



Exploring the psychometric characteristics of the questions of the medical and dental residency entrance exam in 2020 by item responses theory

Hoda sadat Naji¹ , Ali Moghadamzadeh² , Samaneh Panjeh Ali Beik³

1. Ph.D student. Department of Curriculum Planning and Educational Methods, Faculty of Psychology and Education, University of Tehran, Tehran, Iran; (Corresponding Author), Email: s.hodanaji@ut.ac.ir
2. associate professor. Department of Curriculum Planning and Educational Methods, Faculty of Psychology and Education, University of Tehran. Email: amoghadamzadeh@ut.ac.ir
3. Chaiman of research, development and education, National Center for Medical Education Assessment, Tehran, Iran. Email: s.alibeik@sanjeshp.ir

Article Info

Article Type:
Research Article

Received:2022/03/14

Received in revised form:2022/09/04

Accepted:2022/09/13

Published online:
2022/09/14

ABSTRACT

Objective: This study aims to investigate the quality of medical and dental residency programs entrance examinations based on questions' psychometric properties and item-response theory.

Methods: The research method is descriptive-analytical. The analyzed data includes all the participants' answers in the 2021 assistantship exam, which has been analyzed based on the item-response theory.

Results: Results show that dental and medical assistant exams are not satisfying regarding psychometric properties, but designers need to be more careful in formulating questions.

Conclusion: To increase the quality of residency entrance examination and observe the taxonomy of questions, designers must consider the parameter of diagnosis and difficulty of the question and the quality of deviant options, matching questions with each other and the subject of the relevant lesson increased.

Keywords: Residency examination, Medicine, Dentistry, Item-Response theory, Psychometrics

Cite this article: Naji, Seydeh Hoda. Moghadamzadeh, Ali. Panjeh Ali Beik, Samaneh. (2022). Exploring the psychometric characteristics of the questions of the medical and dental residency entrance exam in 2020 by item responses theory. *Educational Measurement and Evaluation Studies*, 12 (38): 10 pages.

DOI: 10.22034/emes.2022.550571.2365



© The Author(s).

Publisher: National Organization of Educational Testing (NOET)



بررسی ویژگی‌های روانسنجی سوالات آزمون دستیاری پزشکی و دندانپزشکی سال ۱۳۹۹،

بر اساس نظریه‌ی پرسش-پاسخ

سیده هدی ناجی^۱، علی مقدم زاده^۲، سمانه پنجه علی بیک^۳

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 سوالات و با استفاده از نظریه پرسش-پاسخ است.
مقاله پژوهشی	روش پژوهش: روش پژوهش از نوع توصیفی-تحلیلی است. داده‌های مورد تحلیل شامل پاسخنامه‌ی تمام افراد شرکت کننده در آزمون دستیاری سال ۱۳۹۹ است که بر اساس نظریه پرسش-پاسخ مورد تحلیل قرار گرفته است.
دریافت: ۱۴۰۰/۱۲/۲۳	یافته‌ها: یافته‌ها نشان می‌دهد آزمون‌های دستیاری دندانپزشکی و پزشکی از نظر ویژگی‌های روانسنجی از وضعیت مطلوبی برخوردار نیستند، اما لازم است طراحان در تدوین سوالات دقت بیشتری داشته باشند.
اصلاح: ۱۴۰۱/۰۶/۱۳	نتیجه‌گیری: برای افزایش کیفیت آزمون‌های دستیاری تخصصی، علاوه بر رعایت تاکسونومی سوالات، لازم است طراحان پارامتر تشخیص و دشواری سوال و نیز کیفیت گزینه‌های انحرافی، همسانی سوالات با یکدیگر و با موضوع درس مربوطه را نیز مدنظر قرار دهند تا از این طریق بتوان کیفیت آزمون‌های دستیاری تخصصی را افزایش داد.
پذیرش: ۱۴۰۱/۰۶/۲۲	واژه‌های کلیدی: آزمون دستیاری، پزشکی، دندانپزشکی، نظریه پرسش-پاسخ، روانسنجی
انتشار: ۱۴۰۱/۰۶/۲۳	

استناد: ناجی، سیده هدی. مقدم زاده، علی. پنجه علی بیک، سمانه (۱۴۰۱). بررسی ویژگی‌های روانسنجی سوالات آزمون دستیاری پزشکی و دندانپزشکی سال ۱۳۹۹، بر اساس نظریه‌ی پرسش-پاسخ. *مطالعات اندازه‌گیری و ارزشیابی آموزشی*، ۱۲ (شماره ۳۸)، ۱۰ 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 می‌توانند بهترین خدمات بهداشتی را ارائه دهند برگزار می‌شود، دوره دستیاری زمانی برای آموزش و تربیت دانشجویان پزشکی است که در طی آن دانشجویان پزشکی به پزشکی توانا تبدیل می‌شوند (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، از اهمیت کلیدی در نظام آموزش عالی کشور برخوردار است و از این رو تحول و ارتقای کیفیت آن‌ها از دغدغه‌های جدی مسئولین اجرایی و صاحب‌نظران است. در این بین همانند هر سیستم 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 از جایگاه ویژه‌ای برخوردار است (خوشرننگ و همکاران، ۱۳۹۲).

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

آزمون‌های دستیاری تخصصی در ایران بصورت چهار گزینه‌ای تدوین و اجرا می‌شود. در خصوص آزمون‌های چند گزینه‌ای می‌توان گفت، با توجه به اینکه این نوع از آزمون‌ها می‌تواند بخش وسیعی از دانش داوطلبان را در مدت کوتاهی مورد ارزیابی قرار دهد و علاوه بر حساسیت کم نسبت به حدس، به آسانی قابل تصحیح است، بهترین نوع آزمون عینی بشمار می‌رود و در سنجش پزشکی بسیار کاربرد دارد. علارغم تصور عموم درباره‌ی آزمون‌های چند گزینه‌ای، اگر این آزمون‌ها با دقت و اصولی تهیه شوند، علاوه بر آنکه می‌توانند سطوح بالای شناختی را اندازه بگیرند، بخوبی می‌توانند افراد قوی را از ضعیف تشخیص دهند. برای آنکه بتوان آزمون با کیفیتی تهیه کرد، لازم است سوالات آزمون با کیفیت باشند، زیرا کیفیت آزمون از کیفیت سوالات آن ناشی می‌شود. در منابع مختلف ویژگی‌های یک سوال خوب بیان شده است و به هر میزان که یک سوال از این معیارهای برخوردار باشد، می‌توان گفت سوال بهتری است. از جمله‌ی این معیارها، باید به دشواری سوال، قدرت تشخیص و همخوانی سوال 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 خود به بررسی کیفیت آزمون‌ها با توجه به سطوح شناختی اندازه‌گیری شده براساس طبقه بندی بلوم پرداخته‌اند. این پژوهش نشان داد که بیشتر سوالات آزمون دستیاری، سطوح پایین شناختی را مورد سنجش قرار می‌دهد.

پژوهش خوشرننگ و همکاران (۱۳۹۲) نیز نشان داد که شرکت در کارگاه‌های آموزش طراحی سوالات چند گزینه‌ای می‌تواند سبب افزایش کیفیت سوالات از نظر سطوح شناختی مورد اندازه‌گیری بر مبنای اصول میلمن^۱ شود. در این پژوهش نیز، کیفیت سوالات تنها از نظر سطوح شناختی مورد اندازه‌گیری بررسی شده است و سایر ویژگی‌های سوال مورد بررسی قرار نگرفته است.

مونتگومری^۲ و همکاران (۲۰۱۰) برای بررسی کیفیت آزمون بازنگری شده‌ی COPE که به عنوان یک آزمون تکوینی برای دستیاران روانپزشکی استفاده می‌شود، از نظریه‌ی کلاسیک اندازه‌گیری استفاده کرده‌اند. مونتگومری در پژوهش خود برای بررسی آزمون از ضریب آلفای کرونباخ برای محاسبه‌ی پایایی، از آنوا برای بررسی روایی، و برای بررسی ویژگی‌های هر سوال از ضریب دشواری و ضریب تشخیص استفاده کرده است. در ضریب دشواری مقدار ۰,۲۵ را برای دشواری بالا (سوال سخت) و مقدار ۰,۹۵ را برای دشواری پایین (سوال آسان) در نظر گرفته است و برای محاسبه‌ی ضریب تشخیص، از همبستگی دو رشته‌ای نقطه‌ای استفاده کرده است. یافته‌های این پژوهش نشان می‌دهد که آزمون بازمینی شده قابلیت اطمینان و روایی مناسب، همچنین ضریب تشخیص مناسبی را دارا است.

یافته‌های این پژوهش صبوری و همکاران (۱۳۸۱) که به بررسی ویژگی‌های روانسنجی بیست و نهمین آزمون دستیاری پرداختند، نشان می‌دهد با توجه به میزان اعتبار و خطای استاندارد اندازه‌گیری، آزمون به خوبی و دقیق توانسته است میزان دانش نظری شرکت‌کنندگان را ارزیابی

1 Milmen

2 Montgomery

نماید. و به عبارت دیگر نتایج اجرای این آزمون (رتبه و نمره شرکت کنندگان) معتبر است. با توجه به عدم وجود قدرت تمیز منفی در بین سوالات شیوه طراحی سوالات مناسب بوده است. میزان درجه دشواری سوال تا حدودی متناسب با تعداد دستیار مورد نظر جهت انتخاب می باشد. همچنین صبوری پیشنهاد می‌دهد که باتوجه به شرایط مطلوب لازم است ضریب دشواری سوالات کاهش یابد.

شرستا^۱ و همکاران (۲۰۱۹) به بررسی روانسنجی آزمون قضاوت موقعیتی^۲ (SJT) به عنوان یک آزمون معتبر برای سنجش توانایی‌های شناختی در پذیرش داوطلبان تحصیلات تکمیلی پزشکی، پرداختند. این مطالعه مقطعی در آکادمی علوم بهداشتی پاتان (PAHS) در ماه مه ۲۰۱۷ انجام شد. نتایج پژوهش آن‌ها نشان می‌دهد آزمون قضاوت موقعیتی ابزاری قابل اعتماد برای سنجش خصوصیات غیر دانشگاهی در آزمون تحصیلات تکمیلی پزشکی است و قابلیت اطمینان و همسانی درونی ابزار را می‌توان با افزایش تعداد سوالات با کیفیت، بهبود بخشید.

اسمیت^۳ و همکاران در سال ۲۰۲۰ به انجام یک مطالعه با دو گام در آزمون رادیولوژی دستیاران پرداختند، در این مطالعه گام اول در سال ۲۰۱۸ به عنوان آزمایشی و گام دوم به فاصله یک سال بعد انجام شد. در هر دو گام، بوردهای تخصصی مربوطه تمام سوالات آزمون را با استفاده از داده‌های آماری از قبیل تعداد سوال و آزمون، ضریب دشواری سوال، ضریب تمیز، ارزیابی کردند و سوالات معیوب شناسایی شده اصلاح، حذف و یا جایگزین شدند. نتایج تحقیق تاثیر مثبت اصلاح را بر کیفیت آزمون نشان داد.

در پژوهشی دیگر (کوثر^۴ و همکاران، ۲۰۱۶) بررسی کیفیت آزمون فارماکولوژی، از طریق انجام تحلیل سوال مورد بررسی قرار گرفت. محققان این پژوهش نتیجه گرفتند که تحلیل سوال یک ابزار ارزشمند برای کمک به حفظ سوالات با کیفیت و حذف سوالات غیر مفید و در نهایت افزایش مهارت در ساخت آزمون است، همچنین مشخص می‌کند که کدام سوال به شفافیت و تاکید بیشتری نیاز دارد.

اگرچه در ایران کیفیت آزمون‌های دستیاری تخصصی مورد پژوهش قرار گرفته، اما در این پژوهش‌ها به ویژگی‌های روانسنجی سوالات (از جمله دشواری و قدرت تشخیص) پرداخته نشده است و پژوهش منتشر شده ای در خصوص بررسی ویژگی‌های روانسنجی وجود ندارد و کیفیت آزمون‌ها صرفاً با توجه به تاکسونومی و سطوح شناختی مورد اندازه‌گیری بررسی شده است و این در حالی است که یک سوال خوب از نظر سنجش سطوح شناختی بالا، می‌تواند دلیل داشتن قدرت تشخیص پایین و یا سختی نامناسب، سوال خوبی برای قرار گرفتن در آزمون‌های کلانی مانند آزمون ورودی دستیاری تخصصی نباشد. بنابراین لازم است علاوه بر بررسی تاکسونومی سوالات، ویژگی‌های روانسنجی سوالات نیز مورد بررسی قرار گیرد، در نتیجه پژوهش پیش رو با هدف بررسی کیفیت آزمون‌های دستیاری تخصصی براساس ویژگی‌های روانسنجی سوالات آزمون‌های دستیاری تخصصی انجام شده است.

روش پژوهش

روش پژوهش از نوع توصیفی-تحلیل است. داده‌های مورد تحلیل شامل پاسخنامه‌ی 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 نظریه‌ی پرسش-پاسخ، ویژگی سوال بگونه‌ای برآورد می‌شود که نابسته به گروه است. در این نظریه از معادله‌های پیچیده‌ی ریاضی برای برآورد پارامتر سوال (سختی، تشخیص و حدس‌پذیری) و توانایی افراد استفاده می‌شود. از جمله نقاط قوت نظریه‌ی پرسش-پاسخ نسبت به نظریه‌ی کلاسیک، برآورد دقیق‌تر سطح دشواری و تشخیص سوالاتی است که هیچ یک از افراد به آن پاسخ درست نداده‌اند ($p=0$ در نظریه‌ی کلاسیک) و یا همه افراد به آن جواب درست داده‌اند ($q=0$ در نظریه‌ی کلاسیک) (همبلتون و همکاران، ۱۳۸۹/۱۹۹۱، امبرستون و همکاران، ۱۳۸۸/۲۰۰۰). معادله‌ی اصلی نظریه‌ی پرسش-پاسخ بصورت زیر است:

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

b پارامتر دشواری، a پارامتر تشخیص و c حدس‌پذیری سوال و θ توانایی افراد، D مقدار ثابت و $P_i(\theta)$ احتمال پاسخ صحیح به سوال i است. تعریف پارامتر تشخیص و دشواری در نظریه‌ی پرسش-پاسخ، معادل با ضریب تشخیص و دشواری در نظریه‌ی کلاسیک است، اما حدس‌پذیری در نظریه‌ی پرسش-پاسخ یعنی احتمال انتخاب پاسخ صحیح توسط افراد با توانایی بسیار پایین. در نظریه‌ی پرسش-پاسخ، دامنه‌ی

1 Shrestha

2 Situational judgement test

3 Smith

4 Kaur

عملی دشواری از ۴- تا ۴+ است و هرچه مقدار آن بیشتر باشد، یعنی سوال سخت‌تر است. دامنه‌ی عملی پارامتر تشخیص از ۰ تا ۳+ است و هر چه مقدار آن بیشتر باشد، یعنی سوال توانایی بیشتری در تفکیک گروه قوی و ضعیف دارد. این نظریه با توجه به تعداد پارامترهای در نظر گرفته شده برای برآورد توانایی، دارای مدل‌های مختلفی است که مشهورترین آن‌ها مدل یک پارامتری، دو پارامتری و سه پارامتری است. در مدل یک پارامتری، برای تمام سوالات فقط یک پارامتر تشخیص برآورد می‌شود ولی برای هر سوال یک پارامتر دشواری جداگانه برآورد می‌شود. به عبارت دیگر پارامتر تشخیص برای تمام سوال‌ها برابر در نظر گرفته می‌شود اما در مدل دو پارامتری، علاوه بر پارامتر دشواری، برای هر سوال یک پارامتر تشخیص برآورد می‌شود، و در مدل سه پارامتری، علاوه بر پارامتر تشخیص و دشواری، برای هر سوال پارامتر حدس نیز برآورد می‌شود (همبلتون و همکاران، ۱۳۸۹/۱۹۹۱، امبرستون و همکاران، ۱۳۸۸/۲۰۰۰).

نظریه‌ی پرسش-پاسخ برای محاسبه‌ی پایایی از فرمول خاص خود استفاده می‌کند. یکی از ضرایب پایایی مورد استفاده در این نظریه، پایایی تجربی است که نشان دهنده‌ی دقت آزمون در برآورد توانایی افراد است. مقدار پایایی تجربی از ۰ تا ۱ است و هرچه مقدار آن به یک نزدیکتر باشد، دقت ابزار در برآورد توانایی افراد بیشتر است (دی‌آیالا، ۲۰۰۸).

از آنجا که نظریه‌ی 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 نمود. برای برازش مدل از آماره‌ی χ^2 استفاده می‌شود:

$$\chi^2 = -2\text{LOG LIKELIHOOD (مدل پیچیده‌تر)} - (-2\text{LOG LIKELIHOOD (مدل ساده‌تر)})$$

علاوه بر ویژگی روانسنجی سوالات، یکی دیگر از ملاک‌های بررسی کیفیت یک آزمون، مساله‌ی بعدیت است. آزمون‌های توانایی معمولاً با هدف سنجش یک توانایی تهیه می‌شوند به این معنا که یک توانایی غالب باید عملکرد آزمودنی را تحت تاثیر قرار دهد. به این ویژگی تک بعدی بودن گفته می‌شود. نقض مفروضه‌ی تک بعدی بودن می‌تواند برآورد پارامترهای سوال و توانایی را به‌طور جدی دچار تورش نماید، علاوه‌براین تک بعدی بودن سوال‌ها برای تفسیر روشن و بدون ابهام نمره‌ها و انجام مقایسه بین افراد مهم است (مینایی و فلسفی‌نژاد، ۱۳۸۹). در نظریه‌ی پرسش-پاسخ، پیش از تحلیل داده‌ها باید تک بعدی بودن آزمون مورد بررسی قرار گیرد. برای بررسی تک بعدی بودن از شاخص‌های برازش تاناکا و RMSE استفاده می‌شود.

یافته‌ها

برای پاسخ به پرسش پژوهشی، تک تک سوالات آزمون‌های دستیاری تخصصی پزشکی و دندانپزشکی با استفاده از نظریه‌ی پرسش-پاسخ مورد تحلیل قرار گرفته و پارامترهای سوال، پایایی و بعدیت آزمون بررسی شده است. آزمون‌های دستیاری تخصصی از چند درس تشکیل شده‌اند، بنابراین برای تحلیل روانسنجی سوالات باید هر درس را بصورت جداگانه مورد بررسی قرار داد. مشخصات توصیفی آزمون‌ها در جدول زیر آمده است.

جدول ۱. مشخصات توصیفی آزمون دستیاری پزشکی و دندانپزشکی

نام آزمون	تعداد دروس	تعداد سوالات	تعداد داوطلبان
دستیاری دندانپزشکی	۱۳	۲۵۰	۱۹۶۱
دستیاری پزشکی	۱۷	۲۰۰	۱۱۶۵۵

ابتدا برای بررسی مفروضه‌ی تک بعدی بودن، داده‌های هر درس بصورت جداگانه توسط نرم‌افزار NOHARM مورد بررسی قرار گرفت. 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 دهنده‌ی مدل مناسب است. مقدار قابل قبول برای شاخص RMSE کمتر از ۰,۰۱ است (مینایی و فلسفی‌نژاد، ۱۳۸۹؛ شرملی انگل^۱ و همکاران، ۲۰۰۳).

¹ Schermelleh-Engel

مقادیر تاناکا و RMSE برای درس ارتودانتیکس نشان دهنده‌ی تک بعدی بودن این درس است. این شاخص‌ها برای تک تک دروس در هر دو آزمون دستیاری تخصصی دندانپزشکی و پزشکی بررسی و تک بعدی بودن تمام دروس، تأیید شده است. برای انتخاب مدل مناسب، برآورد پارامتر سوالات و پایایی هر درس، از نرم‌افزار R و پکیج mirt استفاده شده است. نمونه‌ای از فرایند انتخاب مدل مرتبط با درس ارتودانتیکس در زیر آمده است:

جدول ۲. مقادیر لگاریتم درست‌نمایی

یک پارامتری	دو پارامتری	سه پارامتری
-22192/73	-21928/3	-۲۱۹۰۴/۵۹

برای انتخاب مدل مناسب باید معناداری اختلاف میان شاخص درست‌نمایی بین دو مدل را بررسی نمود. 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 سوالات از مدل سه پارامتری استفاده شد.

پس از تعیین مدل، پارامترهای سوالات هر درس با استفاده از مدل مناسب برآورد می‌شود. برای نمونه پارامترهای سوالات درس ارتودانتیکس که با مدل سه پارامتری تحلیل شده، در ادامه آمده است:

جدول ۳. پارامتر سوالات درس ارتودانتیکس

سوال	تشخیص	قدرت تشخیص	دشواری	درجه‌ی دشواری	حدس پذیری	درجه‌ی حدس پذیری
q1	1/56	زیاد	-0/34	نسبتاً آسان	0	
q2	1/49	زیاد	0/90	نسبتاً سخت	0	
q3	1/25	متوسط	-0/85	نسبتاً آسان	0/04	
q4	0/63	کم	-1/36	آسان	0	
q5	0/76	متوسط	1	سخت	0	
q6	2/15	خیلی زیاد	-0/24	نسبتاً آسان	0/04	
q7	1/15	متوسط	0	نسبتاً آسان	0	
q8	1/30	متوسط	0/23	نسبتاً سخت	0/41	بسیار نامطلوب
q9	1/31	متوسط	0/97	نسبتاً سخت	0	
q10	2/42	خیلی زیاد	0/93	نسبتاً سخت	0/10	نسبتاً نامطلوب
q11	2/58	خیلی زیاد	0/26	نسبتاً سخت	0/03	
q12	1/17	متوسط	3	خیلی سخت	0/06	
q13	2/30	خیلی زیاد	2/24	خیلی سخت	0/23	نامطلوب
q14	1/44	زیاد	0/75	نسبتاً سخت	0/02	
q15	1/35	زیاد	1/05	سخت	0	
q16	0/90	متوسط	0/72	نسبتاً سخت	0	

q17	1/49	زیاد	-0/53	نسبتا آسان	0/03
q18	1/06	متوسط	1/95	سخت	0
q19	1/28	متوسط	1/71	سخت	نسبتا نامطلوب 0/12
q20	0/92	متوسط	-0/04	نسبتا آسان	0/09
میانگین دشواری آزمون		0/70			
میانگین تشخیص آزمون		1/17			
		نسبتا سخت			
		متوسط			

برای تعیین سوالات مطلوب در هر آزمون، باید به پارامترهای سوال توجه کرد. از آنجا که در آزمون‌های دستیاری تخصصی، هدف ورود قوی‌ترین افراد به دوره است، بنابراین باید سطح دشواری سوالات بالا باشد تا احتمال پاسخ صحیح افراد قوی به آن‌ها افزایش یابد بنابراین سطح مطلوب دشواری در حدود ۱ 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 سوالات آزمون‌های دستیاری تخصصی دندانپزشکی و پزشکی آمده است:

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

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

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

پایایی	سوال مناسب	حدس پذیری بالا	تشخیص منفی	میانگین تشخیص	میانگین دشواری	مدل مناسب	تعداد ابعاد	تعداد سوال	عنوان درس پزشکی
۰/۸۳	-	۲	۱	۱/۳۸	۱/۰۴	۳	۱	۲۴	جراحی
۰/۸۸	۸	-	۱	۱/۲۷	۰/۸۲	۳	۱	۴۵	داخلی
۰/۸۲	۲	۲	۲	۱/۱۷	۰/۳۵	۳	۱	۲۶	کودکان
۰/۷۹	۲	-	-	۱/۲۲	-۰/۴۱	۳	۱	۱۸	زنان و زایمان
۰/۲۳	۵	-	۱	۰/۶۰	۱/۴۱	۲	۱	۶	رادیولوژی
۰/۶۱	۱	-	-	۱/۱۳	۰/۸۴	۲	۱	۶	پوست
۰/۵۹	۳	-	-	۱/۳۳	۱/۵۶	۲	۱	۶	جراحی دهان و فک و صورت
۰/۷۳	۲	-	-	۱/۷۰	۱/۳۶	۲	۱	۹	آسیب شناسی
۰/۱۰	۱		-	۰/۷۲	۰/۵۲	تحلیل کلاسیک		۲	اخلاق پزشکی
۰/۷۱	-	-	-	۱/۹۳	۰/۶۷	۳	۱	۸	جراحی استخوان و مفاصل
۰/۷۳	۳	-	-	۱/۱۵	۰/۵۹	۲	۱	۶	گوش و حلق و بینی
۰/۵۹	۱	-	-	۱/۱۷	-۰/۱۱	۳	۱	۸	روان پزشکی
۰/۷۳	۲	-	-	۱/۴۶	۰/۵۵	۳	۱	۱۰	عفونی
۰/۶۳	۲	-	-	۱/۵۶	۰/۸۷	۳	۱	۶	جراحی کلیه و مجاری ادراری و تناسلی
۰/۷۲	-	-	-	۱/۶۹	۱/۰۸	۳	۱	۸	مغز و اعصاب
۰/۴۳	۴	-	-	۰/۹۳	۱/۲۸	۲	۱	۶	آمار و اپیدمیولوژی
۰/۶۶	۱	-	-	۱/۹۰	-۰/۷۱	۳	۱	۱۰	فارماکولوژی

در آزمون دستیاری تخصصی دندانپزشکی، ۶ درس با مدل ۳ پارامتری و ۷ درس با مدل ۲ پارامتری، و در آزمون پزشکی ۱۰ درس با مدل ۳ پارامتری و ۶ درس با مدل ۲ پارامتری تحلیل شده است. در آزمون دستیاری پزشکی، تحلیل پرسش-پاسخ درس اخلاق بدلیل کم بودن تعداد سوالات امکان پذیر نبود، بنابراین سوالات این درس براساس نظریه‌ی کلاسیک تحلیل شده است. در آزمون دستیاری تخصصی دندانپزشکی، در تمام دروس شرط تک بعدی بودن رعایت شده است و از مجموع ۲۵۰ سوال، ۶۵ سوال مطلوب و ۵ سوال دارای پارامتر تشخیص منفی بودند و میزان پایایی برای تمام دروس در این آزمون، در سطح قابل قبولی قرار داشته است.

در آزمون دستیاری تخصصی پزشکی نیز شرط تک بعدی بودن در تمام دروس رعایت شده است و از مجموع ۲۰۰ سوال، ۶۳ سوال در سطح مطلوب و ۵ سوال دارای پارامتر تشخیص منفی بوده است. در این آزمون میزان پایایی برای ۱۵ درس از مجموع ۱۷ درس در سطح قابل قبول و برای دو درس (رادیولوژی، اخلاق پزشکی) در سطح غیر قابل قبولی قرار داشته است.

جدول ۶. خلاصه وضعیت آزمون دستیاری پزشکی و دندانپزشکی

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

بحث

همانگونه که پژوهش‌ها نشان داده است، روانسنجی سوالات ابزاری ارزشمند برای کمک به حفظ سوالات با کیفیت و حذف یا اصلاح سوالات نامناسب است و این مسأله تأثیر مثبتی بر کیفیت آزمون دارد (اسمیت و همکاران، ۲۰۲۰؛ کوثر و همکاران، ۲۰۱۶).

در آزمون دستیاری تخصصی دندانپزشکی، بطور کلی پایایی از وضعیت مطلوبی برخوردار بوده و این به معنای اعتبار نتایج آزمون است. سطح دشواری آزمون نسبتاً سخت بوده و تشخیص در سطح مطلوبی است یعنی 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 بوده و تشخیص در سطح مطلوبی قرار دارد و به خوبی توانسته است افراد قوی و ضعیف را از یکدیگر جدا کند. پایایی دروس آزمون دستیاری تخصصی پزشکی، بجز در سه درس رادیولوژی و اخلاق پزشکی، و آمار و اپیدمیولوژی، در سطح متوسط تا قابل قبولی قرار دارد.

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 است، 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 یک آزمون اثر دارد از جمله تعداد سوالات، تعداد آزمون‌شوندگان و همسانی سوالات با یکدیگر. در رابطه با درس اخلاق پزشکی می‌توان تعداد بسیار کم سوالات (۲ سوال) را دلیل پایایی بسیار پایین آن دانست و با افزایش تعداد سوالات، می‌توان پایایی در این درس را افزایش داد. در درس آمار و اپیدمیولوژی نیز می‌توان کم بودن تعداد سوالات را دلیل پایایی پایین آن دانست، همانگونه که ملاحظه می‌شود، دروسی که تعداد سوالاتی بیشتری دارند، دارای پایایی بالاتری نیز هستند. در درس رادیولوژی، علاوه بر کم بودن تعداد سوالات (۶ سوال)، وجود یک سوال با پارامتر تشخیص منفی، می‌تواند سبب پایین بودن پایایی باشد، در نتیجه با افزایش تعداد سوالات و نیز اصلاح و یا حذف سوال با پارامتر تشخیص منفی می‌توان پایایی را در این درس افزایش داد.

پارامتر تشخیص نشان می‌دهد که یک سوال تا چه اندازه می‌تواند داوطلبان قوی و ضعیف را از هم تشخیص دهد، و منفی شدن آن یعنی داوطلبان با توانایی پایین به سوال پاسخ صحیح داده‌اند در حالی که داوطلبان با توانایی بالا نتوانسته‌اند به سوال پاسخ صحیح دهند و این یعنی سوال بگونه‌ای برعکس عمل کرده است. برای وجود پارامتر تشخیص منفی چند علت می‌تواند وجود داشته باشد، اول اشتباهات ناشی از تدوین سوال شامل اشکال در بدنه‌ی سوال و یا گزینه‌های انحرافی، دوم اشتباهاتی که هنگام تهیه دفترچه آزمون و کلید صورت می‌گیرد مانند اشتباهات تایپی، و سوم اشتباهاتی که در آموزش داوطلبان صورت گرفته است، مانند آموزش نادرست یک مفهوم (بیکر، ۲۰۰۱). با توجه به اینکه مخاطبین آزمون‌های دستیاری، پزشکان و دندانپزشکان هستند، احتمال آموزش نادرست و اشتباه یک مفهوم بسیار اندک است، بنابراین برای بررسی علت پارامتر منفی تشخیص باید بدنه‌ی سوال، گزینه‌ها و فایل کلید از دو جنبه‌ی منطقی (تدوین سوال) و فنی (اشکالات تایپی) مورد بررسی قرار گیرد. در هر دو آزمون حدود ۲ و ۲/۵ درصد از سوالات دارای پارامتر تشخیص منفی است، از سوی دیگر تنها در حدود ۳۰ درصد از سوالات هر دو آزمون در وضعیت مطلوبی از نظر ویژگی‌های روانسنجی قرار دارند

نتیجه‌گیری

بنابراین می‌توان گفت آزمون‌های دستیاری دندانپزشکی و پزشکی از نظر ویژگی‌های روانسنجی در وضعیت مطلوبی قرار ندارند. پیشنهاد می‌شود برای افزایش کیفیت آزمون‌های دستیاری تخصصی، طراحان هنگام طراحی سوالات در کنار توجه به اصول طراحی سوال، علاوه بر توجه به هر دو پارامتر تشخیص و دشواری سوال و کیفیت گزینه‌های انحرافی، همسانی سوالات با یکدیگر و با موضوع درس مربوطه را نیز مدنظر قرار دهند تا از این طریق 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 جهت 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
- 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 :49-50URL: <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. Najj.(2018)Comparative study of student assessment and admission system in countries. Master's Thesis, Tehran universitey, 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>