

## بررسی تحلیلی شاخص‌های کمی آزمون‌های چندگزینه‌ای گروه ایمنی شناسی دانشکده پزشکی دانشگاه علوم پزشکی جندی‌شاپور اهواز

عبدالحسین شکورنیا\*: عضو هیأت علمی، گروه ایمنی شناسی، دانشکده پزشکی، دانشگاه علوم پزشکی جندی‌شاپور اهواز، اهواز، ایران.

مهری غفوریان: عضو هیأت علمی، گروه ایمنی شناسی، دانشکده پزشکی، دانشگاه علوم پزشکی جندی‌شاپور اهواز، اهواز، ایران.

علی خدادادی: عضو هیأت علمی، گروه ایمنی شناسی، دانشکده پزشکی، دانشگاه علوم پزشکی جندی‌شاپور اهواز، اهواز، ایران.

عطاءالله غدیری: عضو هیأت علمی، گروه ایمنی شناسی، دانشکده پزشکی، دانشگاه علوم پزشکی جندی‌شاپور اهواز، اهواز، ایران.

افشین آماری: عضو هیأت علمی، گروه ایمنی شناسی، دانشکده پزشکی، دانشگاه علوم پزشکی جندی‌شاپور اهواز، اهواز، ایران.

**چکیده:** با توجه به استفاده گسترده از سوالات چندگزینه‌ای در ارزیابی دانشجویان پزشکی، پژوهش حاضر به منظور بررسی تحلیلی آزمون‌های چندگزینه‌ای گروه ایمنی شناسی دانشگاه علوم پزشکی جندی‌شاپور اهواز در سال ۱۳۹۶ انجام گرفت. در این مطالعه توصیفی مقطعی، اطلاعات مربوط به آزمون‌های چندگزینه‌ای گروه ایمنی شناسی که با دستگاه OP-Scan تصحیح شده بود مورد بررسی قرار گرفت (۸۰۸ سوال). ضریب دشواری، ضریب تمیز و گزینه‌های انحرافی محاسبه گردید. داده‌ها با استفاده از آمار توصیفی و ضریب همبستگی پیرسون تحلیل شد. میانگین ضریب دشواری سوالات  $0/59 \pm 0/25$  بود و  $46/2$  درصد سوالات ضریب دشواری مناسب داشتند. میانگین ضریب تمیز سوالات  $0/25 \pm 0/24$  بود و  $57/3$  درصد سوالات ضریب تمیز مناسب داشتند. تجمیع ضریب دشواری و تمیز سوالات نشان داد که فقط تعداد ۲۴۸ سوال (۳۰/۷ درصد) از سوالات چندگزینه‌ای ایده‌آل می‌باشند. تعداد ۱۵۲۵ گزینه (۶۲/۹۰٪) از ۲۴۲۴ گزینه مورد بررسی گزینه انحرافی مناسب (Functional distractor=FD) و ۸۸۹ (۳۷/۰٪) گزینه انحرافی نامناسب (Non-functional distractors=NFDs) بودند. یافته‌ها نشان داد که در مجموع سوالات از کیفیت لازم برخوردار نبوده و نیازمند بازبینی و اصلاح می‌باشند.

**واژگان کلیدی:** سوالات چندگزینه‌ای، تحلیل آزمون، ضریب دشواری، ضریب تمیز، گزینه‌های انحرافی، ارزشیابی آموزشی.

\***نویسنده‌ی مسؤوول:** عضو هیأت علمی، گروه ایمنی شناسی، دانشکده پزشکی، دانشگاه علوم پزشکی جندی‌شاپور اهواز، اهواز، ایران.

Email: shakurnia@yahoo.com

## مقدمه

ارزشیابی جزء لاینفک آموزش و یادگیری است. در آموزش پزشکی ارزشیابی دانشجویان به روشهای مختلف انجام می شود. یکی از این روش ها استفاده از آزمون های چند گزینه ای می باشد. آزمون های چندگزینه ای رایج ترین نوع آزمون کتبی می باشند که از حدود پنج دهه قبل در آزمون های کتبی بوردهای تخصصی در جهان مورد استفاده قرار می گیرند. در ایران نیز در دانشگاههای علوم پزشکی آزمون های چندگزینه ای متداول ترین آزمون ها در ارزشیابی دانشجویان می باشند (وندربلت و همکاران ۲۰۱۷، کوهپایه زاده و همکاران ۲۰۱۲، ابوزید و همکاران ۲۰۱۷).

لازمه تهیه سوالات چند گزینه ای مناسب، آشنایی کامل با اصول طراحی این سوالات بویژه نحوه تدوین صحیح پایه سوال و انتخاب گزینه های انحرافی برای هر سوال می باشد. به همین دلیل بعد از برگزاری آزمون های چندگزینه ای، بررسی کیفیت این آزمون و تجزیه و تحلیل سوالات از نظر شاخص های کمی و کیفی برای تعیین سوالات مناسب و تشکیل بانک سؤال و بویژه استاندارد سازی این آزمون ها در دانشگاهها و مراکز آموزشی ضروری است (مکوبری ۲۰۱۴، کارکال و کانداپور ۲۰۱۶، شکونیا و همکاران ۲۰۱۰a، پاتل ۲۰۱۷).

در تحلیل سوالات چند گزینه ای با تعیین شاخص های کمی سوالات از قبیل ضریب دشواری، ضریب تمیز و گزینه های انحرافی، کیفیت سوالات چند گزینه ای مورد ارزیابی قرار می گیرد. و سپس سوالات نامناسب اصلاح و یا حذف می شوند. گزارشات مختلف از بررسی سوالات چند گزینه ای در مراکز آموزشی نشان می دهد که درصد بالایی از این سوالات نیازمند بازبینی و اصلاح می باشند (پاتیل و همکاران ۲۰۱۷، درخشان و همکاران ۲۰۱۶، گاجر و همکاران ۲۰۱۴).

نتایج پژوهش های مختلف نشان داده است که با تحلیل آزمون ها و ارائه بازخورد به طراحان سوال و آموزش روش های مناسب طراحی سوال میتوان به ارتقاء و بهبود کیفیت سوالات آزمون ها کمک کرد (درخشان و همکاران ۲۰۱۵، مهیاری و بیگلرخانی

۲۰۱۳، امین و همکاران ۲۰۱۱). در این فرایند، مدرس با بررسی سوالات امتحانی سعی می کند ضمن تحلیل سوالات امتحانی، سوالات مناسب را انتخاب و سوالات نامناسب را اصلاح و یا حذف نماید. و در نهایت با تجزیه و تحلیل نتایج آزمون ها و گزینش سوالات مناسب و تدوین یک بانک سؤال جامع روایی و پایایی آزمون ها را افزایش دهد. هدف از مطالعه حاضر بررسی شاخص های کمی سوالات چند گزینه ای طراحی شده در بخش ایمنی شناسی دانشکده پزشکی دانشگاه علوم پزشکی جندی شاپور اهواز در سال تحصیلی ۱۳۹۶ بوده است.

## روش کار

در این مطالعه توصیفی- تحلیلی سوالات آزمون های چند گزینه ای که توسط اساتید گروه ایمنی شناسی برای ارزیابی دانشجویان رشته های پزشکی، دندانپزشکی، داروسازی، علوم آزمایشگاهی و پرستاری مامایی در ترم تحصیلی اول و دوم سال ۹۶-۱۳۹۵ در دانشگاه علوم پزشکی جندی شاپور اهواز طراحی شده بود مورد بررسی قرار گرفت.

داده های مطالعه از بانک تصحیح اوراق امتحانی واحد ماشینی معاونت آموزشی دانشگاه بدست آمد. پس از برگزاری آزمون، پاسخنامه های تک تک دانشجویان با دستگاه OP-Scan قرائت و با نرم افزار ویژه تجزیه و تحلیل سوالات چند گزینه ای آنالیز شده و نتایج تحلیلی هر آزمون به صورت یک بانک مجزا در کامپیوتر ذخیره شد.

گزارشات ارزیابی حاصل از تجزیه و تحلیل سوالات چند گزینه ای شامل، مشخصات روانسنجی آزمون اعم از ضریب دشواری (Difficulty Index) و قدرت تمیز (Discrimination Index) هر یک از سوالات آزمون ها، فراوانی پاسخ گویی دانشجویان به هر یک از گزینه های انحرافی (options distractors) و همچنین محل قرار گرفتن پاسخ صحیح در بین گزینه های مختلف بود. سپس این مشخصه ها و آماره های فوق با توجه به استانداردهای آزمون های پیشرفت تحصیلی از قبیل ضریب دشواری و تمیز سوال و گزینه های انحرافی مورد بررسی و تحلیل قرار گرفت.

میزان تاثیر گزینه های انحرافی مورد ارزیابی قرار می گیرد. در هر سوال چهار گزینه ای تعداد گزینه های انحرافی نامناسب از یک تا سه متغیر است. اگر در یک سوال چهار گزینه ای، تعداد گزینه های انحرافی نامناسب (NFDS) به ترتیب سه، دو، یک و یا صفر باشد، ضریب تاثیر یا کارایی (DE) آن سوال  $0\%$ ،  $33\%$ ،  $66\%$  و  $100\%$  در نظر گرفته می شود (تارانت و همکاران ۲۰۰۹).

آزمون ها و سوالات مورد بررسی بدون نام و مشخصات فردی بود و کلیه ملاحظات اخلاقی در این مطالعه از جمله محرمانه بودن اطلاعات فردی رعایت شد. داده ها با کمک نرم افزار SPSS و با استفاده از آمار توصیفی از قبیل فراوانی، درصد، میانگین و انحراف معیار و آزمون های مجذور کای و ضریب همبستگی پیرسون تجزیه و تحلیل شدند و مقادیر  $p$  کمتر از  $0/05$  معنی دار در نظر گرفته شد.

#### یافته ها

در مجموع ۸۰۸ سوال چند گزینه ای درس ایمنی شناسی مورد بررسی قرار گرفت. این سوالات توسط اساتید گروه ایمنی شناسی تهیه شده و جهت ارزشیابی دانشجویان رشته های مختلف در طول یکسال تحصیلی مورد استفاده قرار گرفت. میانگین ضریب دشواری سوالات مورد بررسی  $0/25 \pm 0/59$  بود و  $46/2\%$  درصد سوالات ضریب دشواری مناسب داشتند. میانگین ضریب تمیز سوالات  $0/24 \pm 0/25$  بود و  $25/6\%$  درصد سوالات ضریب تمیز مناسب داشتند (جدول ۱). تجمیع ضریب دشواری و تمیز سوالات نشان داد که فقط تعداد ۲۴۸ سوال ( $30/7\%$  درصد) از سوالات چند گزینه ای ایده آل می باشند. و تعداد ۵۶۰ سوال ( $69/3\%$  درصد) فاقد این ویژگی می باشند.

از تعداد ۲۴۲۴ گزینه انحرافی در سوالات چند گزینه ای مورد بررسی، تعداد ۱۵۲۵ ( $62/90\%$ ) گزینه انحرافی مناسب (FD) و تعداد ۸۹۹ ( $37/10\%$ ) گزینه انحرافی نامناسب (NFD) بودند. بنابراین، در مجموع میانگین ضریب تاثیر گزینه های انحرافی (DE) در سوالات مورد بررسی  $62/8\%$  درصد بود.

شاخص دشواری عبارت است از نسبت افرادی که به یک سوال پاسخ صحیح داده اند. شاخص دشواری یک شاخص وارونه است. هر چه مقدار این شاخص کوچکتر باشد سوال مشکل تر و هر چه مقدار آن بزرگتر باشد سوال آسان تر است. سطح دشواری مطلوب برای یک سوال بدون توجه به نوع سوال حدود  $50\%$  درصد (بین  $30\%$  تا  $70\%$ ) در نظر گرفته می شود. ضریب دشواری کمتر از  $0/3$  خیلی سخت،  $0/3$  تا  $0/7$  مناسب و بالاتر از  $0/7$  خیلی آسان در نظر گرفته می شود (مخرجی و لاهییری ۲۰۱۵).

قدرت تمیز سوال عبارت است از میزان قدرت سوال در تفکیک آزمون شوندگان دارای نمره بالا، از آزمون شوندگان دارای نمره پائین. هر چه مقدار ضریب تمیز بزرگتر باشد سوال از توان بیشتری در جدا کردن گروه قوی و ضعیف برخوردار می باشد. ضریب تمیز کمتر از  $0/2$  ضعیف،  $0/2$  تا  $0/34$  مناسب و بالاتر از  $0/35$  عالی در نظر گرفته شد (کویینگرین و آرهین ۲۰۱۷).

برای نشان دادن سوالات ایده آل از تجمیع دو ضریب دشواری و تمیز سوالات استفاده می شود. سوالات چند گزینه ای ایده آل، سوالاتی هستند که ضریب دشواری  $0/30$  تا  $0/70$ ؛ و ضریب تمیز بالاتر از  $0/24$  داشته باشند (هینگرچو و جلیل ۲۰۱۲).

هر سوال چهار گزینه ای دارای یک گزینه کاملاً درست و سه گزینه نادرست می باشد که به این سه گزینه اصطلاحاً گزینه های انحرافی گفته می شود. گزینه های انحرافی بایستی به گونه ای تهیه شوند که بتوانند نظر تعدادی از آزمون شوندگان (دانشجویان) را به خود جلب کنند. به گزینه هایی انحرافی که بتوانند حداقل  $5\%$  آزمون شوندگان را به خود جلب کنند، گزینه های انحرافی مناسب (functional or effective distractor=FDs) و به گزینه های انحرافی که نظر کمتر از  $5\%$  آزمون شوندگان را به خود جلب کنند گزینه های انحرافی نامناسب (non-functional distractor=NFDs) گفته می شود. در تحلیل تاثیر یا کارایی گزینه های انحرافی (Distractor Efficiency=DE) که بر اساس تعداد گزینه های انحرافی نامناسب (NFDS) از  $0-100\%$  متغیر است

رابطه مثبت معنی داری نشان داد ( $r = 0/528$ ) ،  
 $p = 0/0001$  . یعنی با افزایش تعداد گزینه های  
 انحرافی نامناسب ضریب دشواری سوال نیز افزایش می  
 یابد، به عبارت دیگر هرچه تعداد گزینه های انحرافی  
 نامناسب در یک سوال افزایش پیدا می کند از درجه  
 دشواری آن سوال کاسته می شود و سوال آسان تر می  
 شود. همچنین بین ضریب دشواری و ضریب تمیز  
 سوالات رابطه منفی معنی داری مشاهده داد ( $r = -0/099$  -  
 $p = 0/005$ ) . به این معنی که سوالات آسان تر  
 دارای ضریب تمیز کمتر بودند.

از مجموع ۸۰۸ سوال چند گزینه ای مورد بررسی، در  
 ۲۷۹ سوال (۳۴/۵٪) هر سه گزینه انحرافی مناسب  
 بودند و تعداد ۵۲۹ سوال (۶۵/۵٪) دارای ۱، ۲ و یا ۳  
 گزینه انحرافی نامناسب بودند؛ تعداد ۲۶۲  
 سوال (۳۲/۴٪) دارای ۱ گزینه انحرافی نامناسب، ۱۶۴  
 سوال (۲۰/۳٪) دارای ۲ گزینه انحرافی نامناسب و در  
 ۱۰۳ سوال (۱۲/۷٪) هر سه گزینه انحرافی نامناسب  
 بودند.

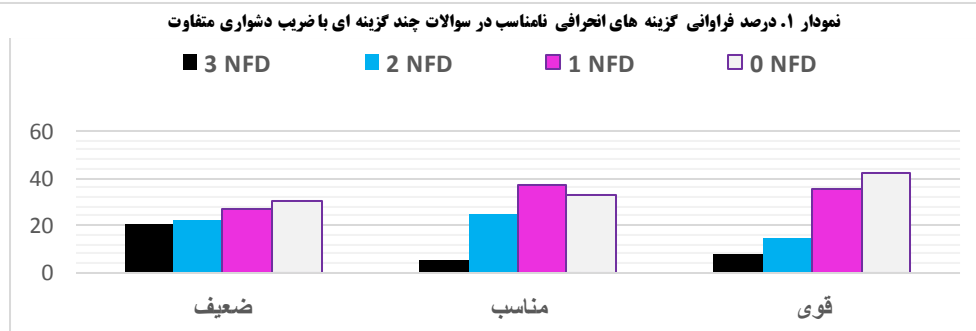
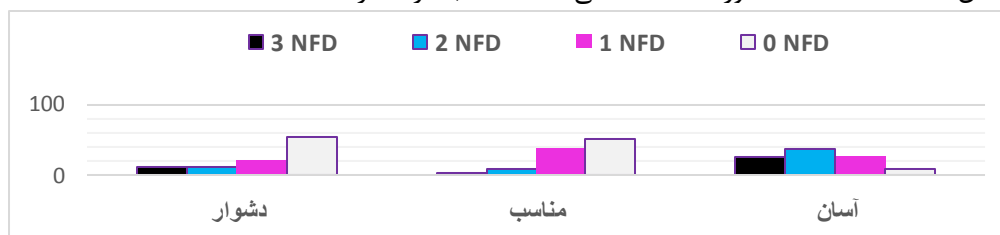
ضریب همبستگی پیرسون بین ضریب دشواری  
 سوالات و تعداد گزینه های انحرافی نامناسب (NFDs)

جدول شماره ۱. توزیع فراوانی ضریب دشواری و ضریب تمیز سوالات چند گزینه ای

ضریب تأثیر گزینه های انحرافی (DE)	درصد	تعداد	تعریف	مقادیر ضریب دشواری
۷۲/۲۵	۱۴/۷	۱۱۹	مشکل	$< 0/30$
۷۸/۸۸	۴۶/۲	۳۷۳	مناسب	$0/30 - 0/70$
۴۰/۴۷	۳۹/۱	۳۱۶	آسان	$> 0/70$
مقادیر ضریب تمیز				
۵۵/۶۲	۴۲/۷	۳۴۵	ضعیف	$< 0/20$
۶۵/۸۳	۲۵/۶	۲۰۷	مناسب	$0/20 - 0/34$
۷۰/۲۸	۳۱/۷	۲۵۶	قوی	$> 0/35$

شود تعداد سوالات با گزینه های انحرافی نامناسب تر  
 در سوالات آسان بیشتر و در سوالات با درجه تمیز  
 بالاتر کمتر است.

در نمودارهای شماره ۱ و ۲ توزیع فراوانی گزینه های  
 انحرافی در سوالات با درجه دشواری و درجه تمیز  
 مختلف نشان داده شده است. همانطور که مشاهده می



توزیع فراوانی چینش پاسخ های درست بین گزینه  
 های انحرافی در سوالات مورد بررسی وضعیت نرمال  
 نداشت. درصد فراوانی پاسخ درست در ۲۲/۴ درصد  
 سوالات گزینه "الف"، در ۳۲/۷ درصد گزینه "ب" در

همکاران (۲۰۱۷) در مطالعه ای با بررسی ۲۰ سوال چند گزینه ای گزارش کرده اند که ۱۵ سوال (۷۵٪) سوالات دارای ضریب دشواری مناسب می باشند که مقادیر گزارش شده در مقایسه با این مطالعه بالاتر است. بطور کلی یک آزمون چند گزینه ای باید به نسبت مناسب هم سوال مشکل داشته باشد و هم سوال آسان. ولی تعداد زیاد سوالات ساده آزمون های یک مدرس ممکن است موجب تلاش کمتر دانشجویان برای یادگیری محتوای درسی بشود و این موضوع می تواند منجر به ارتقاء نامناسب دانشجویان شود و این موضوع می تواند برای رشته های علوم پزشکی که در آینده عهده دار امر بهداشت و درمان هستند نتایج زیانباری را متوجه جامعه کند. در مجموع یافته های این مطالعه همچون مطالعات انجام شده در سایر مناطق نشان می دهد که سوالات آزمون های تهیه شده در بیشتر دانشگاهها کم و بیش از اعتبار و کفایت لازم برخوردار نبوده و نیازمند تحلیل و پایش مستمر می باشند. تشکیل بانک سوالات در گروههای آموزشی از طریق تحلیل ادواری سوالات و حذف سوالات نامناسب راهکار مناسبی در راستای بهبود کیفیت آزمون ها خواهد بود.

در بررسی کیفیت سوالات چند گزینه ای علاوه بر سطح دشواری سوالات، شاخص تمیز نیز نقش مهمی دارد و می تواند کمک بزرگی برای طراحان سوالات باشد. در این مطالعه کمتر از نیمی از سوالات (۴۲/۷٪)، قدرت تمیز پایین یا ضعیف داشتند. اگر قدرت تمیز سوال پایین تر از ۰/۲ و یا منفی باشد، مفهوم آن این است که افراد ضعیف بهتر از افراد قوی توانسته اند به آن سوال پاسخ بدهند و احتمالاً آن سوال اشکال فنی و یا ساختاری داشته است و باید مورد تجدید نظر قرار گرفته، اصلاح شده و یا در آزمون های بعدی کنار گذاشته شود (پاتل ۲۰۱۷). بررسی شاخص های کمی سوالات چند گزینه ای در دانشگاههای علوم پزشکی قزوین، کاشان و گلستان نشان داده است که بیش از نیمی از سوالات طراحی شده از قدرت تمیز مناسبی برخوردار نبودند و علت آن عدم آگاهی کافی اساتید و شیوه نامناسب طراحی سوالات معرفی شده بود

۲۱/۹ درصد گزینه "ج" و در ۲۳ درصد گزینه "د" بود. مقایسه توزیع فراوانی گزینه درست در محل های مختلف (گزینه های الف، ب، ج و د) نشان داد که فراوانی قرار گرفتن پاسخ درست در گزینه "ب" بطور معنی داری از گزینه های دیگر بیشتر است ( $p=0/001$ ).

### بحث

یافته های این مطالعه اطلاعات ارزشمندی در ارتباط با کیفیت سوالات چند گزینه ای گروه ایمنی شناسی فراهم نمود. آزمون های چند گزینه ای یکی از روش های موثر در ارزشیابی دانشجویان می باشند؛ اما این موثر بودن به طراحی صحیح و بهبود مستمر این سوالات از طریق تحلیل آزمون بستگی دارد. این فرآیند میتواند منجر به بهبود کیفیت سوالات و افزایش مهارت مدرسین در آزمون های بعدی شود. (رید و همکاران ۲۰۰۷، شکورنیا و همکاران ۲۰۱۰b، یوسف و تایب ۲۰۱۴).

یافته های این مطالعه نشان داد در مجموع بیش از ۵۰ درصد سوالات مورد بررسی درجه دشواری خیلی ساده و یا خیلی مشکل دارند و در محدوده سوالات مناسب و با ضریب دشواری ۰/۳ تا ۰/۷ قرار نمی گیرند. تشنیزی و همکاران نیز در بررسی سوالات آزمون های چند گزینه ای دانشگاه علوم پزشکی هرمزگان گزارش نمودند که حدود ۵۰ درصد سوالات دارای درجه دشواری مناسب بودند (حسینی تشنیزی و همکاران ۲۰۱۰). بررسی سوالات چند گزینه ای آزمون دستیاری در دانشگاه علوم پزشکی قزوین نشان داد که کمتر از دو سوم سوالات آزمون ها درجه دشواری مناسب داشتند (درخشان و همکاران ۲۰۱۶). نتایج مطالعات مشابه در هند و عربستان نیز درجه دشواری مناسب سوالات چند گزینه ای را در حد ۵۰ درصد گزارش کرده اند (الفاکی و همکاران ۲۰۱۵، گاجر و همکاران ۲۰۱۴). که تا حدودی با یافته های این مطالعه همخوانی دارد. در یافته های برخی مطالعات دیگر فراوانی سوالات با ضریب دشواری مناسب بالاتر گزارش گردیده است که می تواند بیانگر مهارت بیشتر و کیفیت بهتر سوالات آن آزمون ها بوده باشد. منون و

که بالاتر از مقادیر مطالعه ما می باشد و نشاندهنده مهارت بالاتر مدرسین در تهیه چند گزینه ای می باشد (هینگرجو و جلیل ۲۰۱۲).

در طراحی سوال چند گزینه ای ابتدا ریشه سؤال، بعد پاسخ درست سوال (گزینه صحیح) و در نهایت تعدادی پاسخ نادرست (گزینه های انحرافی) نوشته می شوند. مشکل ترین بخش تهیه سوالات چند گزینه ای طراحی گزینه های انحرافی مناسب می باشد. گزینه های انحرافی نقش مهمی در فرایند طراحی و کیفیت سوالات چند گزینه ای دارند و اگر به درستی تهیه شوند اعتبار این آزمون ها را افزایش می دهند. از طرف دیگر گزینه های انحرافی نامناسب منجر به افت کیفیت سوالات شده زیرا قادر به تمایز بین دانشجویان با عملکرد تحصیلی متفاوت نمی باشند. اگر گزینه های انحرافی با توجه به اشتباهات و کج فهمی های متداول دانشجویان تهیه شوند، منبع بسیار مناسبی برای تشخیص مشکلات دانشجویان خواهند بود. تهیه و طراحی گزینه های انحرافی در سوالات چند گزینه ای کار مشکلی است. کیفیت یک سؤال چند گزینه ای به شیوایی گزینه های انحرافی آن بستگی دارد (فو همکاران ۲۰۱۵، مکرچیان ۲۰۱۱).

در این بررسی از تعداد ۲۴۲۴ گزینه انحرافی، ۱۵۲۵ (۶۲/۹۰٪) مناسب (FD) و تعداد ۸۹۹ (۳۷/۱۰٪) نامناسب (NFD) بودند. یعنی بیش از یک سوم گزینه های انحرافی نامناسب بودند. بررسی گزینه های انحرافی سوالات چند گزینه ای در مطالعه مشابهی در کشور هند نیز نشان داد که ۳۵/۳٪ گزینه های انحرافی در سوالات چند گزینه ای نامناسب بودند که تا حدود زیادی با نتایج این مطالعه همخوانی دارد (گاجر و همکاران ۲۰۱۴). در مطالعات مشابه انجام در مناطق مختلف نتایج متفاوتی گزارش گردیده است که می تواند به دلیل تفاوت در جامعه و حجم نمونه مورد بررسی بوده باشد. مطالعه انجام شده توسط نمودو (Namdeo) و همکاران از ۷۵ سوال مورد بررسی تعداد ۳۵ سوال (۴۶/۴٪) دارای گزینه انحرافی مناسب و ۴۰ سوال (۵۳/۴٪) دارای گزینه انحرافی نامناسب بودند (نمودو و روت ۲۰۱۶). در مطالعه دیگری توسط دوسا و همکاران بررسی ۴۸ سوال چند گزینه ای نشان داد که حدود دو سوم (۷۴/۳٪) گزینه های انحرافی در

(درخشان ۲۰۱۶، وکیلی ۲۰۰۷، ساناجو و همکاران ۲۰۱۰).

یافته های این مطالعه همچنین نشان داد که بین دو شاخص ضریب دشواری و ضریب تمیز رابطه خطی معکوسی وجود دارد بنحوی که با افزایش ضریب دشواری (آسان تر شدن سوال) ضریب تمیز سوال کمتر می شود. پورمیرزا در کرمانشاه، مارتا در مالزی و پاندی در اندونزی نیز در مطالعات خود بین این دو شاخص یک رابطه خطی معکوس گزارش کرده اند که با یافته های این مطالعه همخوانی دارد (مارتا و همکاران ۲۰۰۹، پاندی و همکاران ۲۰۱۳، پورمیرزا و همکاران ۲۰۱۵). بر این اساس می توان بیان داشت که با طراحی سوالات آسان، قدرت آزمون ها در تفکیک دانشجویان با عملکرد تحصیلی بالا از دانشجویان با عملکرد تحصیلی پایین که به طور شانس به سوالات پاسخ می دهند کاهش می یابد. نتایج یک مطالعه به ارتباط ضعیف ضریب دشواری و ضریب تمیز و قرار گرفتن بهترین ضریب تمیز در محدوده ۰/۰-۶/۸ یعنی سوالات دارای درجه دشواری متوسط اشاره نموده است و بیان داشته است که این رابطه منعکس کننده این اصل است که طراحان سوالات چند گزینه ای علاوه بر درجه دشواری سوال باید به قدرت تمیز آن نیز توجه کنند زیرا نمی توان به صرف طراحی کردن سوالات دشوار قدرت تمیز آزمون را افزایش داد (پاندی و همکاران ۲۰۱۳). هینگرجو (Hingorjo) معتقد است که یک آزمون چند گزینه ای ایده آل باید دارای درجه دشواری متوسط، ضریب تمیز بالا و گزینه های انحرافی مناسب (functioning distractors) کمتر از ۳ باشد (هینگرجو و جلیل ۲۰۱۲).

یافته های این مطالعه نشان داد که تعداد ۲۴۸ سوال (۳۰/۷ درصد) از سوالات چند گزینه ای مورد بررسی ایده آل می باشند و ضریب دشواری بین ۰/۳ و ۰/۷؛ و ضریب تمیز بالاتر از ۰/۲۴ دارند. گاجر و همکاران (۲۰۱۴) و مهتا و همکاران (۲۰۱۴) درصد سوالات چند گزینه ای ایده آل را به ترتیب ۳۲ و ۲۴ درصد گزارش کرده اند که تا حدودی با یافته های این مطالعه همخوانی دارد (گاجر و همکاران ۲۰۱۴، مهتا و مخاسی ۲۰۱۴). اما در مطالعه دیگری در پاکستان فراوانی سوالات ایده آل ۶۴ درصد گزارش گردیده است

ترتیب ۷۰/۲۸ درصد و ۵۵/۶۲ درصد بود که با نتایج دیگر مطالعات انجام شده همخوانی دارد (مهتا و مخاسی ۲۰۱۴، مخرجی و لاهییری ۲۰۱۵).

تحلیل گزینه های انحرافی در سوالات چند گزینه ای و تهیه گزینه های انحرافی مناسب می تواند درجه دشواری و تمیز سوالات چند گزینه ای را افزایش داده و باعث بهبود کیفیت آزمون های چند گزینه ای شود. بنابراین باید در پایش سوالات یک آزمون در کنار پایایی آزمون، ضریب دشواری و ضریب تمیز به گزینه های انحرافی نیز اهمیت داده شود.

یکی از نگرانی هایی که همواره در مورد سوالات چند گزینه ای وجود داشته و به عنوان یکی از معایب این آزمون ها مطرح گردیده پاسخ حدسی به سوالات است بنحوی که گاهی آزمون شوندگان با دانش کم با حدس به پاسخ درست می رسند و این امر باعث کاهش اعتبار این آزمون ها می شود. راهکارهای مختلفی برای رهایی از این مشکل ارائه شده است. محل قرار گرفتن گزینه درست و توزیع متوازن آن بین گزینه های انحرافی از مواردی است که در طراحی این آزمون ها مورد تاکید قرار گرفته است (کارنجی ۲۰۱۷، اسنل و همکاران ۲۰۱۵، مرتاض هجری و همکاران ۲۰۱۴). نتایج یک مطالعه نشان داده است که محل قرار گرفتن گزینه درست در سوالات چند گزینه ای روی ضریب دشواری سوال تاثیر دارد (حسینی و بقایی ۲۰۱۷). کارشناسان توصیه می کنند که برای جلوگیری از حدس کورکورانه پاسخ سوالات، محل قرار گرفتن گزینه درست بایستی بطور تصادفی انتخاب شده و بین گزینه های مختلف (الف، ب، ج و د) توزیع متعادل وجود داشته باشد. بنحوی که توزیع پاسخ درست در این آزمون ها در هر یک از گزینه های "الف، ب، ج و د" بطور مساوی ۲۵٪ باشد (تارنت و وار ۲۰۱۲). مقایسه توزیع فراوانی پاسخ درست در گزینه های مختلف در این مطالعه تفاوت معنی داری نشان داد بنحوی که حدود یک سوم پاسخ های درست (۳۲/۷٪) در گزینه "ب" قرار گرفته بود. این یافته با نتایج مطالعه ای که در دانشگاه علوم پزشکی جندی شاپور اهواز توسط شکورنیا انجام شد همخوانی دارد (شکورنیا و الهام پور ۲۰۱۵).

سوالات مناسب و ۲۵/۷٪ نامناسب بودند (دوسا و همکاران ۲۰۱۷).

این مطالعه نتایج بررسی های قبلی در ارتباط با تاثیر گزینه های انحرافی نامناسب بر ضریب دشواری سوالات چند گزینه ای را تایید کرد (تارنت و وار ۲۰۱۲). یافته های این مطالعه هماهنگ با نتایج مطالعات مشابه انجام شده در پاکستان و عربستان نشان داد که با افزایش تعداد گزینه های انحرافی نامناسب درجه دشواری سوالات چند گزینه ای کاهش می یابد (هینگرجو و جلیل ۲۰۱۲، عبدالغنی و همکاران ۲۰۱۴a). در مطالعه دیگری نیز نشان داده شده است که سادگی سوالات چند گزینه ای می تواند روی قدرت تمیز این سوالات تاثیر گذار باشد (عبدالغنی و همکاران ۲۰۱۴b).

بطور کلی یافته های این مطالعه نشان می دهد که افزایش تعداد گزینه های انحرافی نامناسب (NFDS) روی کیفیت سوالات چند گزینه ای تاثیر می گذارد. با افزایش تعداد گزینه های انحرافی نامناسب (NFDS) ضریب دشواری سوال افزایش می یابد (سوال آسان می شود) و ضریب تاثیر گزینه های انحرافی (DE) کاهش می یابد؛ برعکس با افزایش تعداد گزینه های انحرافی مناسب (FDS) در سوالات چند گزینه ای ضریب دشواری سوال کاهش می یابد (سوال مشکل می شود) و ضریب تاثیر گزینه های انحرافی افزایش می یابد. در این مطالعه میانگین ضریب تاثیر گزینه های انحرافی در سوالات مشکل ۷۲/۲۵ درصد و در سوالات آسان ۴۰/۴۷ درصد بود که با نتایج مطالعه مخرجی و همکاران که ضریب تاثیر گزینه های انحرافی را در سوالات مشکل و آسان به ترتیب ۸۳/۳۴ درصد و ۲۲/۲۲ درصد گزارش کرده اند همخوانی دارد (مخرجی و لاهییری ۲۰۱۵).

یافته های مطالعه همچنین نشان داد که تعداد گزینه های انحرافی نامناسب (NFDS) روی ضریب تمیز سوالات تاثیر دارد بنحوی که با افزایش تعداد گزینه های انحرافی نامناسب (NFDS) ضریب تمیز سوال کاهش می یابد و قدرت تمیز سوال کم می شود. ضریب تاثیر گزینه های انحرافی در سوالات قوی و ضعیف به

*Journal of Health Specialties*, Vol.2, No.4, Pp. 148-157.

Abdulghani, HM, Ponnampereuma, GG, Ahmad, F, 2014b, A comprehensive, multi-modal evaluation of the assessment system of an undergraduate research methodology course: Translating theory into practice, *Pakistan journal of medical sciences*, Vol.30, No.2, Pp 227-239.

Abozaid, H, Park, YS, Tekian, A, 2017, Peer review improves psychometric characteristics of multiple choice questions, *Medical teacher*, Vol.39, sup1, Pp. S50-S4.

Amin, MM, Shayan, S, Hashemi, H, et al, 2011, Analysis of multiple choice questions based on classical test theory, *Iranian Journal of Medical Education*, Vol.10, No.5, Pp.719-725. [In Persian]

Carnegie, JA, 2017, Does correct answer distribution influence student choices when writing multiple choice examinations? *The Canadian Journal for the Scholarship of Teaching and Learning*, Vol.8, No.1, Pp.11-17.

Derakhshan, F, Ahmady, S, Allami, A, 2016, Quantitative and qualitative indicators evaluation of residency exams in Qazvin University of Medical Sciences (2012-13). *Journal of medical education development*, Vol.8, No.20, Pp.25-32. [In Persian]

Derakhshan, F, Allami, A, Ahmadi, S, 2015, Effect of faculty training programs on improving quality of residency exams in 2013-2014. *Research in Medical Education*, Vol.7, No.1, Pp.19-26. (Persian)

Elfaki, OA, Bahamdan, KA, Al-Humayed, S, 2015, Evaluating the quality of multiple-choice questions used for final exams at the Department of Internal Medicine, College of Medicine, King Khalid University. *Sudan Medical Monitor*, Vol.10, No.4, Pp.123-134.

با توجه به یافته‌های این مطالعه و مطالعات مشابه مبنی بر عدم کفایت لازم در سوالات چند گزینه‌ای آزمون‌های برگزار شده و همچنین اهمیت و نقش ارزشیابی در فرایند آموزش و لزوم آگاهی و توانمندی اساتید دانشگاهها و بهبود عملکرد آنان در ارزیابی موثر فراگیران، برگزاری دوره‌های آموزشی مستمر و پیش‌منظم سوالات با هدف ارتقا و بهبود کیفیت آزمون‌ها پیشنهاد می‌گردد. برای تحقق این هدف برگزاری کارگاههای طرح سوال و ایجاد بانک سوالات استاندارد ضروری است. همچنین شایسته است که مراکز مطالعات و توسعه آموزش پزشکی با بررسی ادواری سوالات چند گزینه‌ای اساتید و ارائه بازخوردهای مناسب به بهبود کیفیت سوالات چند گزینه‌ای کمک کنند.

### نتیجه گیری

نتایج این مطالعه نشان داد که شاخص‌های کمی سوالات مورد بررسی از کفایت لازم برخوردار نمی‌باشد. ضریب دشواری و تمیز آزمون‌ها ضعیف بوده و ضریب تاثیر گزینه‌های انحرافی پایین است بنحوی که بیش از یک سوم سوالات دارای گزینه انحرافی نامناسب می‌باشند. همچنین بین توزیع پاسخ درست در بین گزینه‌های انحرافی تعادل مطلوبی وجود ندارد و در مجموع سوالات آزمون‌ها از کیفیت لازم برخوردار نبوده و نیازمند بازبینی و اصلاح می‌باشند. از آنجایی که تحلیل سوالات آزمون‌ها اهمیت زیادی در ارتقای کیفیت آزمون‌ها دارد، نویسندگان امیدوارند یافته‌های این مطالعه به عنوان الگویی مناسب توجه اساتید دانشگاهها را به اهمیت فرایند تحلیل آزمون‌ها جلب نموده و منجر به ارتقای کیفیت و بهبود شرایط آزمون‌ها گردد.

جهت انجام این مطالعه مجوز لازم از کمیته اخلاق در پژوهش‌های زیست پزشکی دانشگاه علوم پزشکی جندی شاپور اهواز با کد IR.AJUMS.REC.1397.609 دریافت شد.

### References

Abdulghani, HM, Ahmad, F, Ponnampereuma, GG, et al 2014a, The relationship between non-functioning distractors and item difficulty of multiple choice questions: A descriptive analysis.



- of Nursing Education*, Vol.9, No. 3, Pp 109-114.
- McCoubrie, P, 2004, Improving the fairness of multiple-choice questions: a literature review. *Medical teacher*, Vol.26, No.8, Pp.709-712.
- Mehta, G, Mokhasi, V, 2014, Item analysis of multiple choice question-an assessment of the assessment tool. *Int J Health Sci Res*, Vol.4, No. 7, Pp.197-202
- Menon AR, Kannambra PN, 2017, Item analysis to identify quality Multiple Choice Questions. *National Journal of Laboratory Medicine*, Vol.6, No. 2, Pp. 7-10
- Meyari, A, Beiglarkhani, M, 2012, Improvement of design of multiple choice questions in annual residency exams by giving feedback item analysis to identify quality multiple choice questions. *Strides in Development of Medical Education*, Vol.10, No.1, Pp.109-118. [In Persian]
- Mitra, N, Nagaraja, H, Ponnudurai, G, et al, 2009, The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester 1, multidisciplinary summative tests. *IeJSME*, Vol.3, No.1, Pp.2-7.
- Mkrtchyan, A, 2011, Distractor quality analyze in multiple choice uestions based on information retrieval model. *EDULEARN Proceedings*, PP.1624-1631.
- Mortaz Hejri, S, Khabaz Mafinezhad, M, Jalili, M, 2014, Guessing in Multiple Choice Questions: Challenges and Strategies. *Iranian Journal of Medical Education*, Vol.14, No.7, Pp.594-604. (Persian)
- Mukherjee, P, Lahiri, S, 2015, Analysis of multiple choice questions (MCQs): Item and test statistics from an assessment in a medical college of Kolkata, West Bengal. *IOSR J Dent Med Sci*, Vol.1, Pp.47-52.
- Gajjar, S, Sharma, R, Kumar, P, et al, 2014, Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, Vol.39, No.1, Pp.17-20.
- Hingorjo, MR, Jaleel, F, 2012, Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *JPMMA-Journal of the Pakistan Medical Association*, Vol.62, No.2, Pp.142-149.
- Hohensinn, C, Baghaei, P, 2017, Does the position of response options in multiple-choice tests matter? *Psicológica*, Vol.38, No.1, Pp.19-24.
- Hosseini Teshnizi, S, Zare, S, Solati, S, 2010, Quality analysis of multiple choice questions (MCQs) examinations of noncontinuous undergraduate medical records. *Bimonthly Journal of Hormozgan University of Medical Sciences*, Vol.14, No.3, Pp.177-183. [In Persian]
- Karkal, YR, Kundapur, GS, 2016, Item analysis of multiple choice questions of undergraduate pharmacology examinations in an International Medical School in India. *Journal of Dr NTR University of Health Sciences*, Vol.5, No.3, Pp.183-186.
- Kouhpayezadeh, J, Dargahi, H, Arabshahi, KS, 2012, Clinical assessment methods in medical sciences universities of Tehran-clinical instructors' viewpoint. *Bimonthly Journal of Hormozgan University of Medical Sciences*, Vol.16, No.5, Pp.395-402. (Persian)
- D'Sa JL, Visbal Dionaldo ML, 2017, Analysis of multiple choice questions: item difficulty, discrimination index and distractor efficiency. *International Journal*

- year two medical students. *Medical education*, Vol.41, No.8, Pp.754-762.
- Sanagoo, A, Jouybari, L, Ghanbari Gorji, M, 2010, Quantitative and qualitative analysis of academic achievement tests in Golestan University of Medical Sciences. *Research in Medical Education*, Vol.2, No.2, Pp.24-32.
- Şenel, S, Pehlivan, EB, Alatlı, B, 2015, Effect of Correction-for-Guessing Formula on Psychometric Characteristics of Test. *Procedia-Social and Behavioral Sciences*, Vol.191, Pp.925-929.
- Shakurnia, A, Khosravi, BA, Mozafari, A, 2010a, Elhampour H. An evaluation of exam questions designed by faculty members, emphasizing on the multiple choice question structure, Ahvaz Jundishapur University of Medical Sciences, 2007. *Strides Dev Med Educ*, Vol.6, No.2 Pp. 129-138. [In Persian]
- Shakurnia A, Mozafari, A, Khosravi, BA, 2010b. Survey on structural of MCQs of residency exam in *AJUMS*, Vol.8, No.4, Pp.491-502. [In Persian]
- Shakurnia, A, Elhampour, H, 2015, Survey on correct response position in multiple choice tests: A descriptive study. *Strides in Development of Medical Education*, Vol.12, No.2, Pp.426-432. [In Persian]
- Tarrant, M, Ware, J, Mohammed, AM, 2009, An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Medical Education*, Vol.9, No.1, Pp.40-49.
- Tarrant, M, Ware, J, 2012, A framework for improving the quality of multiple-choice assessments. *Nurse Educator*, Vol.37, No.3, Pp.98-104.
- Vakili, Z, Fakharian, E, Rasooli nejad, SA, Et al, 2007, The psychometric properties of Namdeo, SK, Rout, SD, 2016, Assessment of functional and nonfunctional distracter in an item analysis. *International Journal of Contemporary Medical Research*, Vol.3, No.7, Pp.1891-1893.
- Pande, SS, Pande, SR, Parate, VR, et al, 2013, Correlation between difficulty & discrimination indices of MCQs in formative exam in Physiology. *South-East Asian Journal of Medical Education*, Vol. 7, No.1, Pp.45-50.
- Patel, RM, 2017, Use of item analysis to improve quality of multiple choice questions in II MBBS. *Journal of Education Technology in Health Sciences*, Vol.4, No.1< Pp.22-29.
- Patil, R, Palve, SB, Vell, K, et al, 2017, Evaluation of multiple choice questions by item analysis in a medical college at Pondicherry, India. *International Journal of Community Medicine and Public Health*, Vol.3, No.6, Pp.1612-1616.
- Pho, V-M, Ligozat, A-L, Grau, B, 2015, Distractor quality evaluation in multiple choice questions. *International Conference on Artificial Intelligence in Education AIED 2015: Artificial Intelligence in Education* pp 377-386.
- Pourmirza, KR, Rezaie, M, Shojee M, 2015, Correlation of quality and quantity index of multiple choice questions exams of residency promotion in Kermanshah University of medical sciences, 2013. Vol. 4, No. 1, Pp.71-78. [In Persian]
- Quaigrain, K, Arhin, AK, 2017, Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, Vol.4, No.1, Pp.1-11.
- Reid, WA, Duvall, E, Evans, P, 2007, Relationship between assessment results and approaches to learning and studying in

*Medical education online*, Vol.18, No.1, Pp. 1-5.

Yusoff, MSB, Taib, F, 2014, Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. *Journal of Taibah University Medical Sciences*, Vol.9, No.2, Pp.109-114.

the multiple-choice tests residency, Kashan University of Medical. *Strides in Development of Medical Education 8th National Congress of Medical Education*, Pp.174-75.

Vanderbilt, A, Feldman, M, Wood, I, 2013, Assessment in undergraduate medical education: a review of course exams.

## Analytical Study of Quantitative Indices of Multiple-choice Questions of Immunology Department in Ahvaz Jundishapur University of Medical Sciences

Shakurnia Abdolhussein<sup>1\*</sup>

Ghafourian Mehri<sup>2</sup>

Khodadadi Ali<sup>3</sup>

Ghadiri Ata<sup>4</sup>

Amari Afshin<sup>4</sup>

<sup>1</sup>: Assistant, School of medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

<sup>2</sup>: Professor, School of medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

<sup>3</sup>: Associated Professor, School of medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

<sup>4</sup>: Assistant, School of medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

**Abstract:** Considering the widespread use of MCQs in assessment of medical students, present study was conducted to item analyze of MCQs in immunology department at Ahvaz Jundishapur University of Medical Sciences in 2017. In this descriptive cross-sectional study, the data of MCQs designed by immunological faculties that corrected by the OP-Scan was assessed and explored the difficulty index, discrimination index and distractor options (808 MCQs). Then, data were analyzed using descriptive statistics and Pearson coefficient. The results showed the average of difficulty index of the MCQs was  $0.59 \pm 0.25$ , and 46.2% of the MCQs had a suitable difficulty. The average of the discrimination index of the MCQs was  $0.25 \pm 0.24$  and 57.3% of the MCQs had a discrimination index. Accordingly, a Combination of the two difficulty and discrimination indices showed that only 248 MCQs (30.7%) were ideal. 1525 distractor option (62.9%) were functional distractor (FD) and 889 (37%) were non-functional distractor (NFDs). The findings showed that the MCQs required to be reviewed and improved.

**Keywords:** MCQs, Item analysis, difficulty index, discrimination index, distractor options.

**\*Corresponding author:** Assistant, School of medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

**Email:** shakurnia@yahoo.com