



## بررسی ویژگی‌های روانسنجی سؤالات آزمون دستیاری پزشکی و دندانپزشکی سال ۱۳۹۹، بر اساس نظریه پرسش-پاسخ

سیده هدی ناجی<sup>۱</sup>، علی مقدم‌زاده<sup>۲</sup>، سمانه پنجه‌علی بیگ<sup>۳</sup>

۱. دانشجوی دکتری گروه روش‌ها و برنامه‌های درسی و آموزشی دانشکده روانشناسی و علوم تربیتی دانشگاه تهران، تهران، ایران؛ (نویسنده مسئول)، پست الکترونیک: s.hodanaji@ut.ac.ir
۲. دانشیار گروه روش‌ها و برنامه‌های درسی و آموزشی دانشکده روانشناسی و علوم تربیتی دانشگاه تهران، پست الکترونیک: amoghadamzadeh@ut.ac.ir
۳. مسئول طرح برنامه و آموزش مرکز سنجش آموزش پزشکی تهران، تهران، ایران، پست الکترونیک: s.alibeik@sanjeshp.ir

اطلاعات مقاله	چکیده
نوع مقاله:	هدف: هدف از پژوهش حاضر، بررسی کیفیت آزمون‌های دستیاری پزشکی و دندانپزشکی بر اساس ویژگی‌های روانسنجی سؤالات و با استفاده از نظریه پرسش-پاسخ است.
مقاله پژوهشی	روش پژوهش: روش پژوهش از نوع توصیفی-تحلیلی است. داده‌های مورد تحلیل شامل پاسخنامه تمام افراد شرکت کننده در آزمون دستیاری سال ۱۳۹۹ است که بر اساس نظریه پرسش-پاسخ مورد تحلیل قرار گرفته است.
دریافت:	یافته‌ها: یافته‌ها نشان می‌دهد آزمون‌های دستیاری دندانپزشکی و پزشکی از نظر ویژگی‌های روانسنجی در وضعیت مطلوبی نیست و لازم است طراحان در تدوین سؤالات بیشتر دقت کنند.
۱۴۰۰/۱۲/۲۲	نتیجه‌گیری: برای افزایش کیفیت آزمون‌های دستیاری تخصصی، علاوه بر رعایت تاکسونومی سؤالات، لازم است طراحان پارامتر تشخیص و دشواری سؤال و نیز کیفیت گزینه‌های انحرافی، همسانی سؤالات با یکدیگر و با موضوع درس مربوط را نیز مد نظر قرار دهند تا از این راه بتوان کیفیت آزمون‌های دستیاری تخصصی را افزایش داد.
اصلاح:	واژه‌های کلیدی: آزمون دستیاری، پزشکی، دندانپزشکی، نظریه پرسش-پاسخ، روانسنجی
۱۴۰۱/۰۶/۱۳	
پذیرش:	
۱۴۰۱/۰۶/۲۲	
انتشار:	
۱۴۰۱/۰۶/۲۳	

استناد: ناجی، سیده هدی، مقدم‌زاده، علی، پنجه‌علی بیگ، سمانه (۱۴۰۱). بررسی ویژگی‌های روانسنجی سؤالات آزمون دستیاری پزشکی و دندانپزشکی سال ۱۳۹۹، بر اساس نظریه پرسش-پاسخ. فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی، ۱۲(۳۸)، ۶۸-۵۵ صفحه، DOI: 10.22034/EMES.2022.550571.2365

ناشر: سازمان سنجش آموزش کشور حق مؤلف © نویسندگان.



## مقدمه

برنامه‌های دستیاری پزشکی و دندان پزشکی در سراسر جهان به منظور تربیت پزشکانی با مهارت‌های بالا که بتوانند بهترین خدمات بهداشتی را ارائه دهند برگزار می‌شود. دوره دستیاری، زمانی برای آموزش و تربیت دانشجویان پزشکی است که در طی آن دانشجویان به پزشکی توانا تبدیل شوند (کاسونه و همکاران، ۲۰۲۰ و توماس و همکاران، ۲۰۲۰). آموزش تخصصی و فوق تخصصی پزشکی بنابر نقش ویژه دستیاران در نظام آموزشی-درمانی دانشگاه‌ها و همچنین نقش بسیار مهم دانش‌آموختگان در نظام سلامت جامعه، در نظام آموزش عالی کشور اهمیت کلیدی دارد و از این رو تحول و ارتقای کیفیت آن از دغدغه‌های جدی مسئولان اجرایی و صاحب‌نظران است. در این میان مانند هر نظام آموزشی، ارزیابی فراگیران جزء مهمی از آموزش است و در میان اجزاء مختلف آموزش تخصصی، آزمون‌ها به عنوان یکی از بخش‌های تأثیرگذار و جهت‌دهنده به روند کلی آموزش جایگاه ویژه‌ای دارد (خوشرننگ و همکاران، ۱۳۹۲).

در کشورهای مختلف دنیا برای ورود به دوره دستیاری تخصصی، از آزمون‌های ورودی استفاده می‌شود. در ایران نیز یکی از آزمون‌های پرطرفدار که مرکز سنجش آموزش پزشکی، هر ساله برگزار می‌کند، آزمون دستیاری تخصصی پزشکی و دندانپزشکی است. کسانی که در این آزمون پذیرفته شوند، بعد از گذراندن دوره چهارساله دستیاری، می‌توانند با عنوان پزشک متخصص مشغول به فعالیت شوند. از این رو پیوستن به یک دوره دستیاری تخصصی و انتخاب مناسب‌ترین پزشکان برای این برنامه خاص، اهمیت بسیاری دارد. با توجه به اینکه کیفیت پذیرفته‌شدگان در دوره‌های آموزشی، بر کیفیت دوره‌های آموزشی اثرگذار است، تلاش می‌شود در فرآیند سنجش و پذیرش افراد، مناسب‌ترین‌ها وارد دوره‌های دستیاری شوند. با توجه به اینکه آزمون‌های دستیاری تخصصی به عنوان تنها ملاک سنجش و پذیرش متقاضیان محسوب می‌شود، کیفیت این آزمون‌ها تأثیر مستقیمی بر کیفیت ورودی‌ها به دوره دستیاری تخصصی و در نتیجه کیفیت دوره‌های آموزشی دارد (ناجی و همکاران، ۱۳۹۷ و خوشرننگ و همکاران، ۱۳۹۲).

آزمون‌های دستیاری تخصصی در ایران به صورت چهارگزینه‌ای تدوین و اجرا می‌شود. در خصوص آزمون‌های چندگزینه‌ای می‌توان گفت، با توجه به اینکه این نوع آزمون‌ها می‌تواند بخش وسیعی از دانش را در مدت کوتاهی ارزیابی کند، نسبت به حدس حساسیت کمی دارد و به آسانی قابل تصحیح است، بهترین نوع آزمون عینی به شمار می‌رود و در سنجش پزشکی بسیار کاربرد دارد. بر خلاف تصور عمومی درباره آزمون‌های چندگزینه‌ای، اگر این آزمون‌ها با دقت و به شکل اصولی تهیه شوند، علاوه بر اینکه می‌توانند سطوح بالایی شناختی را اندازه بگیرند، به خوبی می‌توانند افراد قوی را از ضعیف تشخیص دهند. برای آنکه بتوان آزمون با کیفیت تهیه کرد، لازم است سؤالات آزمون با کیفیت باشد. در منابع مختلف، ویژگی‌های هر سؤال خوب بیان شده است و به هر میزان که سؤال طبق این معیارها طرح شده باشد، سؤال بهتری خواهد بود. از جمله این معیارها، می‌توان به دشواری

سؤال، قدرت تشخیص و همخوانی سؤال با کل آزمون اشاره کرد. علاوه بر این موارد، ابعاد آزمون و پایایی آزمون نیز باید مورد توجه قرار گیرد.

شکورنیا و همکاران (۱۳۸۸) در پژوهش خود کیفیت آزمون‌ها را با توجه به سطوح شناختی اندازه‌گیری شده، براساس طبقه بندی بلوم، بررسی کرده‌اند. پژوهش آنها نشان داد که بیشتر سؤالات آزمون دستکاری، سطوح پایین شناختی را مورد سنجش قرار می‌دهد.

پژوهش خوشرنگ و همکاران (۱۳۹۲) نیز نشان داد که شرکت در کارگاه‌های آموزش طراحی سؤالات چند گزینه‌ای می‌تواند سبب افزایش کیفیت سؤالات از جنبه سطوح شناختی مورد اندازه‌گیری بر مبنای اصول میلمن<sup>۱</sup> شود. در این پژوهش نیز، کیفیت سؤالات، تنها از نظر سطوح شناختی مورد اندازه‌گیری، بررسی شده است.

مونتگومری<sup>۲</sup> و همکاران (۲۰۱۰) برای بررسی کیفیت آزمون بازنگری شده COPE که به عنوان آزمون تکوینی برای دستیاران روانپزشکی استفاده می‌شود، از نظریه کلاسیک اندازه‌گیری استفاده کرده‌اند. آنها برای بررسی آزمون از ضریب آلفای کرونباخ برای محاسبه پایایی، از آنووا برای بررسی روایی و برای بررسی ویژگی‌های هر سؤال از ضریب دشواری و ضریب تشخیص استفاده کرده است. در ضریب دشواری، مقدار  $0/25$  برای دشواری بالا (سؤال سخت) و مقدار  $0/95$  برای دشواری پایین (سؤال آسان) در نظر گرفته شد و برای محاسبه ضریب تشخیص، از همبستگی دورشته‌ای نقطه‌ای استفاده شده است. یافته‌های این پژوهش نشان می‌دهد که آزمون بازبینی شده قابلیت اطمینان و روایی مناسب و همچنین ضریب تشخیص مناسبی داشت.

یافته‌های پژوهش صبوری و همکاران (۱۳۸۱) که به بررسی ویژگی‌های روانسنجی بیست و نهمین آزمون دستکاری پرداختند، نشان می‌دهد با توجه به میزان اعتبار و خطای استاندارد اندازه‌گیری، آزمون به خوبی و به دقت توانسته است میزان دانش نظری شرکت‌کنندگان را ارزیابی کند و به عبارت دیگر، نتایج اجرای این آزمون (رتبه و نمره شرکت‌کنندگان) معتبر است. با توجه به نبودن قدرت تمیز منفی در بین سؤالات، شیوه طراحی سؤالات مناسب بوده است. درجه دشواری سؤال تا حدودی متناسب با ظرفیت پذیرش دستیار بود. صبوری پیشنهاد می‌دهد که با توجه به شرایط مطلوب، لازم است ضریب دشواری سؤالات کاهش یابد.

شرستا<sup>۳</sup> و همکاران (۲۰۱۹) آزمون قضاوت موقعیتی<sup>۴</sup> (SJT) را به عنوان آزمونی معتبر برای سنجش توانایی‌های شناختی در پذیرش داوطلبان تحصیلات تکمیلی پزشکی، بررسی کردند. این مطالعه مقطعی در آکادمی علوم بهداشتی پاتان (PAHS) در ماه مه ۲۰۱۷ انجام شد. نتایج پژوهش آنها نشان می‌دهد که آزمون قضاوت موقعیتی، ابزاری قابل اعتماد برای سنجش خصوصیات غیر دانشگاهی در آزمون تحصیلات تکمیلی پزشکی

1. Milmen  
2. Montgomery  
3. Shrestha  
4. Situational judgement test

است و دو ویژگی قابلیت اطمینان و همسانی درونی ابزار را می‌توان با افزایش تعداد سؤالات با کیفیت، بهبود بخشید.

اسمیت<sup>۱</sup> و همکاران (۲۰۲۰) آزمون رادیولوژی دستیاران را در دو مرحله بررسی کردند. گام اول این مطالعه در سال ۲۰۱۸ به عنوان مرحله آزمایشی و گام دوم یک سال بعد انجام شد. در هر دو گام، بوردهای تخصصی مربوط، تمام سؤالات آزمون را با استفاده از داده‌های آماری از قبیل تعداد سؤال و آزمون، ضریب دشواری سؤال و ضریب تمیز، ارزیابی کردند و سؤالات معیوب، شناسایی، اصلاح، حذف و یا جایگزین شدند. نتایج تحقیق تأثیر مثبت اصلاح را بر کیفیت آزمون نشان داد.

در پژوهشی دیگر (کوثر<sup>۲</sup> و همکاران، ۲۰۱۶) بررسی کیفیت آزمون فارماکولوژی، با انجام تحلیل سؤال بررسی شد. محققان این پژوهش نتیجه گرفتند که تحلیل سؤال ابزاری ارزشمند برای کمک به حفظ سؤالات با کیفیت و حذف سؤالات بی کیفیت و در نهایت افزایش مهارت در ساخت آزمون است. تحلیل سؤال همچنین مشخص می‌کند که کدام سؤال به شفافیت و تأکید بیشتری نیاز دارد.

اگرچه در ایران کیفیت آزمون‌های دستیاری تخصصی مورد پژوهش قرار گرفته است، اما در این پژوهش‌ها به ویژگی‌های روانسنجی سؤالات (از جمله دشواری و قدرت تشخیص) پرداخته نشده است و پژوهش منتشر شده ای در این باره وجود ندارد. در این قبیل پژوهش‌ها کیفیت آزمون‌ها صرفاً با توجه به طبقه‌بندی (تاکسونومی) و سطوح شناختی مورد اندازه‌گیری، بررسی شده است و این در حالی است که از نظر سنجش سطوح بالایی شناختی، ممکن است یک سؤال، به دلیل قدرت تشخیص پایین و یا سختی نامناسب، سؤال خوبی برای قرار گرفتن در آزمون‌های کلانی مانند آزمون ورودی دستیاری تخصصی نباشد. بنابر این لازم است علاوه بر بررسی طبقه‌بندی (تاکسونومی) سؤالات، ویژگی‌های روانسنجی سؤالات نیز بررسی شود. پژوهش حاضر با هدف بررسی کیفیت آزمون‌های دستیاری تخصصی بر اساس ویژگی‌های روانسنجی سؤالات آزمون‌های دستیاری تخصصی انجام شده است.

### روش پژوهش

روش پژوهش از نوع توصیفی-تحلیلی است. داده‌های مورد تحلیل شامل پاسخنامه تمام افراد شرکت کننده در آزمون دستیاری سال ۱۳۹۹ است که براساس نظریه پرسش-پاسخ مورد تحلیل قرار گرفته است. برای تحلیل داده‌ها از نرم‌افزار R و NOHARM استفاده شده است.

در نظریه پرسش-پاسخ، ویژگی سؤال به گونه‌ای برآورد می‌شود که ناپسته به گروه است. در این نظریه از معادله‌های پیچیده ریاضی برای برآورد پارامتر سؤال (سختی، تشخیص و حدس‌پذیری) و توانایی افراد استفاده می‌شود. از جمله نقاط قوت نظریه پرسش-پاسخ نسبت به نظریه کلاسیک، برآورد دقیق‌تر سطح دشواری و

1. Smith  
2. Kaur

تشخیص سؤالاتی است که هیچ یک از افراد به آن پاسخ درست نداده‌اند ( $p=0$ ) در نظریه کلاسیک) و یا همه افراد به آن جواب درست داده‌اند ( $q=1$ ) در نظریه کلاسیک) (همبلتون و همکاران، ۱۳۸۹/۱۹۹۱، امبرستون و همکاران، ۱۳۸۸/۲۰۰۰).

معادله اصلی نظریه پرسش-پاسخ به صورت زیر است:

$$P_i(\theta) = c + \frac{1 - c}{(1 + e^{-Da_i(\theta - b_i)})}$$

$b$  پارامتر دشواری،  $a$  پارامتر تشخیص و  $c$  حدس پذیری سؤال و  $\theta$  توانایی افراد،  $D$  مقدار ثابت و  $P_i(\theta)$  احتمال پاسخ صحیح به سؤال  $i$  است. تعریف پارامتر تشخیص و دشواری در نظریه پرسش-پاسخ، معادل با ضریب تشخیص و دشواری در نظریه کلاسیک است، اما حدس‌پذیری در نظریه پرسش-پاسخ یعنی احتمال انتخاب پاسخ صحیح توسط افراد با توانایی بسیار پایین. در نظریه پرسش-پاسخ، دامنه عملی دشواری از  $-4$  تا  $+4$  است و هر چه مقدار آن بیشتر باشد، یعنی سؤال سخت‌تر است. دامنه عملی پارامتر تشخیص از  $0$  تا  $+3$  است و هر چه مقدار آن بیشتر باشد، یعنی سؤال، توانایی بیشتری در تفکیک گروه قوی از ضعیف دارد. این نظریه با توجه به تعداد پارامترهای در نظر گرفته شده برای برآورد توانایی، مدل‌های مختلفی دارد که مشهورترین آنها مدل یک پارامتری، دو پارامتری و سه پارامتری است. در مدل یک پارامتری، برای تمام سؤالات فقط یک پارامتر تشخیص برآورد می‌شود ولی برای هر سؤال یک پارامتر دشواری جداگانه برآورد می‌شود. به عبارت دیگر پارامتر تشخیص برای تمام سؤالات برابر در نظر گرفته می‌شود اما در مدل دو پارامتری، علاوه بر پارامتر دشواری، برای هر سؤال یک پارامتر تشخیص برآورد می‌شود و در مدل سه پارامتری، علاوه بر پارامتر تشخیص و دشواری، برای هر سؤال پارامتر حدس نیز برآورد می‌شود (همبلتون و همکاران، ۱۳۸۹/۱۹۹۱، امبرستون و همکاران، ۱۳۸۸/۲۰۰۰).

نظریه پرسش-پاسخ برای محاسبه پایایی از فرمول خاص خود استفاده می‌کند. یکی از ضرایب پایایی مورد استفاده در این نظریه، پایایی تجربی است که نشان دهنده دقت آزمون در برآورد توانایی افراد است. مقدار پایایی تجربی از صفر تا یک است و هر چه مقدار آن به یک نزدیکتر باشد، دقت ابزار در برآورد توانایی افراد بیشتر است (دی‌آیالا، ۲۰۰۸).

از آنجا که نظریه پرسش-پاسخ ابطال‌پذیر است، برای انتخاب مدل مناسب برای تحلیل داده‌ها باید برازش مدل با داده را بررسی و براساس آن مدل مناسب را انتخاب کرد. برای برازش مدل از آماره  $\chi^2$  استفاده می‌شود:

$$\chi^2 = -2 \log \text{Likelihood} - (-2 \log \text{Likelihood})_{\text{مدل پیچیده تر}} - (-2 \log \text{Likelihood})_{\text{مدل ساده تر}}$$

علاوه بر ویژگی روانسنجی سؤالات، یکی دیگر از ملاک‌های بررسی کیفیت هر آزمون، مسئله بُعدیت است. آزمون‌های توانایی، معمولاً با هدف سنجش یک توانایی تهیه می‌شوند. ویژگی تک بعدی بودن به این معنا

است که یک توانایی غالب، عملکرد آزمودنی را تحت تأثیر قرار دهد. نقض مفروضه تک بُعدی بودن می‌تواند برآورد پارامترهای سؤال و توانایی را به طور جدی دچار تورش کند. علاوه بر این تک بُعدی بودن سؤال‌ها برای تفسیر روشن و بدون ابهام نمره‌ها و انجام مقایسه بین افراد مهم است (مینایی و فلسفی‌نژاد، ۱۳۸۹). در نظریه پرسش-پاسخ، پیش از تحلیل داده‌ها باید تک بُعدی بودن آزمون بررسی شود. برای بررسی تک بُعدی بودن از شاخص‌های برازش تاناکا و RMSE استفاده می‌شود.

### یافته‌ها

برای پاسخ به پرسش پژوهش، هر یک از سؤالات آزمون‌های دستیاری تخصصی پزشکی و دندانپزشکی بر مبنای نظریه پرسش-پاسخ تحلیل شد و پارامترهای سؤال، پایایی و بُعدیت آزمون بررسی شد. آزمون‌های دستیاری تخصصی از چند درس تشکیل شده و برای تحلیل روانسنجی سؤالات باید هر درس را به صورت جداگانه بررسی کرد. مشخصات توصیفی آزمون‌ها در جدول زیر آمده است.

جدول ۱. مشخصات توصیفی آزمون دستیاری پزشکی و دندانپزشکی

نام آزمون	تعداد درس‌ها	تعداد سؤالات	تعداد شرکت‌کنندگان
دستیاری دندانپزشکی	۱۳	۲۵۰	۱۹۶۱
دستیاری پزشکی	۱۷	۲۰۰	۱۱۶۵۵

ابتدا برای بررسی مفروضه تک بُعدی بودن، داده‌های هر درس به صورت جداگانه با استفاده از نرم‌افزار NOHARM بررسی شد. نمونه‌ای از خروجی نرم‌افزار مرتبط با درس ارتودانتیکس در زیر آمده است:

Tanaka index of goodness of fit = 0.9928357

Root mean square of residuals (lower off-diagonals) = 0.0058394

در برخی منابع مقدار قابل قبول برای شاخص تاناکا ۰/۹۰ بیان شده است و تاناکای بالاتر از ۰/۹۵ نشان دهنده مدل مناسب است. مقدار قابل قبول برای شاخص RMSE کمتر از ۰/۰۱ است (مینایی و فلسفی‌نژاد، ۱۳۸۹؛ شرملی انگل<sup>۱</sup> و همکاران، ۲۰۰۳).

مقادیر تاناکا و RMSE برای درس ارتودانتیکس نشان دهنده تک بُعدی بودن این درس است. این شاخص‌ها برای هر درس در هر دو آزمون دستیاری تخصصی دندانپزشکی و پزشکی بررسی و تک بُعدی بودن تمام درس‌ها، تأیید شد. برای انتخاب مدل مناسب، برآورد پارامتر سؤالات و پایایی هر درس، از نرم‌افزار R و پکیج mirt استفاده

1. Schermelleh-Engel

شده است. نمونه‌ای از فرایند انتخاب مدل مرتبط با درس ارتودانتیکس در زیر آمده است:

جدول ۲. مقادیر لگاریتم درست‌نمایی

سه پارامتری	دو پارامتری	یک پارامتری
-۲۱۹۰۴/۵۹	-۲۱۹۲۸/۳	-۲۲۱۹۲/۷۳

برای انتخاب مدل مناسب باید معناداری اختلاف میان شاخص درست‌نمایی بین دو مدل بررسی شود. در اینجا اختلاف میان شاخص درست‌نمایی بین مدل یک و دو پارامتری معنادار بود. در گام بعد مدل دو و سه پارامتری بررسی شد که اختلاف میان این دو نیز تأیید شد، بنابراین برای تحلیل سؤالات از مدل سه پارامتری استفاده شد. پس از تعیین مدل، پارامترهای سؤالات هر درس با استفاده از مدل مناسب برآورد می‌شود. برای نمونه پارامترهای سؤالات درس ارتودانتیکس که با مدل سه پارامتری تحلیل شده، در ادامه آمده است:

جدول ۳. پارامتر سؤالات درس ارتودانتیکس

سؤال	تشخیص	قدرت تشخیص	دشواری	درجه دشواری	حدس پذیری	درجه حدس پذیری
۱	۱/۵۶	زیاد	-۰/۳۴	نسبتاً آسان	۰	
۲	۱/۴۹	زیاد	۰/۹۰	نسبتاً سخت	۰	
۳	۱/۲۵	متوسط	-۰/۸۵	نسبتاً آسان	۰/۰۴	
۴	-۰/۶۳	کم	-۱/۳۶	آسان	۰	
۵	۰/۷۶	متوسط	۱	سخت	۰	
۶	۲/۱۵	خیلی زیاد	-۰/۲۴	نسبتاً آسان	۰/۰۴	
۷	۱/۱۵	متوسط	۰	نسبتاً آسان	۰	
۸	۱/۳۰	متوسط	۰/۲۳	نسبتاً سخت	۰/۴۱	بسیار نامطلوب
۹	۱/۳۱	متوسط	۰/۹۷	نسبتاً سخت	۰	
۱۰	۲/۴۲	خیلی زیاد	۰/۹۳	نسبتاً سخت	۰/۱۰	نسبتاً نامطلوب
۱۱	۲/۵۸	خیلی زیاد	۰/۲۶	نسبتاً سخت	۰/۰۳	

	۰/۰۶	خیلی سخت	۳	متوسط	۱/۱۷	۱۲
نامطلوب	۰/۲۳	خیلی سخت	۲/۲۴	خیلی زیاد	۲/۳۰	۱۳
	۰/۰۲	نسبتاً سخت	۰/۷۵	زیاد	۱/۴۴	۱۴
	۰	سخت	۱/۰۵	زیاد	۱/۳۵	۱۵
	۰	نسبتاً سخت	۰/۷۲	متوسط	۰/۹۰	۱۶
	۰/۰۳	نسبتاً آسان	-۰/۵۳	زیاد	۱/۴۹	۱۷
	۰	سخت	۱/۹۵	متوسط	۱/۰۶	۱۸
نسبتاً نامطلوب	۰/۱۲	سخت	۱/۷۱	متوسط	۱/۲۸	۱۹
	۰/۰۹	نسبتاً آسان	-۰/۰۴	متوسط	۰/۹۲	۲۰
	نسبتاً سخت	۰/۷۰	میانگین دشواری آزمون			
متوسط		۱/۱۷	میانگین تشخیصی آزمون			

برای تعیین سؤالات مطلوب در هر آزمون، باید به پارامترهای سؤال توجه کرد. از آنجا که هدف آزمون‌های دستیاری تخصصی ورود قوی‌ترین افراد به دوره است، بنابراین باید سطح دشواری سؤالات بالا باشد تا احتمال پاسخ صحیح افراد قوی افزایش یابد بنابراین سطح مطلوب دشواری در حدود یک به بالا در نظر گرفته می‌شود. بیکر نیز سطح مطلوب تشخیص را در حدود ۰/۵ به بالا و حدس‌پذیری را کمتر از ۰/۳۵ بیان کرده است (بیکر، ۲۰۰۱). در ادامه خلاصه وضعیت سؤالات آزمون‌های دستیاری تخصصی دندانپزشکی و پزشکی آمده است:



جدول ۴. خلاصه وضعیت سؤالات آزمون‌های دستیاری تخصصی دندانپزشکی سال ۱۳۹۹

عنوان درس	تعداد سؤال	تعداد ابعاد	مدل مناسب	میانگین دشواری	میانگین تشخیص	تشخیص منفی	حدس‌پذیری بالا	سؤال مناسب	پایایی
ارتودانتیکس	۲۰	۱	۳	۷۰/۰	۱۷/۱	-	۱	۵	۸۳/۰
اندودانتیکس	۲۰	۱	۳	۷۷/۰	۳۸/۱	۲	-	۲	۸۳/۰
بیماری‌های دهان و فک و صورت	۲۰	۱	۳	۶۷/۰	۳۵/۱	-	۱	۲	۷۸/۰
آسیب شناسی دهان و فک و صورت	۲۰	۱	۲	۰۲/۱	۱۴/۱	۱	-	۸	۸۰/۰
پروتزهای دندانی	۳۰	۱	۳	۶۴/۰	۳۷/۱	-	-	۷	۸۷/۰
پریودانتیکس	۲۰	۱	۳	۴۸/۰	۳/۱	-	-	۲	۸۱/۰
جراحی دهان و فک و صورت	۲۰	۱	۲	-۰/۲۶	۱۶/۱	-	-	-	۸۲/۰
دندانپزشکی ترمیمی	۲۰	۱	۲	۳۸/۰	۰۹/۱	-	-	۱	۸۱/۰
دندانپزشکی کودکان	۲۰	۱	۳	-۰/۱۵	۵۰/۱	-	-	۱	۸۴/۰
رادیولوژی دهان و فک و صورت	۲۰	۱	۲	۹۳/۰	۹۰/۰	۱	-	۹	۷۶/۰
مواد دندانی	۲۰	۱	۲	۵۲/۱	۵۱/۱	-	-	۲	۷۲/۰
سلامت دهان و دندانپزشکی اجتماعی	۱۰	۱	۲	۳۵/۰	۵۲/۱	-	-	۱	۷۸/۰
زبان انگلیسی	۲۰	۱	۲	۹۵/۰	۸۸/۰	۱	-	۶	۷۶/۰

جدول ۵. خلاصه وضعیت سؤالات آزمون‌های دستیاری تخصصی پزشکی سال ۱۳۹۹

عنوان درس پزشکی	تعداد سؤال	تعداد ابعاد	مدل مناسب	میانگین دشواری	میانگین تشخیص	تشخیص منفی	حدس‌پذیری بالا	سؤال مناسب	پایایی
جراحی	۲۴	۱	۳	۰۴/۱	۳۸/۱	۱	۲	-	۸۳/۰
داخلی	۴۵	۱	۳	۸۲/۰	۲۷/۱	۱	-	۸	۸۸/۰
کودکان	۲۶	۱	۳	۳۵/۰	۱۷/۱	۲	۲	۲	۸۲/۰
زنان و زایمان	۱۸	۱	۳	-۰/۴۱	۲۲/۱	-	-	۲	۷۹/۰
رادیولوژی	۶	۱	۲	۴۱/۱	۶۰/۰	۱	-	۵	۲۳/۰
پوست	۶	۱	۲	۸۴/۰	۱۳/۱	-	-	۱	۶۱/۰
جراحی دهان و فک و صورت	۶	۱	۲	۵۶/۱	۳۳/۱	-	-	۳	۵۹/۰
آسیب شناسی	۹	۱	۲	۳۶/۱	۷۰/۱	-	-	۲	۷۳/۰

عنوان درس پزشکی	تعداد سؤال	تعداد ابعاد	مدل مناسب	میانگین دشواری	میانگین تشخیص	تشخیص منفی	حدس پذیری بالا	سؤال مناسب	پایایی
اخلاق پزشکی	۲	تحلیل کلاسیک		۵۲/۰	۷۲/۰	-		۱	۱۰/۰
جراحی استخوان و مفاصل	۸	۱	۳	۶۷/۰	۹۴/۱	-	-	-	۷۱/۰
گوش و حلق و بینی	۶	۱	۲	۵۹/۰	۱۵/۱	-	-	۳	۷۳/۰
روان پزشکی	۸	۱	۳	-۰/۱۱	۱۷/۱	-	-	۱	۵۹/۰
عفونی	۱۰	۱	۳	۵۵/۰	۴۶/۱	-	-	۲	۷۳/۰
جراحی کلیه و مجاری ادراری و تناسلی	۶	۱	۳	۸۷/۰	۵۶/۱	-	-	۲	۶۳/۰
مغز و اعصاب	۸	۱	۳	۰۸/۱	۶۹/۱	-	-	-	۷۲/۰
آمار و اپیدمیولوژی	۶	۱	۲	۲۸/۱	۹۲/۰	-	-	۴	۴۳/۰
فارماکولوژی	۱۰	۱	۳	-۰/۷۱	۹۰/۱	-	-	۱	۶۶/۰

در آزمون دستیاری تخصصی دندانپزشکی، شش درس با مدل سه پارامتری و هفت درس با مدل دو پارامتری، و در آزمون پزشکی ده درس با مدل سه پارامتری و شش درس با مدل دو پارامتری تحلیل شد. در آزمون دستیاری پزشکی، تحلیل پرسش-پاسخ درس اخلاق به دلیل کم بودن تعداد سؤالات امکان پذیر نبود و بنابر این، سؤالات این درس بر اساس نظریه کلاسیک تحلیل شد. در آزمون دستیاری تخصصی دندانپزشکی، در تمام درس‌ها شرط تک بعدی بودن رعایت شده بود و از مجموع ۲۵۰ سؤال، ۶۵ سؤال مطلوب و ۵ سؤال دارای پارامتر تشخیص منفی بود و میزان پایایی برای تمام درس‌های این آزمون، در سطح قابل قبولی قرار داشت.

در آزمون دستیاری تخصصی پزشکی نیز شرط تک بعدی بودن در تمام درس‌ها رعایت شده بود و از مجموع ۲۰۰ سؤال، ۶۳ سؤال در سطح مطلوب و ۵ سؤال دارای پارامتر تشخیص منفی بود. در این آزمون، میزان پایایی برای ۱۵ درس از مجموع ۱۷ درس در سطح قابل قبول و برای دو درس (رادیلولوژی، اخلاق پزشکی) در سطح غیر قابل قبولی قرار داشت.

#### جدول ۶. خلاصه وضعیت آزمون دستیاری پزشکی و دندانپزشکی

نام آزمون	تعداد درس	تعداد سؤالات	تعداد سؤالات با پارامتر تشخیص منفی	تعداد سؤالات با کیفیت مطلوب
دستیاری دندانپزشکی	۱۳	۲۵۰	۵	۶۵
دستیاری پزشکی	۱۷	۲۰۰	۵	۶۳

#### بحث

همان گونه که پژوهش‌ها نشان داده است، روانسنجی سؤالات ابزاری ارزشمند برای کمک به حفظ سؤالات

با کیفیت و حذف یا اصلاح سؤالات نامناسب است و این مسئله تأثیر مثبتی بر کیفیت آزمون دارد (اسمیت و همکاران، ۲۰۲۰؛ کوئر و همکاران، ۲۰۱۶).

به طور کلی در آزمون دستیاری تخصصی دندانپزشکی، پایایی وضعیت مطلوبی داشت و این به معنای اعتبار نتایج آزمون است. سطح دشواری آزمون نسبتاً سخت و تشخیص در سطح مطلوبی بود، به این معنی که آزمون به خوبی توانسته است افراد قوی و ضعیف را از یکدیگر متمایز کند. در آزمون دستیاری تخصصی پزشکی نیز سطح دشواری آزمون نسبتاً سخت بود و تشخیص در سطح مطلوبی قرار داشت به این معنی که آزمون به خوبی توانسته است افراد قوی و ضعیف را از یکدیگر جدا کند. پایایی درس‌های آزمون دستیاری تخصصی پزشکی، به جز در سه درس رادیولوژی، اخلاق پزشکی و آمار و اپیدمیولوژی، در سطح متوسط تا قابل قبولی قرار داشت.

مقدار پایایی قابل قبول برای یک آزمون توانایی، ۰/۶۰ به بالا و برای آزمون‌های سرنوشت ساز ۰/۸۰ به بالا است (سیف، ۱۳۹۷) و این در حالی است که پایایی برای درس رادیولوژی ۰/۲۳، درس آمار و اپیدمیولوژی ۰/۴۳ و برای درس اخلاق پزشکی ۰/۱۰ بود، یعنی نتایج حاصل از این سه درس، اعتبار قابل قبولی ندارند و با تکرار آزمون در شرایط یکسان، احتمالاً نتایج متفاوتی حاصل خواهد شد.

عوامل بسیاری بر پایایی آزمون اثر دارد از جمله تعداد سؤالات، تعداد آزمون شونده‌گان و همسانی سؤالات با یکدیگر است. در باره درس اخلاق پزشکی می‌توان تعداد بسیار کم سؤالات (دو سؤال) را دلیل پایایی بسیار پایین آن دانست که با افزایش تعداد سؤالات، می‌توان پایایی در این درس را افزایش داد. در درس آمار و اپیدمیولوژی نیز می‌توان کم بودن تعداد سؤالات را دلیل پایایی پایین آن دانست. همان گونه که ملاحظه می‌شود، درس‌هایی که تعداد سؤالات بیشتری دارند، پایایی بالاتری نیز دارند. در درس رادیولوژی، علاوه بر کم بودن تعداد سؤالات (شش سؤال)، وجود یک سؤال با پارامتر تشخیص منفی، می‌تواند سبب پایین بودن پایایی باشد که بر این اساس با افزایش تعداد سؤالات و نیز اصلاح و یا حذف سؤال با پارامتر تشخیص منفی می‌توان پایایی را در این درس افزایش داد.

پارامتر تشخیص، نشان می‌دهد که هر سؤال تا چه اندازه می‌تواند داوطلبان قوی را از ضعیف تشخیص دهد و منفی شدن این شاخص یعنی داوطلبان با توانایی پایین، به سؤال پاسخ صحیح داده‌اند در حالی که داوطلبان با توانایی بالا نتوانسته‌اند به سؤال پاسخ صحیح دهند و این یعنی سؤال به گونه‌ی عکس عمل کرده است. برای وجود پارامتر تشخیص منفی چند علت می‌توان برشمرد. اول اشتباهات ناشی از تدوین سؤال شامل اشکال در بدنه سؤال و یا گزینه‌های انحرافی، دوم اشتباهاتی که هنگام تهیه دفترچه آزمون و کلید صورت می‌گیرد مانند اشتباهات تایپی و سوم اشتباهاتی که در آموزش دانشجویان صورت گرفته است، مانند آموزش نادرست یک مفهوم (بیکر، ۲۰۰۱). با توجه به اینکه مخاطبان آزمون‌های دستیاری، پزشکان و دندانپزشکان هستند، احتمال آموزش اشتباه یک مفهوم بسیار اندک است، بنابر این برای بررسی علت پارامتر منفی تشخیص باید بدنه سؤال،

گزینه‌ها و فایل کلید از دو جنبه منطقی (تدوین سؤال) و فنی (اشکالات تاییبی) مورد بررسی قرار گیرد. در هر دو آزمون حدود ۲ و ۲/۵ درصد از سؤالات دارای پارامتر تشخیص منفی است. از سوی دیگر تنها در حدود ۳۰ درصد از سؤالات هر دو آزمون از نظر ویژگی‌های روانسنجی، وضعیت مطلوبی داشت.

### نتیجه‌گیری

بر اساس نتایج پژوهش حاضر، می‌توان نتیجه گرفت که آزمون‌های دستیاری دندانپزشکی و پزشکی بررسی شده از نظر ویژگی‌های روانسنجی در وضعیت مطلوبی قرار نداشت. پیشنهاد می‌شود برای افزایش کیفیت آزمون‌های دستیاری تخصصی، طراحان سؤال، هنگام طراحی سؤالات ضمن رعایت اصول طراحی سؤال، علاوه بر توجه به هر دو پارامتر تشخیص و دشواری سؤال و کیفیت گزینه‌های انحرافی، همسانی سؤالات با یکدیگر و با موضوع درس مربوط را نیز مد نظر قرار دهند تا از این راه بتوان کیفیت آزمون‌های دستیاری تخصصی را افزایش داد.

### تقدیر و تشکر

بدینوسیله از ریاست محترم مرکز سنجش آموزش پزشکی به خاطر حمایت معنوی در اجرای پژوهش حاضر سپاس‌گزاری می‌شود. همچنین از آقای دکتر سید سعید هاشمی نظری برای همکاری در این تحقیق، تشکر و قدردانی می‌شود.

### References

- Shakurnia Ah, Mozaffari Ar, Khosravi Brougeni A, (2010). Survey on Structural of MCQs of residency exam in AJUMS, *Jundishapur Scientific Medical Journal*, 8(4), 491. [magiran.com/p695212](http://magiran.com/p695212)
- Saburi M, Shayan S, Salehi A, Honarpisheh H. Review Psychometric Parameters of the 29th Residency Test (1380) According to the Classic Test Theory (CTT). *Iranian Journal of Medical Education*. 2002; 2:49-50 URL: <http://ijme.mui.ac.ir/article-1-551-fa.html>
- Taheri M, khoshrang H, Asadi louyeh A, Hidarzadeh A. Quality of Residents' Promotion Exams before & after Educational Intervention in 2010-2011. *Iranian Journal of Medical Education*. 2013; 13 (7):551-560 URL: <http://ijme.mui.ac.ir/article-1-2350-fa.html>
- Minaei, a., & Falsafinejad, m. (2010). Methods for assessing unidimensionality in dichotomous irt models. *Training measurement*, 1(3), 71-99. <https://www.sid.ir/en/journal/viewpaper.aspx?id=255760>
- S. H. Naji. (2018) Comparative study of student assessment and admission system in countries. Master's Thesis, Tehran university, Tehran
- Hambleton, Ronald K. & Swaminathan, Hariharan. & Rogers, H. Jane. (1991). *Fundamentals*

- of item response theory. Newbury Park, Calif : Sage Publications, Translated in persian, 2010, Alameh tabai university, Tehran.
- Embretson, Susan & Reise, S.. (2000). Item Response Theory For Psychologists. Tranlated in persian, Roshd, Tehran, 2010
- Baker, Frank B. (1985). *The Basics of Item Response Theory*. the ERIC Clearinghouse on Assessment and Evaluation.
- Kaur, M., Singla, S., & Mahajan, R. (2016). Item analysis of in use multiple choice questions in pharmacology. *International journal of applied & basic medical research*, 6(3), 170–173. <https://doi.org/10.4103/2229-516X.186965>
- Khasawneh, W., Obeidat, N., Albiss, B., & El-Salem, K. (2020). Selection criteria and match results for postgraduate residency programs: A cross-sectional model from a major academic center in Jordan. *Annals of medicine and surgery (2012)*, 59, 199–203. <https://doi.org/10.1016/j.amsu.2020.10.001>
- Montgomery, G. P., Crockford, D. N., & Hecker, K. (2010). The Coordinators of Psychiatric Education (COPE) Residency In-Training Exam: a preliminary psychometric assessment. *Academic psychiatry : the journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*. 34(3), 203–207. <https://doi.org/10.1176/appi.ap.34.3.203>
- Shrestha, A., Bhandary, Sh., Shrestha, Sh. (2019). Situational judgement test: Psychometric analysis of a pilot study for selecting post graduate medical student in residency program. *Journal of Patan Academy of Health Sciences*. 6(2): 81-7.
- Smith, E.B., Lewis, P., Benefield, T., Catanzano, T.M., Khan, M.J., Eric Nyberg, Jordan, Sh. (2020). Auditing RadExam: Employing Psychometrics to Improve Exam Quality. *Academic Radiology*. <https://doi.org/10.1016/j.acra.2020.05.037>
- Thomas, T.V., Kuruvilla, T.P., Eldrin Bhanat, E., Parr, A.Q., Ashley Albert, A., Vijayakumar, S. (2020). An Analysis of the Residents' Research, Education, and Wellness Resources in Radiation Oncology Residency Programs in the United States. *Advances in Radiation Oncology*. (2021) 6, 100548. <https://doi.org/10.1016/j.adro.2020.08.005>