مجلـــه‌ی توســـعه‌ی آمـــوزش در علـوم پزشـــکی
دوره‌ی ۸، شماره ۱۸، تابستان ۹٤ ، صفحات ١٤ تا ٢٢

بررسی تاثیر کارگاه آموزش طراحی سوال بر کیفیت آزمون‌های کتبی دانشکده‌ی دندانپزشکی شهید بهشتی

دکتر مریم بهاروند[۱]،دکتر محمد حسین زاده[۲] ،دکتر سید جلیل صدر[۳]،دکتر زهرا جابری انصاری[٤] ، دکتر فرانه همتی علمداری[٥] ، دکتر حامد مرتضوی[۱] ، دکتر سیامک صبور[٦]، دکتر مینا مطلب نژاد[۷]

نویسنده مسوول: تهران، دانشکده‌ی دندانپزشکی، دانشگاه علوم پزشکی شهید بهشتی : hamedmortazavi2013@gmail.com

دریافت: ۹۳/۷/۲٦ پذیرش: ۹۳/۹/۲٤

چکیده

زمینه و هدف: *ارزشیابی عبارت از فرایند نظام مند گردآوری، تحلیل و تفسیر داده ها به منظور ارزش گذاری و قضاوت می باشد. مطالعه‌ی حاضر با هدف بررسی تاثیر کارگاه آموزش طراحی سوال برکیفیت آزمون های کتبی پایان ترم انجام پذیرفت.*

روش بررسی *: در این پژوهش نیمه تجربی ابتدا کلیه‌ی سوالات آزمون‌های دروس نظری۲ دانشکده‌ی دندانپزشکی شهید بهشتی قبل و پس از ارایه‌ی کارگاه و دفترچه راهنما با عنوان " قواعد طراحی سوال " جمع آوری شد. تاکسونومی، ارتباط و پوشش محتوایی آزمون‌ها توسط سه تن از مدرسین هر بخش آنالیز شد. سایر شاخص ها مانند ضرایب تمیز و دشواری،ارزش گزینه های انحرافی و مرزی ،ضریب خطا و میزان رعایت قواعد ساختاری با استفاده از چک لیست میل من (Millman) و فرمول های استاندارد محاسبه گردید. تحلیل داده ها با استفاده از تست های Independent Samples t. Test و Mann-Whitney Test و ANOVAانجام پذیرفت.*

یافته‌ها *: قبل از مداخله شاخص‌های ارتباط محتوایی و ضریب دشواری در حد مطلوب بودند. پس از مداخله نسبت سوالات تستی، ارتباط محتوایی و ضریب تمیز افزایش داشتند در حالی‌که میزان رعایت قواعد ساختاری آزمون ، ضریب دشواری ، میزان گزینه های انحرافی نیازمند بازبینی ،ضریب خطا ، تعداد سوالات نیازمند بازبینی و پوشش محتوایی کاهش نشان دادند . در کل مداخله فقط روی کاهش ضریب خطا تاثیر کاربردی داشت.*

نتیجه‌گیری : کارگاه کوتاه مدت آموزش روش های طراحی سوال بر کیفیت آزمون های پایان ترم تاثیر کاربردی نداشت و به نظر می رسد برای بهبود کیفیت آزمون نیاز به مداخلات دیگری نظیرآموزش مستمر وجود دارد.

واژگان کلیدی: آزمون نظری, آزمون کتبی , کیفیت، دندانپزشکی، کارگاه طراحی سوال

۱- متخصص بیماری‌های دهان،دانشیار دانشگاه علوم پزشکی شهید بهشتی
۲- دانشجوی PhD آموزش پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی شهید بهشتی
۳- متخصص پروتز ثابت، استاد دانشگاه علوم پزشکی شهید بهشتی
٤- متخصص ترمیمی، دانشیار دانشگاه علوم پزشکی شهید بهشتی
٥- دندانپزشک
٦- متخصص اپیدمیولوژی بالینی، دانشیاردانشگاه علوم پزشکی شهید بهشتی، تهران
۷- متخصص بیماری‌های دهان، دانشیار دانشگاه علوم پزشکی بابل، بابل

[Downloaded from edujournal.zums.ac.ir on 2026-11-11]
[DOR: 20.1001.1.22519521.1394.8.18.12.6]

مقدمه

ارزشیابی عبارت از فرایند نظام مند گردآوری، تحلیل و تفسیر داده ها به منظور ارزش گذاری و قضاوت می باشد به طوری که تعیین کند آیا هدف مورد نظر تحقق یافته یا در حال تحقق یافتن است و میزان تحقق یافتن آن چه قدر است(۱). اگر یک آزمون از طراحی مطلوب در زمینه انتخاب تاکسونومی سوالات و رعایت قواعد ساختاری و روایی محتوا و سایر شاخص های استاندارد سازی آزمون برخوردار نباشد در این صورت نه تنها نقش اصلی آزمون، جزء تکمیل کننده و حیاتی آموزش، از دست می رود بلکه این امر بر انگیزه فراگیران اثرات منفی بر جای گذاشته و خطر به هدر رفتن تلاش اساتید و سیستم آموزشی را خواهد داشت لذا بررسی کیفیت سوالات و استاندارد بودن آزمون ضروری است(۲). نتایج امتحانات نه تنها برای قضاوت درباره‌ی عملکرد فراگیران بلکه درباره‌ی اثربخشی آموزش، روش ها و مواد آموزشی نیز مفید هستند (۳).

یافته های شکورنیا در دانشگاه علوم پزشکی اهواز نشان داده است که اساتید مهارت کافی در ساخت آزمون های درسی را ندارند و برای اندازه گیری پیشرفت تحصیلی دانشجویان اغلب مفاهیم ساده و سطحی درس را اندازه گیری می‌کنند و در اکثر مواقع سوالات مبهم است(٤). همچنین مطالعه‌ی اشرفپوردر دانشگاه علوم پزشکی بابل نشان داد که طراحی آزمون های پایان ترم پرستاری وضعیت مناسب و راضی کننده نداشته و عموما آزمون های مورد استفاده در فرآیند ارزشیابی برآیندهای آموزشی دانشجویان دارای اشکال ساختاری اساسی بوده و این آزمون‌ها به‌عنوان ابزار دقیق و معتبر عمل نمی‌نمایند و ضمنا تفاوت‌های واقعی آزمودنی ها را نمی‌سنجند(۳).

مطالعه‌ی جامعی که سوالات آزمون های نظری پایان ترم دروس مختلف را از نظر تاکسونومی، رعایت اصول ساختاری، روایی محتوا، اعتبار صوری، ضریب تمیز و ضریب دشواری بررسی کرده باشند، تاکنون صورت نگرفته است. تحقیقات متعددی فقط در زمینه‌ی بررسی سوالات چند گزینه ای در دانشگاه ها و موسسات آموزش عالی داخل و خارج از کشور صورت گرفته است(۷-۵). از نتایج این بررسی‌ها چنین بر می‌آید که در این آزمون‌ها از سوالات مناسب کمتر استفاده شده، در اجرا و ساخت آزمون‌ها مسایل و مشکلاتی وجود دارد و بسیاری از مدرسان بدون این که آموزش کافی در سنجش و اندازه گیری و شیوه‌ی ساخت آزمون‌ها کسب کرده باشند، تدریس می‌نمایند.

برای رفع این مشکل ابتدا باید با در نظر گرفتن اولویت ها، کاستی ها و نقاط ضعف موجود را در هر بخش شناسایی کرد و با برگزاری دوره های آموزشی، معرفی منابع موجود و انجام دادن مطالعاتی در این زمینه مهارت لازم را در اساتید محترم ایجاد کرد.

با توجه به اهمیت ارزشیابی در نظام آموزش عالی و لزوم آمادگی و توانمند سازی اساتید در تهیه و طراحی آزمون‌های مناسب بر آن شدیم تا با بررسی سوالات آزمون های نظری پایان ترم بخش های مختلف دانشکده‌ی دندانپزشکی شهید بهشتی در یک نیم سال تحصیلی زمینه را برای طراحی نظام ارزشیابی و پایش آزمون‌های کتبی و اجرای آزمایشی آن فراهم آوریم.

روش بررسی

در این تحقیق نیمه تجربی (قبل و بعد) سنجش کیفیت محتوایی(Content Analysis) ، تاکسونومی (سطح شناخت)، سوالات آزمون ها و سطح شناخت مورد انتظار از طریق اتفاق نظر متخصصان، میزان رعایت شدن قواعد ساختاری از طریق چک لیست Millman و سایر متغیرها با استفاده از فرمول‌های استاندارد و توسط محقق انجام شده است(۸و۲).

[Downloaded from edujournal.zums.ac.ir on 2026-11-11]

[DOR: 20.1001.1.22519521.1394.8.18.12.6]

جامعه‌ی مورد بررسی در این تحقیق شامل کلیه‌ی آزمون‌های کتبی پایان ترم بخش های مختلف دانشکده‌ی دندانپزشکی شهید بهشتی در سال تحصیلی ۹۱-۱۳۹۰ می‌باشد. حجم نمونه در این تحقیق شامل کل سوالات آزمون های کتبی دروس ۲ نظری پایان ترم بخش‌های مختلف دانشکده در سال تحصیلی ۹۱-۱۳۹۰ و روش نمونه گیری به صورت سرشماری بود.

پس از برگزاری امتحانات پایان ترم نیمسال اول ۹۱-۱۳۹۰ طی نامه‌ای از طرف معاونت محترم آموزشی دانشکده از مدیر گروه های بخش های مختلف دانشکده خواسته شد تا برگه های تصحیح شده در اختیار مجری این تحقیق قرار گیرد و برگه های تصحیح شده توسط مجری تحقیق از بخش های مختلف دانشکده جمع آوری شد,سپس مداخله تحقیق در قالب کارگاهی با عنوان "قواعد طراحی سوال" برگزار گردید که در طی آن دفترچه‌ای هم با عنوان "مروری بر قواعد طراحی سوال" به اساتید مدرس و طراح سوالات پایان ترم دروس ۲ نظری ارایه گردید.همچنین پرسشنامه ای تهیه شد و در مورد آموزش دیدن در خصوص نحوه‌ی طراحی سوال از جمله شرکت‌در کارگاه‌های مرتبط یا مطالعه‌ی کتاب و مقاله‌ی مرتبط از یک سال قبل از زمان طراحی سوالات نیمسال اول تا زمان طراحی سوالات نیمسال دوم از تمامی اساتید سوال شد که در طی آن معلوم گردید از ٤٠ تن از اساتید دروس نظری ۲ تنها یک نفر و آن هم ٦ ماه قبل از طراحی سوالات نیمسال اول اطلاعاتی دریافت کرده‌اند که این موضوع با توجه به زمان آن و مقایسه‌ای بودن تحلیل اطلاعات قابل صرفنظر می‌باشد.سپس بعد از برگزاری امتحانات نیمسال دوم طی نامه مجددی از مدیر گروه‌های بخش‌های مختلف دانشکده خواسته شد تا برگه‌های تصحیح شده نیمسال دوم نیز در اختیار مجری این تحقیق قرار گیرد و برگه های تصحیح شده مجددا توسط مجری تحقیق از بخش‌های مختلف دانشکده جمع آوری شد. در مورد همه بخش های شرکت کننده اعتبار محتوایی,سطح شناختی (تاکسونومی) و میزان رعایت قواعد ساختاری سوالات و آزمون‌ها بررسی شد.

ولی فقط در مورد آزمون‌هایی که دسترسی به برگه های تصحیح شده وجود داشت علاوه بر متغیرهای فوق ضریب تمیز , ضریب دشواری , ارزش مورد نظر برای هر گزینه انحرافی , میزان گزینه های انحرافی نیاز مند بازبینی , میزان گزینه های مرزی, درصد سوالاتی که از لحاظ ضریب تمیز و ضریب دشواری نیازمند بازبینی‌اند، قابل محاسبه بود چرا که امکان دسترسی به برگه‌های تصحیح شده بخش‌های جراحی فک و صورت و ارتودنسی فراهم نشد.

چگونگی سنجش متغیرها(۲) :

ضریب تمیز سوالات با استفاده از فرمول :

$$\frac{\text{تعداد انتخاب های درست گروه بالا} - \text{تعداد انتخاب های درست گروه پایین}}{\text{تعداد افراد یک گروه(بالا یا پایین)}}$$

ضریب دشواری با استفاده از فرمول :

$$\frac{\text{تعداد انتخاب های درست گروه پایین} + \text{تعداد انتخاب های درست گروه بالا}}{\text{تعداد افراد گروه بالا} + \text{تعداد افراد گروه پایین}}$$

ارزش مورد نظر برای هر گزینه انحرافی با استفاده از فرمول :

$$\frac{\text{تعداد انتخاب های گروه بالا} + \text{تعداد انتخاب های گروه پایین}}{\text{تعداد افراد گروه بالا} + \text{تعداد افراد گروه پایین}}$$

بررسی میزان گزینه های انحرافی نیاز مند بازبینی:

$$\frac{\text{تعداد گزینه هایی که کمتر از 10\% انتخاب ها را به خود اختصاص داده اند}}{\text{تعداد کل گزینه های انحرافی}}$$

بررسی میزان گزینه های مرزی :

$$\frac{\text{تعداد گزینه هایی که 10\% انتخاب ها را به خود اختصاص داده اند}}{\text{تعداد کل گزینه های انحرافی}}$$

چگونگی به دست آوردن نمرات افراد گروه بالا و پایین :

[Downloaded from edujournal.zums.ac.ir on 2026-11-11]

[DOR: 20.1001.1.22519521.1394.8.18.12.6]

لیست نمرات دانشجویان از نمرات بالا به پایین مرتب شده و برای دروسی که تعداد پاسخ دهندگان آن بیشتر از ۳۰ نفر است ۱۰ نمره بالا و ده نمره پایین لیست به ترتیب به عنوان گروه بالا و گروه پایین در نظر گرفته می شوند.لیست نمرات هر آزمون از طریق آموزش در اختیار محقق قرار گرفت (به دلیل رعایت اصل اخلاق از لیست نمرات اصلی کپی گرفته شد و پس از جدا کردن اسامی دانشجویان , لیست نمره ها تنها به همراه شماره های دانشجویی به محقق ارایه گردید) سپس برگه های امتحانی افراد گروه بالا و پایین هر آزمون از بین برگه های آزمون جدا شده و جواب های داده شده به تمامی سوالات و گزینه ها به صورت تک تک مورد بررسی قرار گرفت و در فرمول های فوق قرار داده شد.

چگونگی بررسی میزان رعایت شدن قواعد ساختاری :

چک لیست Millman برای سنجش میزان رعایت شدن قواعد ساختاری مورد استفاده قرار گرفت.[۸]

با استفاده از جدول قواعد ساختاری سوالات و فرمول زیر ضریب خطا برای هر آزمون به دست آمد :

$$\text{درصد ضریب خطا} = \frac{\text{تعداد خطا های کل هر آزمون}}{\text{تعداد سوالات}} \times 100$$

نحوه‌ی بررسی سطح شناختی (تاکسونومی) و ارتباط محتوایی (Content Relevance) سوالات آزمون ها:

این دو متغیر توسط سه نفر از اعضای هیات علمی هر بخش سنجیده شد. در این تحقیق کارگاهی با حضور این اساتید برگزار گردید.پس از تدریس موارد مرتبط , جدول ارتباط محتوایی و تاکسونومی همراه با یک نمونه سوال از هر نیمسال در اختیار هر یک از اساتید قرار داده شد و در پایان جلسه جداول تکمیل شده از اساتید تحویل گرفته شد. در ستون‌های جدول سوالات و در سطرهای جدول اهداف آموزشی قرار داده شدند که در زیر هر سوال و روبروی هدف مربوط به آن سوال تیک زده شد که مشخص می‌کرد سوال مربوط به همان هدف می باشد (Content Relevance) و قرار دادن عدد ۱ یا ۲ در کنار علامت تیک نشان دهنده‌ی تاکسونومی هر سوال بود. عدد ۱ برای سوالاتی که محفوظات دانشجو را می سنجد و فقط جنبه یادآوری دارد و عدد ۲ برای سوالاتی که دانشجو برای پاسخگویی نیاز به تجزیه و تحلیل بیشتری داشته و جنبه کاربردی دارد, در نظر گرفته شد.

Content Relevance مشخص کننده‌ی مرتبط بودن سوالات آزمون با اهداف آموزشی مورد نظر آن درس می باشد و وقتی ۱۰۰ درصد می‌باشد که همه سوالات مرتبط با اهداف آموزشی باشند و سوالی نباشد که هدفی را نسنجیده باشد یا غیر مرتبط با اهداف یا محتوای درس مورد نظر باشد. درصد ارتباط محتوایی از تقسیم کردن تعداد سوالات مرتبط با اهداف درس بر تعداد کل سوالات و ضرب حاصل آن در ۱۰۰ بدست می آید.

Content Coverage مشخص کننده‌ی پوشش اهداف آموزشی می باشد یا به عبارت دیگر برای هر هدف متناسب با اهمیت آن تعداد مناسبی سوال در نظر گرفته شده باشد و هدفی نباشد که برای سنجش آن سوالی در نظر گرفته نشده باشد. درصد پوشش محتوایی از تقسیم کردن تعداد اهداف لحاظ شده در سوالات بر تعداد کل اهداف آموزشی و ضرب حاصل مربوطه در ۱۰۰ به‌دست می‌آید.

طرح تجزیه و تحلیل داده ها:

داده های به‌دست آمده از مطالعه با استفاده از نرم افزار آماری *S.P.S.S.19* و تست هایIndependent Samples T. Test و Mann-Whitney test و ANOVA مورد تجزیه و تحلیل قرار گرفتند.

یافته ها

از مجموع ۱۰ بخش تخصصی دانشکده‌ی دندانپزشکی شهید بهشتی که دارای دروس ۲ نظری بودند, ۸ بخش با در اختیار گذاشتن سـوالات پایان تـرم نیم‌سال اول و دوم سال

[Downloaded from edujournal.zums.ac.ir on 2026-11-11]

[DOR: 20.1001.1.22519521.1394.8.18.12.6]

تحصیلی ۹۱-۱۳۹۰ در این مطالعه شرکت کردند.

از مجموع ۳۲۱ سوال آزمون های نیمسال قبل از مداخله بیشترین فراوانی را سوالات چندگزینه ای(۷۸ درصد) به خود اختصاص دادند پس از آن سوالات صحیح- غلط(۱۲ درصد) و تشریحی (۱۰ درصد) قرار داشتند.

پس از مداخله، از مجموع ۲۸۸سوال بیشترین فراوانی مربوط به سوالات چند گزینه ای(۸۲ درصد) بود و بعد از آن به ترتیب سوالات تشریحی(۱۳ درصد) و صحیح- غلط (۵ درصد) قرار داشتند.

میزان رعایت شدن قواعد ساختاری آزمون در آزمون های نیمسال قبل و بعد از مداخله بخش های مختلف در جدول ۱ نشان داده شده است.بر طبق این جدول میزان رعایت قواعد ساختاری آزمون های نیمسال اول از ۵۰-۱۰۰ درصد و نیمسال دوم از ٤٠ تا ۱۰۰ درصد متغیر بوده است که میزان این تغییر معنا دار نبود (P=۰/۷۵).

ضریب خطا برای آزمون های نیمسال قبل و پس از مداخله بخش های مختلف در جدول ۱ نشان داده شده است.بر طبق این جدول آزمون های درس اندودانتیکس کمترین ضریب خطا را در هر دو نیمسال داشته اند.مداخله روی این متغیر فاقد تاثیر معنی دار از لحاظ آماری بود (P=۰/۳۱) در حالیکه تغییر آن از نظر آموزشی معنی دار بود(P=۰/۰۲).

ضریب خطا ی آزمون ها در نیمسال قبل بین ۰/۳ تا ۱/۲ و بعد از مداخله بین ۰/۱ تا ۱ متغیر بوده است (جدول ۱).

میزان ارتباط محتوایی (Content Relevance) هر دو نیمسال در جدول ۱ آورده شده که با توجه به آن‌ها ارتباط محتوایی آزمون آسیب شناسی بالاترین رشد را نشان می دهد.ولی در مجموع بین دو ترم تغییرات ناشی از مداخله فاقد تاثیر معنی‌دار بود(P=۰/٦٥).

جدول۱ میزان پوشش محتوایی (Content Coverage) را در هر دو نیمسال نمایش می‌دهد که بالاترین پوشش محتوایی مربوط به بخش بیماری‌های دهان بود.تغییرات پوشش محتوایی قبل و پس از مداخله فاقد تاثیر معنی دار بود (P=۰/٦٥) .

درصد تاکسونومی I مورد انتظار بر اساس نظر خبرگان , درصد تاکسونومی I به دست آمده از آزمون های هر دو نیمسال و اختلاف هر یک با تاکسونومی I مورد انتظار در جدول ۱ آورده شده اند که با توجه به این جدول, آزمون های دروس ترمیمی و آسیب شناسی دارای نزدیکترین درصد تاکسونومی I در هر دو نیمسال , به درصد تاکسونومی I مورد انتظار بوده‌اند.تحلیل داده‌ها عدم تاثیر معنی دار مداخله, چه از لحاظ آماری و چه از لحاظ آموزشی را نشان می‌دهد (P=۰/۸۵) .

پس از تست نرمالیته نحوه‌ی توزیع ضریب تمیز نرمال به-دست آمد, در این حالت مقایسه با توجه به میانگین و تحلیل آن با استفاده از Independent Samples Test انجام شد.

با توجه به اطلاعات به دست آمد، از کل آزمون‌ها, دامنه‌ی تغییرات ضریب تمیز از ٤٠- تا ۹۰+ و میانگین کل آن ۲۱/٤٥۸± ۲٤/٦۲ بوده است و به تفکیک نیمسال بدین گونه است که میانگین ضریب تمیز آزمون ها در نیمسال قبل از مداخله ۲۱/۱٤۹ ± ۲۳/۹۸ و دامنه‌ی تغییرات آن از ٤٠- تا ۹۰ + بوده است و پس از مداخله میانگین آن به ۲۱/۸۱۰ ± ۲۵/۳۰ رسید و دامنه‌ی تغییرات آن از ۳۰- تا ۹۰ + بوده است.

پس از مقایسه‌ی نتایج فوق و با وجود افزایش ۱/۳ درصد میانگین ضریب تمیز در نیمسال پس از مداخله, این افزایش بهبود معنی داری را نشان نداد(P=۰/۳۵).

پس از تست نرمالیته , نحوه‌ی توزیع ضریب دشواری غیرنرمال به‌دست آمد لذا برای تحلیل آن از Mann-Whitney Test استفاده شد.

دامنه تغییرات ضریب دشواری در هر دو نیم‌سال از ۵+ تا ۱۰۰+ متغیر بوده است و میانگین کل آن ۲۱/۸۳۳ ± ٦۸/٦۱ می‌باشد و با توجه به نتایج حاصل از Mann-Whitney Test تغییر معنی داری پس از مداخله مشاهده نشد(P=۰/۵۵).

[Downloaded from edujournal.zums.ac.ir on 2026-11-11]

[DOR: 20.1001.1.22519521.1394.8.18.12.6]

درصد سوالاتی که از لحاظ ضریب تمیز نیاز به بازبینی دارند در جدول ۱ آورده شده‌اند که بالاترین میزان بهبودی کیفی سوالات از این حیث مربوط به گروه پریودنتولوژی بود.از لحاظ آماری و آموزشی مداخله تاثیری روی این شاخص نداشته است(P=۰/۵۵).

درصدسوالاتی که از لحاظ ضریب دشواری نیاز به بازبینی دارند در جدول ۱ قابل مشاهده می‌باشد و بر اساس این جدول بالاترین میزان بهبودی کیفی سوالات از این لحاظ قبل و پس از مداخله مربوط به گروه پروتزهای دندانی بود و در کل مداخله‌روی این متغیر تاثیر معنی‌داری‌نداشته است(P=۰/۲۵).

جدول ۱.میزان رعایت اصول طراحی سوال در آزمون بخش‌های مختلف دانشکده‌ی دندانپزشکی شهید بهشتی قبل و بعد از کارگاه طراحی سوال

نام بخش	رعایت قواعد ساختاری آزمون درصد		Content Coverage درصد		Content Relevance درصد		تاکسونومی I درصد		تاکسونومی I مورد انتظار درصد	سوالات نیازمند به باز بینی از نظر ضریب تمیز درصد		سوالات نیازمند به باز بینی از نظر ضریب دشواری درصد		ضریب خطا درصد	
	ق	ب	ق	ب	ق	ب	ق	ب		ق	ب	ق	ب	ق	ب
بیماری‌های دهان	۸۰	۸۰	۷۶/۸	۸۳/۲	۱۰۰	۱۰۰	۹۰	۸۶/۶	۷۲/۳	۱۸/۳	۸/۳	۲۱/۶	۳۱/۶	۱/۱	۱/.
پریودنتولوژی	۸۰	۸۰	۶۸/۹	۵۵/۴	۱۰۰	۱۰۰	۸۸/۵	۸۳/۳	۶۵/۱	۲۰	۰	۲۶/۶	۲۱/۷	۱/۱	۰/۷
اطفال	۸۰	۸۰	۵۲/۷	۵۵/۴	۱۰۰	۹۶/۲	۹۱/۶	۱۰۰	۷۱/۲	۲۵/۸	۳۵/۷	۴۸/۳	۳۹/۲	۱	۰/۶
پروتز	۸۰	۸۰	۳۲/۵	۳۷/۲	۹۵/۸	۹۲/۴	۸۵/۲	۹۳/۹	۷۷/۴	۱۳/۶	۱۶/۶	۲۷/۲	۱۲/۵	۱/۲	۰/۷
رادیولوژی	۸۰	۸۰	۵۱/۳	۴۷/۳	۱۰۰	۱۰۰	۹۰	۹۵	۸۲/۳	۲۰	۲۰	۳۲/۵	۲۷/۵	۰/۳	۰/۱
ترمیمی	۶۶	۶۶	۳۹/۵	۴۷/۴	۹۷/۳	۹۸/۴	۶۲	۶۸/۱	۶۴/۱	۲۷	۳۸/۴	۵۰	۴۸	۱/۲	۰/۷
پاتولوژی	۵۰	۴۰	۶۳/۶	۳۴/۸	۸۲/۸	۱۰۰	۸۰/۹	۸۳/۷	۷۰/۹	۰	۸/۳	۱۱/۷	۱۶/۶	۰/۹	۰/۸
اندو	۸۰	۸۰	۰/۳	۰/۱	۱۰۰	۹۸/۸	۷۴	۷۵/۸	۶۷/۷	–	–	–	–	۰/۳	۰/۱
میانگین کل	۷۴/۵	۷۳/۲	۵۱/۷	۴۸/۱	۹۷	۹۸/۲	–	–	–	۲۲/۸	۱۹/۳	۳۲/۶	۳۰/۳	۱/۱	۰/۷

ق: قبل از کارگاه
ب: بعد ازکارگاه

بحث

پژوهش حاضر به منظور نیازسنجی بر پایه‌ی پیامد برای طراحی نظام ارزیابی کیفیت آزمون های کتبی پایان ترم و سنجش کیفیت آزمون های کتبی ۲ نظری بخش های مختلف دانشکده‌ی دندانپزشکی شهید بهشتی بر اساس شاخص‌های تحلیل کیفیت آزمون‌و سوالات وبررسی تاثیر ارایه‌ی روش‌های استاندارد طراحی سوال به اساتید بر کیفیت آزمون و سوالات در سال تحصیلی ۹۱-۱۳۹۰ انجام پذیرفت. با انجام این تحقیق می‌توان گفت که به جز درمورد متغیرهای معدودی, کارگاه نتیجه قابل ملاحظه‌ای بر کیفیت سوالات آزمون های پایان ترم نداشته و این نتیجه استقرار و طراحی نظام ارزیابی کیفیت آزمون های کتبی پایان ترم در جهت دسترسی بالاتر به

[Downloaded from edujournal.zums.ac.ir on 2026-11-11]
[DOR: 20.1001.1.22519521.1394.8.18.12.6]

اهداف آموزشی را قویا پیشنهاد می‌کند. محمدی یک مطالعه‌ی مقطعی بر روی ۲۲ نفر از اعضای هیات علمی دانشکده‌ی دندانپزشکی دانشگاه علوم پزشکی بابل که کارگاه ارزیابی سوالات را گذرانیده بودند، در سال های ۸۵ تا ۸۸ انجام داد. نتایج مطالعه او نشان داد که بررسی و ارزیابی سوالات امتحانی قبل از انجام امتحانات در داخل گروه های آموزشی در افزایش کیفیت سوالات امتحانی موثر بوده است(۹).

بررسی نوع سوالات امتحانی در مطالعه‌ی ما نشان داد که اعضای هیات علمی با توجه به موضوعی که تدریس می کنند از آزمون های یکسانی استفاده نمی‌کنند.حدود چهار پنجم سوالات نیمسال قبل از مداخله را سوالات چند گزینه ای تشکیل می‌دادند که در نیمسال دوم علی رغم مداخله آموزشی, شاهد رشد ۳ درصد در این نسبت بودیم که نشان می‌دهد بیشتر اساتید دانشگاه همچنان مایلند از سوالات عینی جهت سنجش پیشرفت تحصیلی دانشجویان استفاده کنند و نسبت آزمون های ذهنی یا تشریحی با کاهش مواجه شده‌اند. آزمون های عینی آزمون هایی هستند که در آنها هم سوال ها و هم جواب سوال ها در اختیار آزمون شوندگان قرار داده می‌شود و آزمون شوندگان در مورد جواب‌هایی که در اختیار دارند، تصمیماتی می‌گیرند, وجه تمایز این آزمون ها با آزمون های تشریحی و کوته پاسخ در این است که آزمون های تشریحی تنها وسیله‌ی موجود سنجش توانایی آزمون شونده در پروراندن جواب سوال‌ها و بیان آن‌هاست و این آزمون‌ها توانایی پاسخ دادن به سوال‌ها را می‌سنجند نه توانایی انتخاب پاسخ‌ها را, بنابراین موقعیت‌های واقعی تری را نسبت به آزمون‌های عینی به آزمون شوندگان عرضه می‌کنند.درنتیجه سنجش بهتر دانشجویان از طریق استفاده همزمان از آزمون های تشریحی و عینی حاصل می‌شود که با توجه به سنجش دروس ۲ نظری این موضوع از اهمیت بالاتری برخوردار می باشد. با توجه به جدول ۱ , درصد رعایت شدن قواعد ساختاری آزمون بر حسب آزمون ها (نمره‌ی کیفیت ساختاری آزمون) در نیمسال بعد از مداخله تفاوت معنی داری را نسبت به نیمسال قبل از مداخله نشان نمی‌دهد و حتی یکی از هشت گروه کاهش ۱۰ درصد را نشان می‌دهد. بر طبق جدول ۱ مقدار ضریب خطا پس از مداخله کاهش چشمگیری از لحاظ آموزشی‌را در تمامی گروه‌ها نشان می‌دهد و مداخله باعث بهبود بالای همه گروه‌ها شده است .

با توجه به جدول ۱ ارتباط محتوایی قبل از مداخله, تنها در گروه آسیب شناسی زیر ۹۰ درصدبوده که پس از مداخله همه درس ها به بالای ۹۰ درصد رسیده‌اند، یعنی سوالات آزمون‌ها با توجه به اهداف و محتوای آموزشی هر درس طراحی شده اند و سوالات غیرمرتبط با فصول درسی به ندرت در آزمون ها آورده شده است .

نتایج حاصل از بررسی پوشش محتوایی در نیمسال قبل و نیمسال بعد از مداخله سطح پایین این متغیر را در اغلب گروه‌ها نشان می‌دهد و مداخله تاثیری بر روی این متغیر نداشته‌اند.لذا با توجه به اینکه در تمامی گروه ها تنها از یک آزمون برای سنجش دانشجویان در طول ترم استفاده می‌شود، به نظر می‌رسد افزایش تعداد آزمون ها به افزایش متغیر پوشش محتوایی (Content Coverage) کمک کند که این افزایش موجب سنجش دقیق تر دانشجویان در مورد بیشترین تعداد اهداف آموزشی می شود.

به عنوان یک قاعده‌ی عرفی در بین متخصصان آموزش پزشکی میزان تاکسونومی ۲ باید حداقل ۳۰ درصد باشد,بر این اساس و با توجه به جدول ۱ , میزان تاکسونومی مورد انتظار در ۵ گروه از ۸ گروه مطلوب نمی‌باشد و با توجه به اینکه این بررسی در مورد دروس ۲ نظری می باشد و در این دروس دانشجویان نیاز به کسب توانایی استدلال و تجزیه و تحلیل دارند, می‌توان پیشنهاد کرد تجدید نظری در طرح آموزشی این دروس صورت پذیرد.در این بررسی آزمون های دروس ترمیمی و آسیب شناسی دارای نزدیکترین درصد تاکسونومی I در هر دو نیم‌سال , به درصد تاکسونومی I

[Downloaded from edujournal.zums.ac.ir on 2026-11-11]

[DOR: 20.1001.1.22519521.1394.8.18.12.6]

مورد انتظار بوده اند.

از نتایج تجزیه و تحلیل متغیر تاکسونومی در هر دو نیمسال و مقایسه آن‌ها با یکدیگر و با تاکسونومی مورد انتظار اینگونه درک می شود که مداخله تاثیر معنی داری از لحاظ کاربردی بر روی این متغیر نداشته، لذا نیاز به مداخله ای با زمان بیشتر و به طور مستمر احساس می‌شود.

با توجه به نتایج حاصل از تحلیل داده ها, حد واسط ضریب تمیز سوالات بعد از مداخله به میزان ۱/۳ درصد افزایش نشان می‌دهد ولی این افزایش چه از لحاظ آماری و چه از لحاظ آموزشی معنی دار نمی باشد. بالاترین میزان ضریب تمیز مربوط به گروه آسیب شناسی و پایین ترین میزان مربوط به گروه ترمیمی می باشد.

با توجه به اینکه ضریب تمیز بزرگتر نشان‌دهنده‌ی قوه‌ی تمیز آن سوال است و هر چقدر این ضریب به عدد یک نزدیک تر باشد، مناسب‌تر است؛ می‌توان گفت که در مجموع ضریب تمیز آزمون ها پایین بوده، قدرت تشخیص لازم برای تفکیک دانشجویان قوی از ضعیف را ندارند.

ضریب تمیز صفر حاکی از آن است که سوال مورد بررسی به هیچ وجه نتوانسته بین گروه قوی و ضعیف تمایز قایل شود.در مورد سوالات مورد بررسی به سوالاتی برخوردیم که ضریب تمیز منفی داشتند, این نوع ضریب تمیز نشان می دهد که در آن سوال گروه قوی بدتر از گروه ضعیف عمل کرده و این گونه سوال ها دارای معایب اساسی هستند که یا باید به کلی کنار گذاشته شوند یا در طراحی آن‌ها تجدید نظر اساسی صورت گیرد.اگر سوال دارای اشکال فنی نباشد, دلیل چنین ضریبی می تواند این باشد که افراد گروه بالا یا موضوع مربوط به آن سوال را به طور کلی یاد نگرفته اند و یا آن را به غلط آموخته اند.بنابراین تمامی ضرایب تمیز یک آزمون باید مثبت باشند(۱۰). درصد سوالاتی که از لحاظ ضریب تمیز نیاز به بازبینی دارند در جدول ۱ آورده شده اند که پس از مداخله این میزان فقط در دو گروه بیماری ها و پریودنتیکس کاهش قابل توجهی داشته و میانگین کلی این میزان تنها کاهش ۳ درصد را نشان می دهد که از لحاظ آموزشی اهمیتی ندارد .

با توجه به نتایج به دست آمده در این تحقیق می توان گفت که آزمون ها و سوالات اساتید از ضریب تمیز مناسبی برخوردار نمی باشد و سوالاتی که آن‌ها به کار برده اند توانایی تشخیص دانشجویان قوی از ضعیف را ندارد و این یک ضعف اساسی به شمار می رود و اعمال مداخله نیز تاثیر معنی داری بر این روند نداشته است.

سوالاتی که از لحاظ ضریب دشواری نیاز به بازبینی دارند سوالاتی هستند که میزان ضریب دشواری آن‌ها خارج از دامنه‌ی مطلوب است و لذا بسیار آسان یا بسیار دشوار می باشند که این سوالات فاقد اطلاعات شناختی در مورد آزمون شوندگان هستند.با توجه به جدول ۱ بالاترین میزان سوالاتی که از لحاظ ضریب دشواری نیاز به بازبینی دارند در هر دو نیمسال مربوط به گروه ترمیمی است و از طرفی کاهش بارزی در این میزان (۱٤/۷۷ درصد)در سوالات پس از مداخله در گروه پروتز دیده می‌شود.اما به طور کلی نتایج, تاثیر معنی داری را به سبب مداخله بر کیفیت این متغیر نشان نمی دهند.

با انجام این تحقیق می‌توان گفت که به جز درمورد متغیرهای معدودی, کارگاه نتیجه قابل ملاحظه ای بر کیفیت سوالات آزمون های پایان ترم نداشته است؛ این نتیجه استقرار و طراحی نظام ارزیابی کیفیت آزمون های کتبی پایان ترم در جهت دسترسی بالاتر به اهداف آموزشی قویا پیشنهاد می‌شود.

نتیجه گیری

نتایج حاصل از این مطالعه نشان داد که به جز متغیرهای معدودی, کارگاه نقش موثری بر کیفیت سوالات آزمون های پایان ترم نداشته و این نتیجه استقرار و طراحی نظام ارزیابی کیفیت آزمون های کتبی پایان ترم در جهت دسترسی بالاتر به اهداف آموزشی را قویا پیشنهاد می کند.

[Downloaded from edujournal.zums.ac.ir on 2026-11-11]

[DOR: 20.1001.1.22519521.1394.8.18.12.6]

تقدیر و تشکر

این مقالـه بر اساس پایان نامه‌ی دانشجویی دوره‌ی دندانپزشکی عمومی دانشکـده‌ی دندانپـزشکـی دانشگـاه علوم پزشکی شهید بهشتی بود، که در سال ۹۲-۹۱ انجام گردید. نگارنـدگان از همکـاری صمیمـانه دفتـر توسعه‌ی آموزش و اعضـای هیـات علمی دانشکدـه‌ی دندانپزشکـی که در انجام این تحقیق همکاری نمودند، صمیمانه تشکـر می‌کنند.

References

1- Bazargan A. Educational assessment. Tehran: SAMT; 2001, 79-91.

2- Seif A. Educational measurement, assessment & evaluation. Tehran: AGAH; 2007, 60-132.

3- Ashraf Pour M, Beheshti Z, Molook Zadeh S. Quality of final examination in students of Babol Medical University, 1999-2000. *JBUMS*. 2003; 5 (5) :42-47

4- Shakoornia A, Khosravi A, Shariati A, Zarei A. Survey on multiple choice questions of faculty members of Jondi Shapoor Medical University of Ahvaz. The 8th national congress of medical education Kerman. 2009; 4 (2): 44-59.

5- Tarrant M, Kneirim A: The frequency of item writing flaws in multiple choice questions high-stakes nursing assessments. *Nar Educoda*. 2006; 26(8): 662-71.

6- Sepasi H, Attari YA. The study of psychometric characteristics of Shahid Chamran University Faculty members final test scores. *Journal of Education and Psychology* (Special Issue an Education). 2006; 3(4): 1-20.

7- Haghshenas M, Vahidshahi K, Mahmudi M, Shahbaznejad L, Parvinnejad N, Emadi A. Evaluation of multiple choice questions in the school of medicine, Mazandaran University of Medical Sciences, the first semester of 2007. *Strides in Development of Medical Education*. 2008; 5(2): 119-126.

8- Millman J. Student achievement as a measure of teacher competent. Handbook of teacher evaluation. London: Bev. CA Sage Publication; 1981, 146-166.

9- Mohammadi, H, Javadian Y, Nikbakhsh N, Javanian M, Jahanian Z. Effects of self assessment of the examination questions by educational groups on the quality of final exams. *JBUMS*. 2010; 12 (5): 65-69.

10- Cronbach L J. Essentials of psychological testing. New York: Harper and Row ;1990, 48-49.

[Downloaded from edujournal.zums.ac.ir on 2026-11-11]

[DOR: 20.1001.1.22519521.1394.8.18.12.6]

A Research on the Efficacy of Test Development Workshop on the Quality of Theoretical Exams of Shahid Beheshti Dental School

Baharvand M[1], Hossein- Zadeh M[2], Sadr SJ[3], Jaberi-Ansari Z[4], Hemmati-Alamdari F[5], Mortazavi H[1], Sabour S[6], Motallebnejad M[7]

[1]Dept. of Oral Medicine, Dental School, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2]Medical School, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[3]Dept. of Fixed Prosthodontics, Dental School, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[4]Dept. of Operative Dentistry, Dental School, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[5]Dental School, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[6]School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[7]Dept. of Oral Medicine, Dental School, Babol University of Medical Sciences, Babol, Iran

Corresponding Author: Mortazavi H, Dept. of Oral Medicine, Dental School, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Email: hamedmortazavi2013@gmail.com
Received: 18 Oct 2014 **Accepted:** 15 Dec 2014

***Background and Objective*:** Evaluation is a systematic process of gathering, analyzing and interpreting data for making assessment and judgment. This study aimed to evaluate the efficacy of a test development workshop on the quality of final written exams.

Materials and Methods: In this semi-experimental study, written final exams of all theoretical courses before and after presenting a workshop were gathered and a booklet titled "Standard rules of questions" was distributed. Taxonomy, content relevance, and content coverage of exams were analyzed by three course lecturers of each department. Other indices such as difficulty and differential coefficients, value of distracter and borderline choices, error coefficient, and considering structural rules were calculated by means of Millman's checklist and standard formulae. Independent-Samples T-Test, Mann-Whitney U Test, and ANOVA were used to analyze the data.

Results: Before intervention, content relevance and difficulty coefficient were in acceptable levels. Proportion of multiple choice questions, content relevance, and differential coefficient increased after intervention, whereas consideration of structural rules, difficulty coefficient, number of distracters requiring revision, error coefficient, number of questions requiring revision, and coverage relevance decreased. Generally, the intervention just decreased the error coefficient practically.

Conclusion: A short- term course on standard rules of question development does not affect quality of exam questions considerably and a long term education is suggested in this regard.

Keywords: Theoretical exam, written exam, Quality, Dental, Exam development workshop

[Downloaded from edujournal.zums.ac.ir on 2026-11-11]

[DOR: 20.1001.1.22519521.1394.8.18.12.6]