت علمی تجزیه و تحلیل و با نتایج نیمسال ۸۴-۸۳ مقایسه گردید. از تعداد ۳۳ نفر عضو هیأت علمی ۲۷ نفر قبل از مداخله و ۲۱ نفر بعد از مداخله در تحقیق شرکت کردند. ابزار گردآوری داده‌ها چک لیست پنج قسمتی شامل فرم ظاهری متن سؤالات و گزینه‌ها، فرم ظاهری کل آزمون، روایی محتوا، سطح سؤال در حیطه شناختی و نکات اجرایی آزمون بود. برای تجزیه و تحلیل داده‌ها از آزمون‌های آماری  $t$ ، پی‌رسون، اسپیرمن و آزمون دقیق فیشر استفاده شد.

یافته‌ها: قبل از مداخله ۱۰۵۶ سؤال و بعد از مداخله ۸۰۳ سؤال مورد تجزیه و تحلیل قرار گرفت و نتایج مقایسه شد. بین شکل متن سؤالات و گزینه‌ها ( $p < 0/001$ )، سطح سؤالات در حیطه شناختی ( $p < 0/05$ ) و تحلیل گزینه‌های انحرافی ( $p < 0/001$ ) قبل و بعد از مداخله اختلاف معنادار آماری وجود داشت. اما بین ضریب دشواری و ضریب تمیز سؤالات، شکل آزمون، روایی محتوا و نکات اجرایی آزمون‌ها قبل و بعد از مداخله اختلاف معناداری وجود نداشت. همچنین آزمون آماری پی‌رسون ارتباط آماری معناداری بین متغیرهای سن ( $r = -0/535, p = 0/004$ ) و سابقه کار ( $r = 0/003, p = 0/546$ ) با ضریب دشواری نشان داد.

نتیجه‌گیری: مقایسه وضعیت کمی و کیفی سؤالات طراحی شده پایان ترم توسط اعضای هیأت علمی دانشکده پرستاری و مامایی در دو نیمسال مقطع کارشناسی نشان داد که ارایه نتایج تجزیه و تحلیل سؤالات به همراه جزوه‌های آموزشی به اعضای هیأت علمی در ارتقای طراحی سؤالات از جنبه‌های شکل متن سؤالات و گزینه‌ها، سطح سؤالات در حیطه شناختی و گزینه‌های انحرافی نقش مثبت داشته است.

واژه‌های کلیدی: تجزیه و تحلیل سؤالات، طراحی سؤالات، اعضای هیأت علمی، پرستاری

نویسنده مسئول:  
مرضیه شبان؛ دانشکده  
پرستاری و مامایی  
دانشگاه علوم پزشکی  
تهران

e-mail:  
Shabanma@sina.tums.  
ac.ir

- دریافت مقاله: مهر ماه ۱۳۸۵ - پذیرش مقاله: بهمن ماه ۱۳۸۵

### مقدمه

یکی از مهم‌ترین وظایفی که اعضای هیأت علمی با آن مواجه هستند ارزیابی دانشجویان است. این وظیفه مهم نیاز به مهارت‌های قابل توجهی دارد، زیرا آزمون‌های دانشگاهی به عنوان یک شاخص کلیدی و حساس‌ترین نقطه

فرآیند آموزش و یادگیری در تصمیم‌گیری برای ارتقا یا رد دانشجویان به کار می‌روند (۱). مسلماً کار تهیه آزمون با نوشتن سؤالات آن پایان نمی‌یابد بلکه پس از آن باید به دنبال گردآوری شواهدی برای اطمینان از کیفیت مناسب سؤالات بود. برای چنین هدفی طراح آزمون باید یک فرآیند دقیق نظام‌دار را برای جمع‌آوری اطلاعات صحیح به انجام رسانده و

\* مربی گروه آموزشی پرستاری داخلی جراحی دانشکده پرستاری و مامایی دانشگاه علوم پزشکی تهران  
\*\* مربی گروه آموزشی پرستاری داخلی جراحی دانشکده پرستاری و مامایی دانشگاه علوم پزشکی زنجان

که مجدداً در آزمون‌های بعدی استفاده می‌شوند. همچنین برای حذف سؤالات مبهم و دو پهلو نیز مفید بوده و در مجموع باعث افزایش مهارت مدرسان در طراحی سؤالات آزمون و تعیین حیطه‌های اختصاصی محتوای یک دوره که نیازمند تأکید و شفافیت بیشتر است، می‌گردد (۳). ضمناً مدرسان می‌توانند نقاط ضعف و قوت دانشجویان در درک مطالب و مفاهیم مربوط به درس را تعیین کنند و به نقاط ضعف و قوت تدریس پی ببرند (۶).

با توجه به تحقیقات انجام شده در دانشگاه‌های علوم پزشکی کشور ایران، ارزشیابی‌های کتبی به عمل آمده از دانشجویان که عمدتاً با استفاده از سؤالات چند گزینه‌ای انجام می‌گیرند، نقص‌های قابل ملاحظه‌ای دارند. سالمی می‌نویسد، مهارت طرح سؤال استاندارد در حال حاضر به عنوان یک مسأله قابل بحث آموزشی مطرح است (۷). تحقیقی که در دانشگاه علوم پزشکی بابل در سال ۱۳۸۲ انجام گرفت، نشان داد که طراحی آزمون‌های پایان ترم وضعیت مناسب و راضی‌کننده نداشته و عموماً آزمون‌های مورد استفاده در فرآیند ارزشیابی برآیندهای آموزشی دانشجویان، دارای اشکال ساختاری اساسی بوده و این آزمون‌ها به عنوان ابزار دقیق و معتبر عمل نمی‌نمایند و تفاوت‌های واقعی آزمودنی‌ها را نمی‌سنجند (۸). در بررسی روایی و پایایی امتحان جامع در دانشکده پرستاری و مامایی حضرت فاطمه (س) دانشگاه علوم پزشکی شیراز، از آنجا که ۸۶٪ سؤالات، مستلزم اصلاح تشخیص داده شدند، محققان به این نتیجه رسیدند که توجه بیشتر در آزمون‌سازی و اصلاح گزینه‌ها هنگام استفاده مجدد از سؤالات باید مورد تأکید قرار

بر پایه آن اطلاعات قضاوت نماید. هدف از چنین قضاوت‌هایی صرفاً شناسایی منابع احتمالی خطا و کاهش خطا در اندازه‌گیری است (۲). بدین منظور بعد از اجرای آزمون و تعیین نمرات آن، تجزیه و تحلیل سؤالات روش مناسبی برای قضاوت در مورد کیفیت آن به ویژه در آزمون‌های کوتاه پاسخ (آزمون‌های چندگزینه‌ای) است (۳). تجزیه و تحلیل سؤالات شامل فرآیند بررسی پاسخ‌های دانشجویان به سؤالات آزمون است به عبارت دیگر کیفیت و کمیت سؤالات آزمون بررسی می‌شود (۴). در تجزیه و تحلیل سؤالات اصولاً اندازه‌گیری دو جنبه کیفی و کمی آن مورد توجه قرار می‌گیرد. از نظر کیفی شکل، روش تهیه، دقت در ساختن متن سؤال، سطح سؤال در حیطه شناختی و سرانجام گزینه درست و گزینه‌های نادرست آن ارزیابی می‌شود و به گونه کلی شواهدی که در تعیین روایی محتوا و روایی سازه تست مؤثر است، مورد بررسی قرار می‌گیرد و از لحاظ کمی مواردی مثل میزان اعتبار آزمون، قدرت تشخیص هر سؤال و تأثیر آن در جدا ساختن افراد قوی از ضعیف، درجه دشواری هر سؤال و میزان مشابهت گزینه‌های نادرست هر سؤال و قدرت آن در جلب توجه آزمودنی مورد بررسی قرار می‌گیرد (۵).

تجزیه و تحلیل سؤالات آزمون مزیت‌های زیادی در ارزشیابی مؤثر دانشجویان دارد. مقاله‌ای تحت عنوان «تجزیه و تحلیل سؤالات چهار گزینه‌ای» که توسط Kehoe نوشته شده است، نشان می‌دهد، تجزیه و تحلیل سؤالات تأثیر به‌سزایی در ارتقای کیفی سؤالاتی دارد

## روش بررسی

بررسی حاضر یک مطالعه مداخله‌ای نیمه تجربی است. جمعیت مورد مطالعه ۳۳ نفر اعضای هیأت علمی دانشکده پرستاری و مامایی دانشگاه علوم پزشکی تهران می‌باشند، که در مقطع کارشناسی پرستاری تدریس می‌نمایند. نمونه‌ها با استفاده از روش سرشماری انتخاب شدند. از تعداد ۳۳ نفر، ۲۷ نفر در مرحله قبل از مداخله و ۲۱ نفر (از ۲۷ نفر) در مرحله بعد از مداخله در مطالعه شرکت کردند. با توجه به این که تجزیه و تحلیل آماری سؤال تحت تأثیر طبیعت و ویژگی‌های گروه آزمون (دانشجویان) قرار می‌گیرد، دانشجویان مورد آزمون از نظر سن، جنس، رشته تحصیلی و تعداد مشابه در نظر گرفته شده‌اند.

ابزار گردآوری داده‌ها برای تجزیه و تحلیل کیفی، چک لیست پنج قسمتی شامل شکل ظاهری متن سؤالات و گزینه‌ها، شکل ظاهری کل آزمون، روایی محتوا، سطح سؤال در حیطه شناختی و نکات اجرایی آزمون می‌باشد. قسمت اول این چک لیست شامل ۲۹ متغیر، قسمت دوم تا چهارم آن هر کدام شامل ۳ متغیر، و قسمت پنجم آن شامل ۵ متغیر می‌باشد. به منظور تعیین وضعیت شکل ظاهری متن هر سؤال و گزینه‌های آن، برای هر کدام از ۲۹ متغیر قسمت اول چک لیست، با توجه به اهمیت آن‌ها امتیازی بین ۰/۵ تا ۲ در نظر گرفته شد، که این امتیازها نشان‌دهنده میزان رعایت آن مورد یعنی مناسب می‌باشد. در صورت عدم رعایت هر کدام از موارد شکل ظاهری متن سؤالات و گزینه‌ها (قسمت اول چک لیست)، امتیاز صفر در نظر گرفته شد که به معنای نامناسب بودن آن است.

گیرد (۹). در دانشگاه علوم پزشکی همدان نیز مشخص گردید، با توجه به این که نقص ظاهری سؤالات ۳۷٪ و سنجش محفوظات ۷۷/۷٪ می‌باشد، سؤالات امتحانات دانشکده پزشکی نیاز به تحول دارد (۱۰). همچنین تحلیل سؤالات جامع آزمون نهایی پرستاری در دانشکده علوم پزشکی شاهرود نشان می‌دهد که عمده‌ترین مشکلات طراحی سؤال، طراحی سؤالات با ضریب دشواری خیلی پایین و یا خیلی بالا، و دارای گزینه‌های فاقد کارایی برای متمایز ساختن دانشجویان ضعیف از قوی می‌باشد (۱۱). Bachhoff و همکاران نیز در بررسی قدرت تمیز و سطح دشواری سؤالات آزمونی، میانگین ضریب دشواری سؤال‌ها را ۰/۳۴ و میانگین ضریب تمیز سؤال‌ها در کل آزمون را ۰/۲۸ اعلام نمودند (۱۲).

در تحقیقات ذکر شده، تجزیه و تحلیل آزمون‌ها اشکالات فراوانی را در طراحی سؤالات بسیاری از آزمون‌های دانشگاهی نشان می‌دهد. آگاهی مدرسان از این نقص‌ها شاید قدم اول در حل این مشکلات باشد. بنابراین طراحان سؤال به طور مستمر باید تجزیه و تحلیل سؤالات را به عنوان معیار کیفیت قرار داده و سؤالات غیر قابل قبول را حذف نمایند و سؤالات بهتر با ملاک‌های مطلوب را جایگزین نموده و نگهداری کنند (۲). مطالعه حاضر با هدف تعیین میزان تأثیر تجزیه و تحلیل سؤالات چهار گزینه‌ای آزمون‌های پایان ترم بر نحوه طراحی سؤال توسط اعضای هیأت علمی دانشکده پرستاری و مامایی دانشگاه علوم پزشکی تهران انجام گرفته است. پژوهشگران طرح امیدوارند از این طریق در جهت ارتقای روند ارزشیابی دانشجویان گام مؤثری برداشته شود.

ابزار با استفاده از رابطه اسپیرمن براون (۰/۸۸) محاسبه گردید.

به منظور جمع‌آوری اطلاعات، سؤالات پایان ترم طرح شده به همراه پاسخ‌های دانشجویان و طرح درس‌های مربوطه از اعضای هیأت علمی دریافت شد و مورد تجزیه و تحلیل کمی و کیفی قرار گرفت. سؤالاتی که مورد تجزیه و تحلیل قرار گرفتند سؤالات آزمون یکی از دروس تدریس شده توسط هر عضو هیأت علمی بود. به منظور انجام مداخله، پس از تجزیه و تحلیل سؤالات پایان ترم نیمسال دوم ۸۴-۸۳، نتایج به اطلاع اعضای هیأت علمی رسانده شد و به همراه آن توصیه‌های آموزشی به صورت یک جزوه در اختیار هر فرد قرار گرفت و در مورد تغییر و اصلاح سؤالات با توجه به نتایج حاصل از تجزیه و تحلیل، به سؤالات آنان پاسخ داده شد. سپس در پایان نیمسال اول ۸۵-۸۴ سؤالات طرح شده توسط اعضا هیأت علمی ذکر شده مورد تجزیه و تحلیل قرار گرفت و نتایج آن با نتایج نیمسال دوم ۸۴-۸۳ مقایسه گردید.

اطلاعات جمع‌آوری شده با استفاده از آمار توصیفی نظیر میانگین، جداول توزیع فراوانی، انحراف معیار، و آمار استنباطی نظیر آزمون‌های آماری  $t$ ، ضریب همبستگی پیرسون و اسپیرمن، و آزمون دقیق فیشر، با نرم‌افزار SPSS مورد تجزیه و تحلیل قرار گرفت و  $P$  کمتر از ۰/۰۵ معنادار تلقی شد.

### یافته‌ها

قبل از مداخله ۱۰۵۶ سؤال مورد تجزیه و تحلیل قرار گرفت، که از این تعداد بیشترین

قسمت دوم چک لیست (شکل ظاهری کل آزمون) نیز به همین صورت امتیازبندی گردید. در مورد قسمت‌های سوم تا پنجم چک لیست یعنی روایی محتوا، سطح سؤال در حیطه شناختی و نکات اجرایی آزمون سه وضعیت مناسب با امتیاز ۲، نسبتاً مناسب با امتیاز ۱، و نامناسب با امتیاز صفر در نظر گرفته شد.

برای تجزیه و تحلیل کمی از فرمول‌های ضریب دشواری و ضریب تمیز سؤالات چندگزینه‌ای استفاده گردید، و برای تحلیل گزینه‌های انحرافی، درصد انتخاب آن‌ها محاسبه شد. برای تفسیر ضریب دشواری، میزان صفر تا ۰/۳۹، نامناسب و ۰/۴ تا ۰/۸، مناسب در نظر گرفته شد. برای تفسیر ضریب تمیز میزان کمتر از ۰/۱۹، ضعیف، ۰/۲ تا ۰/۲۹ متوسط، ۰/۳ تا ۰/۳۹ خوب و ۰/۴ تا ۱ خیلی خوب ارزیابی گردید. برای تحلیل گزینه‌های انحرافی، با محاسبه درصد انتخاب آن‌ها مقادیر مناسب، نسبتاً مناسب، و نامناسب در نظر گرفته شد. به علاوه پرسشنامه‌ای شامل برخی اطلاعات فردی اعضای هیأت علمی دانشکده پرستاری و مامایی دانشگاه علوم پزشکی تهران شامل سن، سابقه کار به عنوان عضو هیأت علمی، سابقه شرکت در کارگاه‌های تجزیه و تحلیل آزمون‌ها، و آشنایی با روش‌های تجزیه و تحلیل سؤالات چندگزینه‌ای تهیه شد. برای تعیین روایی چک لیست مربوط به تجزیه و تحلیل کیفی سؤالات از روش روایی محتوا استفاده شد. بدین منظور از اعضای هیأت علمی دانشگاه علوم پزشکی تهران و ایران نظر خواهی گردید. همچنین برای تعیین پایایی این ابزار از روش ثبات درونی استفاده شد. برای این کار ابتدا ضریب همبستگی دو نیمه پرسشنامه محاسبه شد و سپس ضریب پایایی کل

نظر شکل ظاهری متن سؤالات و گزینه‌ها مناسب بودند. بعد از مداخله نیز نتایج نشان داد که بیشتر سؤالات مورد بررسی (به ترتیب ۴۷/۷٪ و ۳۱٪) در مورد «نوشتن جداگانه پرسش‌ها و گزینه‌ها در یک خط افقی» و «مشخص نمودن کلمه منفی در ریشه سؤال» نامناسب، و در بقیه موارد مناسب هستند. در مجموع اکثر سؤالات مورد بررسی (۹۹/۶٪) از نظر شکل ظاهری متن سؤالات و گزینه‌ها مناسب بودند.

در مرحله قبل از مداخله ۸۵/۲٪ آزمون‌های مورد بررسی از نظر شکل ظاهری کل آزمون مناسب بودند. این در حالی است که متن ۸۱/۵٪ آزمون‌ها دست‌نویس بوده و تایپ نشده بودند. بعد از مداخله، ۹۰/۴٪ آزمون‌های مورد بررسی از نظر شکل ظاهری کل آزمون مناسب بودند، و فقط ۶۶/۷٪ آزمون‌ها دست‌نویس بوده و تایپ نشده بودند. در مورد روایی محتوا نتایج نشان داد که قبل از مداخله بیشتر سؤالات آزمون‌های مورد مطالعه (۶۶/۷٪) بر اساس اهداف آموزشی درس مربوطه طرح شده‌اند. در اکثر آزمون‌ها (۹۲/۶٪) تعداد سؤالات طرح شده بر اساس زمان تدریس آن‌ها بود. در ۵۹/۳٪ آزمون‌ها نیز سؤالات طرح شده بر اساس محتوای برنامه درسی بود. در مجموع فقط ۵۵/۶٪ آزمون‌ها از نظر روایی محتوا مناسب بودند. اما بعد از مداخله فقط ۵۲/۴٪ سؤالات آزمون‌ها بر اساس اهداف آموزشی درس مربوطه طرح شده بودند. همچنین در اکثر آزمون‌ها (۹۵/۲٪) طراحی سؤالات بر اساس زمان تدریس، نسبتاً مناسب بود و در ۵۲/۴٪ آزمون‌ها، سؤالات طرح شده بر اساس محتوای برنامه درسی بود. در مجموع یافته‌ها نشان داد که روایی محتوای ۴۷/۶٪

سؤالات مورد بررسی (۶/۸٪) مربوط به درس پرستاری اطفال (۱) و کمترین سؤالات مورد بررسی (۱/۸٪) مربوط به درس پرستاری بیماری‌های چشم بود. بعد از مداخله ۸۰۳ سؤال مورد تجزیه و تحلیل قرار گرفت، که از این تعداد بیشترین سؤالات مورد بررسی (۹٪) مربوط به درس پرستاری اطفال (۱) و (۲)، و کمترین سؤالات مورد بررسی (۱/۵٪) مربوط به درس پرستاری سوختگی بود. قابل ذکر است که نتایج تجزیه و تحلیل سؤالات قبل و بعد از مداخله با یکدیگر مورد مقایسه قرار گرفتند.

نتایج پژوهش نشان داد که بیشتر اعضای هیأت علمی (۳۷٪) در گروه سنی ۳۶-۴۰ قرار دارند. بیشتر آن‌ها (۴۰/۸٪) دارای سابقه کار ۱۱-۱۵ سال هستند. ۶۶/۷٪ آن‌ها در هیچ‌کدام از کارگاه‌های تجزیه و تحلیل آزمون شرکت نکرده‌اند. اکثر نمونه‌های پژوهش (۸۵/۲٪) در مورد تجزیه و تحلیل سؤالات چند گزینه‌ای آگاهی نسبی داشتند. بیشتر اعضای هیأت علمی (۷۴/۱٪) در مورد سطح سؤالات در حیطه شناختی آگاهی کامل داشتند. در حالی که در مورد خطای استاندارد آزمون فقط ۱۸/۵٪ از نمونه‌های پژوهش آگاهی کامل داشتند. همچنین ۷۴/۱٪ از آن‌ها در مورد تجزیه و تحلیل کیفی سؤالات آگاهی نداشتند.

بررسی وضعیت کیفی سؤالات قبل از مداخله نشان داد که بیشتر سؤالات (به ترتیب ۶۲/۴٪ و ۵۴/۸٪) در مورد «نوشتن جداگانه پرسش‌ها و گزینه‌ها در یک خط افقی» و «مشخص نمودن کلمه منفی در ریشه سؤال» نامناسب، و در بقیه موارد مناسب هستند. به طور کلی اکثر سؤالات مورد بررسی (۹۷/۹٪) از

مقایسه وضعیت کیفی سؤالات نشان داد که بین شکل ظاهری متن سؤالات و گزینه‌ها ( $p < 0/001$ ) و سطح سؤالات در حیطة شناختی ( $p < 0/05$ ) قبل و بعد از مداخله اختلاف معنادار آماری وجود دارد، اما بین شکل ظاهری کل آزمون، روایی محتوا و نکات اجرایی آزمون‌ها قبل و بعد از مداخله اختلاف معناداری وجود نداشت. همچنین اختلاف معناداری بین ضریب دشواری سؤالات و ضریب تمیز سؤالات قبل و بعد از مداخله وجود نداشت (جدول شماره ۱).

در مورد ارتباط بین برخی متغیرهای فردی با نتایج تجزیه و تحلیل سؤالات قبل از مداخله، آزمون آماری پیرسون ارتباط آماری معناداری بین متغیرهای سن ( $p = 0/004$ ) و سابقه کار ( $p = 0/003$ ) با ضریب دشواری نشان داد. اما بین متغیرهای شرکت در کارگاه و میزان آگاهی از تجزیه و تحلیل سؤالات و متغیر «ضریب دشواری سؤالات» ارتباط معناداری وجود نداشت. همچنین ارتباط آماری معناداری بین متغیرهای سن، سابقه کار، شرکت در کارگاه و میزان آگاهی از تجزیه و تحلیل سؤالات با شکل ظاهری متن سؤالات و گزینه‌ها، شکل ظاهری آزمون، نکات اجرایی، سطح سؤالات در حیطة شناختی، ضریب تمیز سؤالات، و گزینه‌های انحرافی وجود نداشت (جدول شماره ۲).

آزمون‌ها مناسب بود و به همین میزان یعنی ۴۷/۶٪ آزمون‌ها روایی محتوای نامناسب داشتند. ۴/۸٪ نیز نسبتاً مناسب بودند.

در مورد سطح سؤالات در حیطة شناختی (طبق طبقه‌بندی بلوم) نتایج نشان داد که قبل از مداخله بیشتر سؤالات آزمون‌های طرح شده (۵۹/۳٪) نامناسب بودند. به طوری که ۷۲/۶٪ از کل سؤالات در سطح دانش طرح شده بودند. اما بعد از مداخله سطح ۵۲/۴٪ سؤالات آزمون‌های طرح شده در حیطة شناختی، نسبتاً مناسب بود و فقط ۳۳/۳٪ سؤالات نامناسب بودند. به عبارت دیگر ۴۸/۶٪ از کل سؤالات در سطح دانش طرح شده بودند. در مورد آخرین مقوله بررسی کیفی سؤالات، یعنی نکات اجرایی آزمون‌ها، ۶۲/۹٪ آن‌ها قبل از مداخله مناسب بودند، در حالی که بعد از مداخله ۷۱/۴٪ آزمون‌ها نکات اجرایی مناسبی داشتند.

نتایج تجزیه و تحلیل کمی سؤالات نشان داد که قبل از مداخله ۵۶/۵٪ سؤالات ضریب دشواری مناسب داشتند و بعد از مداخله ضریب دشواری ۵۹/۳٪ سؤالات مناسب (۰/۸-۰/۴) ارزیابی شد. نتایج مطالعه در مورد ضریب تمیز سؤالات قبل و بعد از مداخله نشان داد که به ترتیب ۴۴/۲٪ و ۴۰/۶٪ سؤالات ضعیف (۰/۱۹ <) بودند. همچنین نتایج نشان داد که ۳۴/۷٪ گزینه‌های انحرافی قبل از مداخله مناسب بودند. در حالی که ۳۶/۲٪ گزینه‌های انحرافی بعد از مداخله مناسب بودند.

**جدول ۱:** مقایسه وضعیت کیفی و کمی سؤالات آزمون‌های پایان ترم قبل و بعد از مداخله در دانشکده پرستاری و مامایی دانشگاه علوم پزشکی تهران در سال ۱۳۸۴

نتایج آزمون	بعد از مداخله		قبل از مداخله		وضعیت کیفی و کمی سؤالات	وضعیت کیفی
	انحراف معیار	میانگین	انحراف معیار	میانگین		
Paired $t$ test= $۷/۴۸$ df= $۸۰۲$ $p < ۰/۰۰۱$	۲/۲۹	۹۸	۲/۳۱	۹۷/۷	شکل ظاهری متن سؤالات و گزینه‌ها	وضعیت کیفی
Paired $t$ test= $۰/۸۴$ df= $۲۰$ $p = ۰/۴۱۱$	۱۶/۱۸	۸۰/۹	۱۹/۷۶	۷۵/۶	شکل ظاهری کل آزمون	
Paired $t$ test= $۰/۸۵$ df= $۲۰$ $p = ۰/۴۰۷$	۱۶/۶۵	۶۶/۶۵	۱۶/۲۳	۶۹/۸۲	روایی محتوا	
Paired $t$ test= $۲/۶۵$ df= $۲۰$ $p < ۰/۰۵$	۱۲/۷۹	۶۲/۶۶	۱۵/۱۶	۵۲/۳۵	سطح سؤالات در حیطه شناختی	
Paired $t$ test= $۰/۱۸$ df= $۲۰$ $p = ۰/۸۶۱$	۱۳/۳۶	۸۲/۷۳	۱۳/۵	۸۱/۹	نکات اجرایی آزمون‌ها	
Paired $t$ test= $۰/۵۲$ df= $۸۰۰$ $p = ۰/۶۰۵$	۰/۲۲	۰/۷۱	۰/۲۳	۰/۷۱	ضریب دشواری سؤالات	وضعیت کمی
Paired $t$ test= $۱/۸۷$ df= $۸۰۰$ $p = ۰/۰۶۲$	۰/۲۱	۰/۲۳	۰/۲	۰/۲۲	ضریب تمیز سؤالات	

**جدول ۲:** ارتباط بین وضعیت کیفی و کمی سؤالات در مرحله قبل از مداخله با مشخصات فردی اعضای هیأت علمی دانشکده پرستاری و مامایی دانشگاه علوم پزشکی تهران در سال ۱۳۸۴

مشخصات فردی و حرفه‌ای نوع آزمون	سن	سابقه کار	سابقه شرکت در کارگاه تجزیه و تحلیل سؤالات	آگاهی قبلی در مورد تجزیه و تحلیل سؤالات	وضعیت کیفی و کمی سؤالات
پیرسون	پیرسون	پیرسون	آزمون دقیق فیشر	اسپیرمن	شکل ظاهری متن سؤالات و گزینه‌ها
۰/۰۸	۰/۰۸	۰/۰۶	۰/۵۶	۰/۰۶	شکل ظاهری کل آزمون
۰/۱۲	۰/۰۷	۰/۶۶	۰/۶۶	۰/۲۷	روایی محتوا
۰/۱۵	۰/۰۵	۰/۷۵	۰/۷۵	۰/۰۷	سطح سؤالات در حیطه شناختی
۰/۰۸	۰/۱۵	۰/۵۵	۰/۵۵	۰/۰۷	نکات اجرایی آزمون‌ها
*-۰/۵۴	*-۰/۵۵	**۰/۵۵	۰/۴۵	۰/۰۲	ضریب دشواری سؤالات
۰/۱۷	۰/۲۱	۰/۲۲	۰/۲۲	۰/۱۵	ضریب تمیز سؤالات
۰/۱۶	۰/۰۲	۰/۴۲	۰/۴۲	۰/۲۲	تحلیل گزینه‌های انحرافی

\*  $p=0.004$ , \*\*  $p=0.003$

## بحث

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.

مقایسه وضعیت کیفی و کمی سؤالات پایان ترم نیمسال دوم ۸۴-۸۳ و سؤالات پایان ترم نیمسال اول ۸۵-۸۴ نشان داد که تجزیه و تحلیل سؤالات، بازخورد به اعضای هیأت علمی، و ارایه جزوات آموزشی به آن‌ها بر روی طراحی سؤالات از نظر شکل متن سؤالات و گزینه‌ها، سطح سؤالات در حیطه شناختی و طراحی گزینه‌های انحرافی مؤثر بوده است. این مطلب بیانگر آن است که بازخورد نتایج تجزیه و تحلیل سؤالات به اعضای هیأت علمی می‌تواند باعث طرح سؤالات بهتر شود. فتاحی بافقی و کریمی در تحقیقی تحت عنوان «تأثیر آموزش بر اعضای هیأت علمی در طرح سؤالات چند گزینه‌ای از نظر اعتبار صوری و محتوایی» به نتایج مشابهی دست یافتند. آن‌ها بیان نمودند که آموزش کارگاهی در زمینه‌های آموزش طراحی سؤال تأثیرات مطلوبی بر عملکرد اعضای هیأت علمی داشته است. به عنوان مثال قبل از آموزش ۲۱٪ سؤالات به صورت منفی طراحی شده بود، در صورتی که بعد از آموزش این میزان به ۹٪ کاهش یافت (۱۳). همچنین غلامی و دادگستر نیا در تحقیقی با عنوان «دیدگاه اساتید بالینی درباره تأثیر ارایه بازخورد به اساتید در مورد طراحی سؤال و تهیه بانک سؤالات بر کیفیت آزمون‌ها» در دانشگاه علوم پزشکی اصفهان بیان نمودند که

همه اساتید مایل به دریافت جزوه آموزشی درباره اصول طراحی سؤال بودند با این که ۷۷٪ آنان قبلاً در کارگاهی با این عنوان شرکت کرده بودند. همچنین ۶۶٪ معتقد بودند که نظارت کارشناسان خطاهای احتمالی را کاهش خواهد داد (۱۴).

در مورد سایر موارد تجزیه و تحلیل سؤالات نظیر شکل آزمون، نکات اجرایی، روایی محتوا، ضریب دشواری و ضریب تمیز سؤالات اختلاف معنادار آماری وجود نداشت اما ارزیابی انجام شده تفاوت‌هایی را بین قبل و بعد از مداخله نشان می‌دهد. به عبارت دیگر بعد از مداخله نتایج ارزیابی متغیرهای ذکر شده به جز روایی محتوا بهتر شده بود. به نظر پژوهشگران احتمالاً با توجه به این که اکثر اعضای هیأت علمی در مورد تجزیه و تحلیل کیفی سؤالات آگاهی نداشتند و ۶۶٪ آن‌ها در هیچ‌کدام از کارگاه‌های تجزیه و تحلیل آزمون شرکت نکرده بودند. لذا، صرف اطلاع‌رسانی در مورد نتایج تجزیه و تحلیل سؤالات، ارایه جزوات آموزشی و پاسخ دادن به برخی اشکالات آن‌ها در مورد سؤالات، نمی‌توانسته باعث رفع برخی مشکلات آن‌ها در مورد طراحی سؤال و آزمون شود. در تحقیقی که توسط سالمی و رفیعی تحت عنوان «بررسی تأثیر کارگاه‌های آموزشی روش تدریس در میزان به کارگیری مهارت‌های طرح سؤال در بین اساتید» در دانشگاه علوم پزشکی اراک انجام گردید، نتایج مشابهی حاصل شد. نتایج پژوهش فوق نشان داد، اساتیدی که این کارگاه را گذرانده‌اند از لحاظ ضریب دشواری و ضریب تمیز سؤالاتی که طرح کرده‌اند، با اساتیدی که دوره را نگذرانده‌اند، هیچ تفاوتی



ندارند (۷). به نظر پژوهشگران مهارت طرح سؤال استاندارد نیاز به یک برنامه‌ریزی اساسی از سوی مدیران و معاونین آموزشی دارد. شاید برنامه‌ریزی برای ایجاد انگیزه در اساتید، و افزایش آگاهی و بهبود عملکرد آن‌ها برای طرح سؤالات استاندارد باعث ارتقای آزمون‌ها شود.

با توجه به نتایج، روایی محتوای برخی آزمون‌ها قبل و بعد از مداخله نامناسب بودند. در این مورد برخی سؤالات بر اساس اهداف آموزشی درس مربوطه و محتوای برنامه درسی طرح نشده بودند. برای رفع این مشکل احتمالاً استفاده از جدول مشخصات آزمون یا جدول بودجه‌بندی هنگام طراحی سؤالات می‌تواند مفید باشد.

در مورد سطح سؤالات در حیطه شناختی قبل و بعد از بازخورد اختلاف آماری معناداری مشاهده شد. به عبارت دیگر بعد از مداخله سؤالات در سطوح بالاتر حیطه شناختی طرح شده بودند. اما در مجموع اکثر سؤالات در سطح دانش طرح شده بودند. عرب در پژوهش خود نتایج مشابهی را ذکر می‌کند. وی بیان نمود که ۷۷/۷٪ سؤالات مورد بررسی در سطح دانش طرح شده‌اند. در حقیقت این تعداد از سؤالات فقط محفوظات فراگیران را سنجیده بود (۱۰). پروانه و همکاران در پژوهشی تحت عنوان «ارزیابی سؤالات امتحانات فیزیوپاتولوژی و بالینی رشته پزشکی دانشگاه علوم پزشکی بابل» بیان نمودند که از مجموع ۷۷۰ سؤال، ۵۰/۶٪ آن‌ها در سطح دانش طرح شده بودند. با توجه به یافته‌های مطرح شده، بیش از نیمی از سؤالات طرح شده، دانش

فراگیران را ارزشیابی می‌کنند. از آنجا که در رشته‌های پرستاری و پزشکی، تجزیه و تحلیل، قضاوت و کاربرد دانسته‌ها باعث افزایش مهارت دانشجویان و افزایش دستیابی آن‌ها به صلاحیت‌های بالینی حرفه‌ای می‌گردد، توصیه می‌شود سؤالات، سطوح بالاتر حیطه شناختی را مورد ارزیابی قرار دهند. بنابراین برنامه‌ریزی جهت افزایش مهارت اعضای هیأت علمی برای طراحی سؤالات در سطوح بالاتر تاکسونومی بلوم می‌تواند کمک‌کننده باشد (۱۵).

بر اساس یافته‌ها پیشنهاد می‌شود که علاوه بر تجزیه و تحلیل مرتب سؤالات پایان ترم، برای استفاده از نتایج آن برنامه‌ریزی لازم صورت گیرد. ایجاد بانک سؤال برای اساتید و یا اصلاح آن، افزایش آگاهی آنان به طور مستمر و مناسب و به‌کارگیری کارشناسان جهت ارزیابی سؤالات و ایجاد انگیزه برای اعضای هیأت علمی می‌تواند باعث شود تا قضاوتی مناسب از یادگیری دانشجویان به عمل آید و در جهت بهبود روش تدریس و فرآیند یادگیری یاددهی اقدام مناسب صورت گیرد.

### نتیجه‌گیری

بررسی ضریب دشواری سؤالات قبل و بعد از مداخله بیانگر آن بود که بسیاری از سؤالات طرح شده خیلی آسان و یا خیلی دشوار بودند. همچنین وجود ارتباط آماری معنادار بین متغیرهای سن و سابقه کار با ضریب دشواری نشان می‌دهد که افراد با سن و سابقه کار کمتر تمایل دارند سؤالات با ضریب دشواری بالاتر (آسان‌تر) و افراد با سن و سابقه کار بیشتر تمایل دارند سؤالات

در مورد سطح سؤالات در حیطه شناختی قبل و بعد از بازخورد اختلاف آماری معناداری مشاهده شد. به عبارت دیگر بعد از مداخله سؤالات در سطوح بالاتر حیطه شناختی طرح شده بودند. اما در مجموع اکثر سؤالات در سطح دانش طرح شده بودند. عرب در پژوهش خود نتایج مشابهی را ذکر می‌کند. وی بیان نمود که ۷۷/۷٪ سؤالات مورد بررسی در سطح دانش طرح شده‌اند. در حقیقت این تعداد از سؤالات فقط محفوظات فراگیران را سنجیده بود (۱۰). پروانه و همکاران در پژوهشی تحت عنوان «ارزیابی سؤالات امتحانات فیزیوپاتولوژی و بالینی رشته پزشکی دانشگاه علوم پزشکی بابل» بیان نمودند که از مجموع ۷۷۰ سؤال، ۵۰/۶٪ آن‌ها در سطح دانش طرح شده بودند. با توجه به یافته‌های مطرح شده، بیش از نیمی از سؤالات طرح شده، دانش

فراگیران را ارزشیابی می‌کنند. از آنجا که در رشته‌های پرستاری و پزشکی، تجزیه و تحلیل، قضاوت و کاربرد دانسته‌ها باعث افزایش مهارت دانشجویان و افزایش دستیابی آن‌ها به صلاحیت‌های بالینی حرفه‌ای می‌گردد، توصیه می‌شود سؤالات، سطوح بالاتر حیطه شناختی را مورد ارزیابی قرار دهند. بنابراین برنامه‌ریزی جهت افزایش مهارت اعضای هیأت علمی برای طراحی سؤالات در سطوح بالاتر تاکسونومی بلوم می‌تواند کمک‌کننده باشد (۱۵).

بر اساس یافته‌ها پیشنهاد می‌شود که علاوه بر تجزیه و تحلیل مرتب سؤالات پایان ترم، برای استفاده از نتایج آن برنامه‌ریزی لازم صورت گیرد. ایجاد بانک سؤال برای اساتید و یا اصلاح آن، افزایش آگاهی آنان به طور مستمر و مناسب و به‌کارگیری کارشناسان جهت ارزیابی سؤالات و ایجاد انگیزه برای اعضای هیأت علمی می‌تواند باعث شود تا قضاوتی مناسب از یادگیری دانشجویان به عمل آید و در جهت بهبود روش تدریس و فرآیند یادگیری یاددهی اقدام مناسب صورت گیرد.

انحرافی مؤثر بوده است ولی در مورد سایر موارد تجزیه و تحلیل سؤالات به لحاظ آماری اختلاف معناداری وجود نداشته است. اما با این حال ارزیابی انجام شده تفاوت‌هایی را بین قبل و بعد از مداخله نشان می‌دهد. به عبارت دیگر نتایج ارزیابی متغیرهای ذکر شده به جز روایی محتوا بهتر شده است. لازم به ذکر است که اکثر اعضای هیأت علمی در مورد تجزیه و تحلیل کیفی سؤالات آگاهی نداشتند و اکثریت آن‌ها در هیچ کارگاهی در این زمینه شرکت نکرده بودند.

### تشکر و قدردانی

این مقاله نتیجه طرح تحقیقاتی مصوب دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران به شماره قرارداد ۱۳۲/۲۸۳۲ مورخ ۸۴/۴/۱۵ می‌باشد، بدین وسیله از حمایت مربوطه تشکر می‌نماییم. همچنین از همکاری مستمر اعضای هیأت علمی دانشکده پرستاری و مامایی به عنوان نمونه‌های پژوهش نیز سپاسگزاریم.

با ضریب دشواری پایین‌تر (دشواری) طرح نمایند. که به هر صورت تمایل به طراحی این نوع سؤالات می‌تواند باعث کاهش ضریب تمیز سؤالات گردد.

ضریب تمیز درصد تعدادی از سؤالات ضعیف بودند. این در حالی است که درصد کمی از گزینه‌های انحرافی قبل و بعد از مداخله مناسب بودند.

همچنین بیشتر سؤالات در سطح دانش حیطه شناختی طرح شده بودند، بین سن و سابقه کار و ضریب دشواری سؤالات ارتباط معنادار و منفی وجود داشت.

مقایسه وضعیت کیفی و کمی سؤالات طراحی شده پایان ترم توسط اعضای هیأت علمی دانشکده پرستاری و مامایی در دو نیمسال در مقطع کارشناسی نشان داد که تجزیه و تحلیل سؤالات، بازخورد به اعضای هیأت علمی و ارایه جزوات آموزشی به آن‌ها روی طراحی سؤالات از نظر شکل متن سؤالات و گزینه‌ها، سطح سؤالات در حیطه شناختی و طراحی گزینه‌های

### References

- 1 - Zurawski RM. Making the most of exams: procedures for item analysis. The National Teaching & Learning Forum. 1998; 7(6).
- ۲ - ذوالفقاری بهزاد، ادیبی ندا، درخشان‌فر ثریا و همکاران. آزمون‌های پیشرفت تحصیلی در علوم پزشکی، اصفهان: انتشارات مدیریت مطالعات و توسعه آموزش پزشکی دانشگاه علوم پزشکی اصفهان، ۱۳۷۹.
- 3 - Kehoe J. Basic item analysis for multiple-choice tests. Practical Assessment, Research & Evaluation. Available at: <http://pareonline.net/getvn.asp?v=4&n=10>. 2007; 4(10).
- 4 - Matlock-Hetzel S. Basic concepts in Item and Test Analysis. Texas A & M University . URL: <http://ericae.net/ft/tamu/Espy.htm>; 1997.
- ۵ - هومن حیدرعلی. اندازه‌گیری‌های روانی و تربیتی (فن تهیه تست و پرسشنامه)، تهران: انتشارات پارسا، ۱۳۸۱.
- 6 - Item Analysis. Office of educational assessment university of washington. URL: [http://www.washington.edu/oea/pdfs/resources/item\\_analysis.pdf](http://www.washington.edu/oea/pdfs/resources/item_analysis.pdf); 2001.
- ۷ - سالمی لادن، رفیعی مریم. بررسی تأثیر کارگاه‌های آموزشی روش تدریس در میزان به کارگیری مهارت‌های طرح سؤال در بین اساتید، ویژه‌نامه خلاصه مقالات هفتمین همایش کشوری آموزش پزشکی، تبریز: انتشارات معاونت پژوهشی دانشگاه علوم پزشکی تبریز، ۱۳۸۳: ۵۴-۵۵.

- ۸- اشرفپور منوچهر، بهشتی زهرا، ملوکزاده سرور. بررسی کیفیت آزمون‌های پایان ترم فراگیران دانشگاه علوم پزشکی بابل در نیمسال دوم تحصیلی ۱۳۷۸-۷۹. مجله دانشگاه علوم پزشکی بابل، شماره ۲، ۱۳۸۲؛ ۴۷-۴۲.
- ۹- معطری مرضیه، هاشمی فاطمه، پارسایکتا زهره. بررسی روایی و پایایی امتحان جامع در دانشکده پرستاری و مامایی حضرت فاطمه (س) دانشگاه علوم پزشکی شیراز، ویژه‌نامه چهارمین همایش کشوری آموزش پزشکی، تهران: انتشارات معاونت پژوهشی دانشگاه علوم پزشکی تهران، ۱۳۷۸؛ ۶۱.
- ۱۰- عرب ملیحه. بررسی نقایص سؤالات چهار گزینه‌ای امتحانات دانشکده پزشکی دانشگاه علوم پزشکی همدان در نیمه دوم سال تحصیلی، ویژه‌نامه چهارمین همایش کشوری آموزش پزشکی، تهران: انتشارات معاونت پژوهشی دانشگاه علوم پزشکی تهران، ۱۳۷۸؛ ۵۱.
- ۱۱- دادگری علی، ابراهیمی حسین، حسنی محمد رضا و همکاران. تحلیل سؤالات جامع آزمون نهایی پرستاری در دانشکده علوم پزشکی شاهرود. مجله دانشگاه علوم پزشکی شهرکرد، سال ۱۳۸۱؛ ۴(۴): صفحات ۱۰-۴.
- 12 - Bachhoff E, Larrazolo N, Rosas M. The level of Difficulty and Discrimination Power of the Basic knowledge and skills Examination (EXHCOBA). Electronic Journal of Educational Research. 2000; 2(1).
- ۱۳- فتاحی بافقی علی، کریمی حسین. تأثیر آموزش بر اعضای هیأت علمی در طرح سؤالات چند گزینه‌ای از نظر اعتبار صوری و محتوایی، ویژه‌نامه خلاصه مقالات هفتمین همایش کشوری آموزش پزشکی، تبریز: انتشارات معاونت پژوهشی دانشگاه علوم پزشکی تبریز، ۱۳۸۳؛ ۶۴-۶۳.
- ۱۴- غلامی وجیهه، دادگستر نیا محمد. دیدگاه اساتید بالینی درباره تأثیر رایج فیدبک به اساتید در مورد طراحی سؤال و تهیه بانک سؤالات بر کیفیت آزمون‌ها، ویژه‌نامه خلاصه مقالات هفتمین همایش کشوری آموزش پزشکی، تبریز: انتشارات معاونت پژوهشی دانشگاه علوم پزشکی تبریز، ۱۳۸۳؛ ۵۲.
- ۱۵- پروانه منیره و همکاران. ارزیابی سؤالات امتحانات فیزیوپاتولوژی و بالینی رشته پزشکی دانشگاه علوم پزشکی بابل، ویژه‌نامه خلاصه مقالات هفتمین همایش کشوری آموزش پزشکی، تبریز: انتشارات معاونت پژوهشی دانشگاه علوم پزشکی تبریز، ۱۳۸۳؛ ۷۵.