

The use of RASCH Model (Item-mMapping) in Determining the Standard and Cut-off Score of the Criterion-Reference Tests

Noorali Farrokhi  *

Shirindokht
Habibzadeh 

Asghar Minaei 

Mohammad Jalili 

Corresponding Author, Professor, Department of Assessment Measurment, Allameh tabataba'i University, Tehran, Iran. E-mail: farrokhinoorali@gmail.com

Education department, pardis bentolhoda Faculty, Farhangiyen University, Tehran, Iran. E-mail: habibzadehshirindokht@yahoo.com

Assistant Professor, Department of Assessment Measurment, Allameh tabataba'i University, Tehran, Iran. E-mail: minaeiasghar@yahoo.com

Professor, Department of Emergency Medicine, Department of Medical Education, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran. E-mail: mjalili@tums.ac.ir

Abstract

The Item Response Theory (IRT) has been extensively used in the development of tools in recent years. The present study aimed at determining the standard and cut- off score of the medical basic sciences comprehensive test using the IRT model, the item-mapping method. The statistical population in this study was all the candidates for the medical basic sciences comprehensive test of the tenth pole of Iran (n=324), and the responses of all candidates were analyzed to determine the standard and cut- off score. The tool used in this research was the basic science comprehensive test with 200 four-option items that was taken in September of 2016. In this cross-sectional study, Win steps was used to analyze the items. The findings of this study showed that the test has an item reliability of PR=0/98 and person reliability of IP=0/92 which indicated that the sample variance and test length were appropriate and the items were appropriately selected from the nine domains. Performance of the reviewers' panel and evaluation of the items with the item map method showed that the agreement of the reviewers after three review stages included a difficulty parameter of b=1/3 and a raw score of 103. The result of this research showed that the item map method can result in higher agreement among reviewers to determine the cut score.

Keywords: Cut off score, modeling, Item-mapping, basic science comprehensive exam

Cite this Article: Farrokhi, N., Habibzadeh, Sh., Minaei, A., & Jalili, M. (2025). The use of RASCH Model (Item-mMapping) in Determining the Standard and Cut-off Score of the Criterion-Reference Tests. *Educational Measurement*, 16(61), 7-30. <https://doi.org/10.22054/jem.2024.37764.1856>



© 2016 by Allameh Tabataba'i University Press
Publisher: Allameh Tabataba'i University Press

Introduction

Assessing students' professional performance is an important part of medical education programs and is very important in the field of medical education. For this purpose, medical students must be able to pass three sections and three comprehensive exams during their studies. By successfully passing the basic science reference test, the scientific competence of medical students to enter the physiopathology and clinical stages is confirmed (Ministry of Health, Treatment and Medical Education, 2014). Determining the standard and cut-off score is considered to be a basic and important, yet difficult and controversial step in order to determine acceptance in criterion-reference tests (Schnabel, 2018; Wang, 2003). There are different types of cut scores (Cizek, 2012; Barron, Sirci and Slater, 2020), the most widely used standard setting method for multiple-choice tests with the aim of obtaining a certificate is the Angoff method (cetin and Gelbal, 2013). A newer method of determining the standard and cut off score that does not include Item-by-Item judgments is the bookmark, Item-mapping method.

In this method, using the question-answer theory of single-parameter Rasch model, the parameter of the individual's ability and the difficulty of the question are estimated. This method was designed to "reduce the cognitive complexity of Angoff' method" (Peterson, Schulz and Engelhard, 2011). In the bookmark method, the judges are given a booklet in which the test questions are arranged from easy to difficult (according to the Item-response Theory (IRT) Model). Then the judges put their mark on a question in this booklet that they think a subject with minimal ability cannot answer the Items after that correctly (Glaser et al., 2017; Hambleton and Pitoniak, 2006; Kang, 2022). Hine and Skaggs (2009) found that the judges faced the problem in the process of determining the bookmark standard that the difficulty of the Items was different from their order in the Item book (the non-order of the Items), so they had trouble placing the mark.

A type of bookmarking method is the Item-mapping method. Wang (2003) in the Item-mapping method, test Items are measured using the Rasch single-parameter measurement model. In this method, a histogram (Item-mapping) is provided to the judges. The histogram starts from the easiest Item on the right side (from the last column on the right side of the histogram) and as we move to the left side of the histogram, the Items become more difficult. After examining the

relevant Items, the judges are asked the following question: Can the candidate with the minimum qualification answer the Item correctly with at least 50% probability? This process is repeated until the judges agree on which Item (or Items) are on the map that the minimum qualified candidate has exactly a 50% chance of answering correctly. In this method, 50% is chosen because when the one-parameter Rasch model is used to grade the Items, the information of the Item is maximized at the probability value of 50%. In this histogram, the Items with the almost similar difficulty parameter are grouped in the unit column. Instead of estimating the probability of the candidates answering each test Item, the judges check all the Items at once. The average number of correct answers is then converted into the minimum score (Wang, 2003).

Research Question(s)

Considering the advantages of the Item-mapping method, this research examines the question of what will be the standard and cut off score of the basic sciences based on the Item-Mapping Method?

Literature Review

In their research, McCann and Stanley (2006) investigated "the use of Rush modeling to determine standards" and used Angoff's common bookmark methods. The results show how informational data can be presented to judges, before and after Angoff measurement. In addition, in this research, an example of the method of determining the bookmark cut score was presented in the form of Item-mapping. The use of non-Rach methods was also used to determine the standard.

In the research entitled "Using Rasch Model in Standard Setting: Item-Mapping Method" conducted by Wang (2003), logical and empirical evidence for using Item-mapping method to determine passing scores for certification exams and multiple-choice tests. Was presented, the results of this research showed that the Item-mapping method provides higher internal group homogeneity.

Buckendahl, et al. (2002) investigated the methods of Angoff and bookmark scores and came to the conclusion that although the Angoff method has been widely used, the bookmark method can help in determining education standards. Emphasize the expectations of students' performance in the test.

In his research, Engelhard (2011) examined the judges' judgments based on the bookmark method and concluded that the use of the Rasch model has an effect on the judges' judgments. Jiao et al. (2011) concluded that the Rush model can help judges in identifying candidates with minimal competence or borderline candidates.

Lewis et al (1996) Stated several issues regarding the bookmark method. A ceiling or floor effect may occur if the test does not include questions that represent the full range of ability level desired for the cut score. In this way, if there are no difficult questions in the exam, the bookmark does not correctly show the content that these students know and can do.

Methodology

The current is a combination of quantitative and qualitative studies. The statistical population of the study consists of all students of basic sciences who participated in comprehensive examinations of medical sciences in September 2015, in the number of 324 people (155 women, 169 men).

The instrument of the research is the comprehensive basic science exam of Qutb-Dah of the country, which was implemented in September 2015 for medical fields in Tehran, ShahidBeheshti, Iran, Baqhiyatullah, Artesh, Behzisty va tavanbakhshy Universities. This exam includes 200 four-choice questions with 1 correct and 0 incorrect scores in the subjects of physiology, pathology, clinical biochemistry, medical physics, clinical psychology, parasitology Mycology, entomology, bacteriology, epidemiology, anatomy, nutrition, embryology, histology, public health, genetics, immunology, English language, Islamic revolution. By using Item-response theory (Item-mapping) in estimating the cutoff score of the comprehensive exam of medical sciences, it is possible to achieve the estimation of question parameters and ability only by using quantitative methods. Therefore, in this research, the test Items were quantitatively analyzed using WINSTEPS 3.92.1 software.

Conclusion

The basic purpose of the tests that are used to grant the certificate is whether the candidates have the appropriate levels of knowledge and skills to perform the necessary activities in their specialized field. One of the most difficult steps in the implementation of these tests is

determining a score that indicates the level of acceptance that indicates competence. It is very important that the passing score reflects the appropriate performance standard, the cut-off score should be able to differentiate between candidates with competence and candidates who have not achieved the appropriate level of competence. The common methods of determining the cutoff score, such as Angoff's method, have problems with the cognitive understanding of candidates with minimal competency and for this reason, and to solve the cognitive deficiency of Angoff's method, the Item-mapping method was used in this research. This method is based on the Item-response theory of the Rasch model and by estimating the difficulty parameter of the Item and the person's ability to identify a question in the test where the probability of the candidates answering that Item correctly is 0/5 (Mitzel et al., 2006).

The issues that the judges mainly face during the cut-off score process include: the correct cognitive understanding of the candidate with the minimum competency and also the accurate knowledge of the cut-off score determination method, which is the method of determining the cut-off score of the Item-mapping, the judges need to understand the concept of the candidate with the least competency. It help the results of this research showed that the Item-mapping method was able to help the judges in the field of determining candidates with minimum competency. The results of the present study are consistent with the results of Hai Wang (2003) and Buckendhal et al. (2002). Considering that the Item-mapping uses the Rasch model to determine the difficulty parameter and the ability parameter, it is able to put this Item difficulty parameter and the ability parameter on the same scale. Also, in the Item-mapping method, individual questions are not examined and all Items are examined together. In this method, the passing grade was determined by identifying a point on the histogram of the difficulty parameter of the Item, which showed that the candidate with the minimum competency had a 50% chance to provide the correct answer to the Item, and considering that in determining the standard and cutoff score, determining the level of ability The candidate has a minimum competency of 50/. The best point for the judges to determine the cut-off score is that the Item-mapping method with emphasis on the Rush model provides this estimate to the judges, which helps to determine the cut-off score more accurately. The results of this research are also consistent with the results of Schnabel (2018).

In the process of determining the standard for the basic science comprehensive exam, during three stages (rounds), the exam Items were judged and evaluated. In the first and second stages of judging, there was no agreement between the judges about the cut-off score and the standard (Item that the candidate with the minimum competency is not able to answer the Items after) but in the third stage (round), by providing the Item-mapping The judges had a high consensus on the difficulty parameter of $1/3$ and specified this point as the cut-off point and standard of the test. At this cut-off point, the ability parameter was $1/31$, and the score of the subject with the minimum competency was 103, and the acceptance percentage of the candidates was estimated at 65/84. Actually, based on this finding, candidates with difficulty parameter less than $1/3$, ability parameter less than $1/31$ and raw score lower than 103 are rejected. The study of Buckendhal et al. (2002) and Schnabel (2018) and Engelhard (1996) also showed that although Angoff's method has been widely used, the Item-mapping method can be used in the cut off score and performance expectations of candidates in certification exams. Those involved in education should help. Among the limitations of this research, it can be pointed out that because the Item-mapping method was used for the first time to determine the standard in medical sciences, the judges had no prior knowledge of this method.

In order to determine the optimal cut-off score, it is suggested to conduct a comparative study of the method of determining the standard and the cut-off score of the Item-mapping with methods such as Angoff and the traditional method. Also, future researches can examine the impact of alternative values of 50% probability with other common values such as 0/67 in determining the cutoff score. It is suggested that in future researches, a judging session with two groups of expert judges or more should be held and the results of the two groups should be compared with each other.

کاربرد مدل راش (نقشه سؤال) در تعیین استاندارد و نمره برش آزمون‌های ملاک مرجع

نورعلی فرخی *

نویسنده مسئول، استاد، گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی،
تهران، ایران.. رایانامه: farrokhinoorali@gmail.com

شیرین دخت حبیب زاده

مربی گروه علوم تربیتی، دانشگاه فرهنگیان، تهران، ایران. رایانامه:
habibzadehshirindokht@yahoo.com

اصغر مینایی

دانشیار، گروه سنجش اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران.
رایانامه: minaeiasghar@yahoo.com

محمد جلیلی

استاد، گروه طب اورژانس و گروه آموزش پزشکی، دانشگاه علوم پزشکی
تهران، تهران، ایران. رایانامه: mjalili@tums.ac.ir

چکیده

تعیین استاندارد قبولی و نمره برش برای قبولی یا رد در آزمون‌های ملاک مرجع امری ضروری است. در سال‌های اخیر نظریه پرسش-پاسخ (IRT) در ساخت و اجرای آزمون‌ها کمک شایانی نموده است. هدف از پژوهش حاضر، تعیین استاندارد و نمره برش آزمون ملاک مرجع جامع پایه علوم پزشکی با استفاده از مدل راش، روش نقشه سؤال بود. جامعه آماری این پژوهش کلیه داوطلبان آزمون جامع علوم پایه پزشکی قطب ده کشوری بودند که تعداد آنان ۳۲۴ نفر بود و پاسخ کلیه داوطلبان در این آزمون جهت تعیین استاندارد و نمره برش مورد تحلیل قرار گرفته است. ابزار این پژوهش آزمون جامع علوم پایه با ۲۰۰ سؤال چهارگزینه‌ای بوده است که در شهریورماه ۱۳۹۵ اجرا شد. پژوهش حاضر از نوع پژوهش‌های مقطعی و برای تحلیل سؤالات از نرم‌افزار WINSTEPS استفاده گردید. یافته‌های پژوهش نشان داد که این آزمون از اعتبار سؤال ۰/۹۸ و اعتبار فرد ۰/۹۲ برخوردار بوده که نشان می‌دهد واریانس نمونه و طول تست مناسب بوده و سؤالات از حوزه‌های نه‌گانه به تناسب انتخاب شده است. هم‌چنین آزمون با مدل تک پارامتری راش برازش دارد. اجرای پل داوران و بررسی سؤالات با روش نقشه سؤال نشان داد توافق داوران پس از سه مرحله داوری شامل پارامتر دشواری برابر ۱/۳ و نمره خام برابر ۱۰۳ بوده است. یافته‌های این پژوهش نشان داد که استفاده از روش نقشه سؤال توانسته است توافق بالایی بین داوران نسبت به تعیین نمره برش ایجاد نماید.

کلیدواژه‌ها: نمره برش، مدل راش، نقشه سؤال، ملاک مرجع، آزمون جامع علوم پایه

استناد به این مقاله: فرخی، نورعلی، حبیب زاده، شیرین دخت، مینایی، اصغر، و جلیلی، محمد. (۱۴۰۴). کاربرد مدل راش (نقشه سؤال) در تعیین استاندارد و نمره برش آزمون‌های ملاک مرجع. فصلنامه اندازه‌گیری تربیتی، ۱۶(۶۱)، ۷-۳۰

<https://doi.org/10.22054/jem.2024.37764.1856>

© ۲۰۱۶ دانشگاه علامه طباطبائی

ناشر: دانشگاه علامه طباطبائی



مقدمه

سنجش عملکرد حرفه‌ای دانشجویان بخش مهمی از برنامه‌های آموزش پزشکی بوده و در حوزه آموزش پزشکی از اهمیت بسیاری برخوردار است. به همین منظور دانشجویان رشته‌های پزشکی در طول دوران تحصیل باید بتوانند از سه مقطع و سه آزمون جامع ملاک مرجع عبور نمایند. با کسب موفقیت در آزمون ملاک مرجع علوم پایه، صلاحیت علمی دانشجویان پزشکی برای ورود به مراحل فیزیوپاتولوژی و بالینی تأیید می‌شود (وزارت بهداشت، درمان و آموزش پزشکی ۱۳۹۶). تعیین استاندارد^۱ و نمره برش^۲ جهت تعیین قبولی در آزمون‌های ملاک مرجع^۳ یک گام اساسی و مهم و درعین حال دشوار و بحث‌برانگیز به حساب می‌آید (Wang, 2003; Schnabel, 2018). تعیین قبولی/شکست یا سنجش عملکردهایی که بر اساس نمرات برش انجام می‌شود، نه تنها بر پیشرفت تحصیلی فردی دانش‌آموزان تأثیر می‌گذارد، بلکه بر پیشرفت موسسه آموزشی، کشور و کل سیستم آموزشی نیز تأثیر می‌گذارد (Cetin & Gelbal, 2013). به همین دلیل بسیار مهم است که نمره برش بتواند استاندارد عملکرد مناسبی را نشان دهد، این نمره باید بتواند داوطلبان با شایستگی حرفه‌ای را از داوطلبانی که شایستگی لازم را کسب ننموده‌اند تمیز دهد (Cizek, 2012; Berk, 1996).

هدف اساسی در آزمون‌هایی که جهت اعطای گواهینامه به کار می‌روند، این است که آیا داوطلبان، دارای سطوح مناسب دانش و مهارت اجرای فعالیت‌های لازم در حوزه تخصصی خود هستند؟ یکی از دشوارترین مراحل در اجرای این آزمون‌ها، تعیین نمره‌ای است که سطح قبولی را که نشان‌دهنده شایستگی است، نشان دهد. نمره برش باید بتواند میان داوطلبان با شایستگی و داوطلبانی که سطح شایستگی مناسبی را احراز نکرده‌اند، تفاوت قائل شود (Mitzel et al., 2006).

نمرات برش انواع گوناگونی دارند (Cizek, 2012; Baron et al., 202)، پرکاربردترین روش تعیین استاندارد برای آزمون‌های چندگزینه‌ای با هدف اخذ گواهینامه، روش آنگوف^۴ است (عباسی، ۱۳۹۳; Cetin & Gelbal, 2013; Kane, 1995). در این روش پس از تشکیل پانلی از داوران متخصص، هر داور به صورت مستقل به این سؤال پاسخ می‌دهد که: چقدر

1. Standard Setting
2. Cut off scores
3. Criterion-Reference
4. Angoff

احتمال دارد یک دانشجو با حداقل صلاحیت علمی بتواند به یک سؤال پاسخ صحیح بدهد؟ سپس میانگین احتمالاتی که داورهای مختلف به هر سؤال داده‌اند محاسبه می‌شود و میانگین استانداردهای تمامی سؤالات محاسبه می‌گردد (Alshawwa, 2023; Wang, 2003; Buckendahl et al., 2002؛ مرتاض هجری، ۱۳۹۰). دو انتقاد عمده به روش آنگوف مطرح گردیده است: ۱- در فرایند تعیین نمره برش داوران به‌درستی مفهوم داوطلب با حداقل صلاحیت را درک نمی‌کنند و ۲- برآورد داوران در مورد پارامتر دشواری سؤال بر اساس داوطلب با حداقل صلاحیت درست نبوده و در برخی موارد ناهم‌سو هستند (Shepard, 1995; MacCann, 2019). ایراد عمده این روش عدم درک شناختی داوران در برآورد درست احتمال پاسخ صحیح داوطلبان با حداقل صلاحیت به سؤالات آزمون است (Mckinley et al., 1996).

روش جدیدتر تعیین استاندارد و نمره برش که شامل داورهای سؤال به سؤال نمی‌شود، روش بوک مارک^۱ است. در این روش با استفاده از نظریه سؤال-پاسخ مدل تک پارامتری راش، پارامتر توانایی فرد و دشواری سؤال برآورد می‌شود. این روش به این منظور طراحی شد تا از پیچیدگی‌های شناختی روش آنگوف کاسته گردد (Peterson et al., 2011). در روش بوک‌مارک، به داوران کتابچه‌ای داده می‌شود که در آن سؤالات آزمون از آسان تا دشوار به ترتیب آورده شده است (مطابق مدل نظریه پرسش-پاسخ (IRT)). سپس داوران نشانه و علامت خود را بر روی سؤالی در این کتابچه قرار می‌دهند که به نظر آنان آزمون‌دنی با حداقل توانایی^۲ به سؤالات پس‌از آن به‌درستی نمی‌تواند پاسخ دهد (Glaser, 2017; Hambleton & Pitoniak, 2006; Kang, 2022). Hein and Skaggs (2009) دریافتند که داوران در روند تعیین استاندارد بوک‌مارک با این مشکل مواجه می‌شوند که مسئله دشواری سؤالات متفاوت از ترتیب آن‌ها در کتابچه سؤالات است (غیرترتیبی بودن سؤال‌ها)، از این رو در قراردادن نشانه به مشکل برمی‌خورند.

یک نوع از روش بوک‌مارک، روش نقشه سؤال^۳ است. Wang (2003) در روش نقشه سؤال، سؤالات آزمون با استفاده از مدل اندازه‌گیری تک پارامتری راش اندازه‌گیری می‌شوند. در این روش هیستوگرامی (نقشه سؤال) در اختیار داوران گذاشته می‌شود.

1. Bookmark

2. Minimum Competency Candidate

3. Item Mapping

هیستوگرام از آسان‌ترین سؤال در سمت راست آغاز می‌شود (از آخرین ستون سمت راست هیستوگرام) و هرچه به سمت چپ هیستوگرام پیش می‌رویم سؤالات دشوارتر می‌شوند. پس از بررسی سؤالات آزمون، از داوران این سؤال پرسیده می‌شود که: آیا داوطلب با حداقل صلاحیت، دست کم به احتمال ۵۰ درصد می‌تواند به سؤال پاسخ درست دهد؟ این فرآیند آن قدر تکرار می‌شود تا داوران به این اتفاق نظر برسند که چه سؤال (یا چه سؤالاتی) در نقشه سؤال قرار گرفته که داوطلب با حداقل صلاحیت دقیقاً ۵۰ درصد شانس آن را دارد که به آن پاسخ صحیح دهد. در این هیستوگرام، سؤالات با پارامتر دشواری تقریباً مشابه در ستونی واحد گروه‌بندی می‌شوند. داوران به جای برآورد احتمال پاسخگویی داوطلبان برای هر سؤال آزمون، به یکباره تمام سؤالات را بررسی می‌کنند. میانگین تعداد پاسخ‌های درست سپس به حداقل نمره تبدیل می‌شود (Wang, 2003).

مدل تک پارامتری راش، مبنای نظری روش تعیین نمره برش به روش نقشه سؤال است. پاسخ داوطلب به سؤالات آزمون تحت تأثیر دو عامل قرار دارد: توانایی غیرقابل مشاهده داوطلب و دشواری سؤالات (Wang, 2003). در مدل راش برای ارزیابی عملکرد آزمودنی در آزمون و پارامتر دشواری سؤال‌ها از رویکردی متفاوت از نظریه پرش-پاسخ استفاده می‌شود. این رویکرد در نهایت پارامتر توانایی آزمودنی‌ها و پارامتر دشواری سؤال‌ها را در مقیاسی مشترک قرار می‌دهد که به واحدهای آن‌ها لوجیت اطلاق می‌شود. مقیاس لوجیت فواصل موجود در این زنجیره را حفظ می‌کند و به شکل‌گیری ویژگی جمع‌پذیری منتهی می‌شود (Schanbel, 2018). در این مدل احتمال اینکه داوطلبی با توانایی θ پاسخ صحیحی به سؤال دهد را می‌توان با تابع ریاضی زیر مدل‌سازی نمود.

$$p(\theta) = \frac{e^{(\theta-b)}}{1 + e^{(\theta-b)}}$$

در این فرمول $p(\theta)$ برابر احتمالی است که داوطلب با توانایی θ به سؤال خاصی پاسخ درست دهد. پارامتر θ معادل توانایی داوطلب و b پارامتر دشواری سؤال بر اساس مدل راش است (Schnabel, 2018؛ Boone et al., 2014). در این مدل وقتی توانایی داوطلب برابر با پارامتر دشواری سؤال باشد، احتمال پاسخ صحیح به سؤال ۵۰ درصد است. در مدل تک پارامتری راش، آگاهی سؤال در هنگامی که احتمال موفقیت ۵۰ درصد باشد به حداکثر می‌رسد؛ و باعث می‌گردد تا در فرایند تعیین استاندارد، نه برشی انتخاب گردد که به بهترین

نحو می‌تواند بین داوطلب با حداقل صلاحیت با توانایی پایین‌تر از ۵۰ درصد و داوطلب با توانایی پاسخگویی درست به سؤال با احتمال بالای ۵۰ درصد، تیز قائل گردد (Hambleton & Pitoniak, 2006). مهم‌ترین دلیل برای استفاده از مدل راش را می‌توان به همین مسئله و سهولت در کاربرد آن در تعیین نمره برش دانست. همچنین مزیت دیگر استفاده از روش نقشه سؤال، تیم گرافیکی سؤالات بر اساس پارامتر دشواری آنان است. رابطه‌ی بین پارامتر توانایی داوطلب و پارامتر سؤال به داوران در تعیین نمره برش کمک شایانی می‌کند. روش نقشه سؤال دو مشکل عمده روش آنگوف را مرتفع نموده است. این روش به درک شناختی بهتر داوران نسبت به داوطلب با حداقل صلاحیت و هم‌چنین به داوران در درک و برآورد احتمال موفقیت داوطلبان به سؤالات آزمون کمک می‌نماید (Wang, 2003).

Wang (2003) در پژوهش خود جهت تعیین نمره برش از روش نقشه سؤال استفاده نموده و بیان کرد که استفاده از روش نقشه سؤال به درک داوران نسبت به داوطلب با حداقل توانایی کمک می‌کند و هم‌چنین در تعیین احتمال پاسخ صحیح داوطلب با حداقل صلاحیت نیز توافق بیشتری بین داوران ایجاد می‌شود. Schnabel (2018) در پژوهشی به تعیین نمره برش به روش نقشه سؤال پرداخت و بیان کرد که داوران فرایند تعیین احتمال را به‌خوبی درک کرده‌اند. Hein and Skaggs (2009) روش نقشه سؤال را به‌مراتب شهودی‌تر از روش‌های دیگر خصوصاً روش آنگوف دانسته و داوران با استفاده از این روش توانسته‌اند داوطلب با حداقل صلاحیت را به‌خوبی شناسایی نمایند.

آموزش مبتنی بر کسب شایستگی حرفه‌ای که هدف آن افزایش دانش‌ها و مهارت‌ها و نگرش‌های دانشجویان در رشته پزشکی است، از اهمیت بسیاری در حوزه آموزش برخوردار است. آموزش بر پایه کسب شایستگی به دانشجویان کمک می‌نماید تا بتوانند خدمات مناسب حرفه‌ای به جامعه ارائه نمایند (Epstein & Hundert, 2002). برنامه دوره دکتری عمومی دانشگاه‌های علوم پزشکی نیز شامل سه دوره مهم عمومی و علوم پایه، نشانه‌شناسی و فیزیوپاتولوژی، کارآموزی و کارورزی است. به دلیل حساسیت بالای این رشته و تأثیر مستقیم توانمندی‌های حرفه‌ای دانشجویان این رشته بر سلامت جسمانی و روانی جامعه؛ لذا در طول دوران تحصیل، دانشجویان رشته‌های پزشکی باید بتوانند از این سه مقطع با کسب موفقیت در سه آزمون عبور نمایند. مقطع علوم پایه در پنج نیمسال اول تحصیلی ارائه می‌گردد و کسب موفقیت در امتحان جامع علوم پایه، مجوز ورود به مراحل

فیزیوپاولوژی و بالینی است. آزمون جامع علوم پایه با ۲۰۰ سؤال چهارگزینه‌ای برگزار می‌شود. چنانچه دانشجویان بعد از سه بار شرکت در این امتحان موفق به اخذ نمره قبولی نشوند، مردود شده و از ادامه تحصیل در این رشته بازمی‌مانند که پیامدهایی چون اتلاف وقت و هزینه توسط دانشجو و دانشگاه، افت سطح علمی دانشگاه‌ها و محرومیت از ادامه تحصیل دانشجویان را به همراه خواهد داشت (وزارت بهداشت و درمان آموزش پزشکی، ۱۳۹۶). لذا این مسئله لزوم توجه و دقت در طراحی آزمون و تعیین استاندارد و نمره برش دقیق جهت تعیین صلاحیت و شایستگی داوطلب را نشان می‌دهد.

آزمون‌های علوم پایه پزشکی با هدف تصمیم‌گیری در مورد ارتقاء دانشجویان به مقطع بالاتر تحصیلی طراحی و اجرا می‌گردد، به همین جهت تعیین نمره‌ای در آزمون که آزمودنی‌های را به‌درستی از هم تفکیک کند، بسیار حائز اهمیت است. انتخاب دقیق این نمره که جزو مهم‌ترین اهداف آزمون‌های ملاک مرجع است، می‌تواند به طراحان آزمون‌ها در تعیین میزان شایستگی دانشجویان در طول دوران تحصیل کمک نماید. در روش‌های تعیین نمره برش که تاکنون در دانشگاه‌های علوم پزشکی در داخل کشور اجرا گردیده است، از روش سنتی تعیین استاندارد و نمره برش استفاده شده است (عباسی، ۱۳۹۳). تعیین حدنصاب قبولی یا نمره برش به شیوه‌ای موجود، برای آزمون‌های ملاک مرجع مناسب نبوده است و بیشتر با هدف مقایسه افراد با یکدیگر کاربرد دارد. لذا با توجه به مزیت‌هایی که روش نقشه سؤال دارد، این پژوهش به بررسی این سؤال می‌پردازد که تعیین استاندارد و نمره برش آزمون علوم پایه بر اساس روش نقشه سؤال چه نمره‌ای خواهد بود؟

پیشینه پژوهش

MacCann and Stanley (2006) در پژوهش خود به بررسی «به‌کارگیری مدل‌سازی راش برای تعیین استاندارد» پرداخته و از روش‌های متداول آنگوف و بوک مارک استفاده کردند. نتایج نشان می‌دهد چگونه داده‌های اطلاعاتی را می‌توان برای داوران، قبل و بعد از سنجش آنگوف، ارائه نمود. به‌علاوه در این پژوهش یک نمونه روش تعیین نمره برش بوک مارک به شکل نقشه سؤال ارائه گردید. به‌کارگیری روش‌های غیر راش نیز برای تعیین استاندارد بکار گرفته شد.

در پژوهش با عنوان «استفاده از مدل راش در تعیین استاندارد: روش نقشه سؤال» که توسط Wang (2003) انجام شد، شواهد منطقی و تجربی برای به‌کارگیری روش نقشه سؤال

جهت تعیین نمرات قبولی برای امتحانات گواهینامه و آزمون‌های چندگزینه‌ای ارائه شد. نتایج این پژوهش نشان داد که روش نقشه سؤال، همسانی درونی گروهی بالاتری را ارائه می‌نماید. Buckendahl و همکاران (2002) نیز روش‌های نمره برش آنگوف و بوک‌مارک را بررسی نمودند و به این نتیجه رسیدند که گرچه روش آنگوف به صورت گسترده مورد استفاده قرار گرفته است اما روش بوک‌مارک می‌تواند در تعیین استانداردهای آموزش کمک نماید تا بر روی انتظارات عملکرد دانش‌آموزان در آزمون تأکید کنند.

Engelhard (2011) در پژوهش خود قضاوت‌های داوران بر اساس روش بوک‌مارک را مورد بررسی قرار داد و به این نتیجه رسید که استفاده از مدل راش بر قضاوت داوران تأثیر دارد. Jiao و همکاران (2011) به این نتیجه دست یافتند که مدل راش می‌تواند در شناسایی داوطلبان با حداقل شایستگی یا داوطلبان با حداقل صلاحیت به داوران کمک نماید.

Lewis و همکاران (1996) چندین مسئله را در مورد روش بوک‌مارک بیان کردند. اگر آزمون شامل سؤالاتی که محدوده کامل سطح توانایی موردنظر نمره برش را نشان دهد، نباشد، ممکن است اثر سقف یا کف اتفاق افتد. به این صورت که اگر در آزمون سؤالات دشوار وجود نداشته باشد، بوک‌مارک به درستی محتوایی که این دانش‌آموزان می‌دانند و می‌توانند انجام دهند را نشان نمی‌دهد.

روش

پژوهش حاضر در زمره مطالعات مقطعی بوده و ترکیبی از مطالعات کمی و کیفی است. جامعه آماری پژوهش عبارت از کلیه دانشجویان علوم پایه شرکت کننده در آزمون‌های جامع علوم پزشکی شهریورماه ۱۳۹۵ به تعداد ۳۲۴ نفر (۱۵۵ نفر زن، ۱۶۹ نفر مرد) است.

ابزار پژوهش عبارت از آزمون جامع علوم پایه قطب ده کشوری است که در شهریورماه ۱۳۹۵ برای رشته‌های پزشکی در دانشگاه‌های علوم پزشکی تهران، شهید بهشتی، ایران، بقیه‌الله، ارتش، بهزیستی و توان‌بخشی اجرا می‌گردید، بود. این آزمون شامل ۲۰۰ سؤال چهارگزینه‌ای با نمره‌گذاری صحیح ۱ و غلط ۰ در موضوعات فیزیولوژی، آسیب‌شناسی، بیوشیمی بالینی، فیزیک پزشکی، روان‌شناسی بالینی، انگل‌شناسی، قارچ‌شناسی، حشره‌شناسی، باکتری‌شناسی، اپیدمیولوژی، تشریح، تغذیه، جنین‌شناسی، بافت‌شناسی، بهداشت عمومی، ژنتیک، ایمنی‌شناسی، زبان انگلیسی، انقلاب اسلامی و ریشه‌های آن است. در پژوهش حاضر با توجه به به کارگیری نظریه سؤال- پاسخ (نقشه سؤال) در برآورد نمره

برش آزمون جامع علوم پزشکی، دستیابی به برآورد پارامترهای سؤال و توانایی تنها با استفاده از روش‌های کمی امکان‌پذیر است. لذا در این پژوهش، ابتدا با استفاده از نرم‌افزار WINSTEPS 3.92.1 سؤالات آزمون به شیوه کمی مورد تحلیل تجزیه و تحلیل قرار گرفت.

جدول ۱. برآورد اعتبار سؤالات آزمون و فرد بر اساس مدل تک پارامتری راش

اعتبار سؤال	اعتبار شخص	شاخص جداسازی شخص	ضریب جداسازی سؤالات	ضریب آلفای کرونباخ
۰/۹۸	۰/۹۲	۳/۳۶	۷/۴۸	۰/۹۲

همان‌گونه که در جدول ۱ ملاحظه می‌شود اعتبار سؤال آزمون جامع علوم پایه پزشکی برابر ۰/۹۸ برآورد شد که نشان می‌دهد نمونه در حد قابل قبول بوده و پارامتر دشواری سؤال از گستردگی لازم برخوردار است. همچنین در این پژوهش ضریب اعتبار شخص برابر ۰/۹۲ است. که نشان می‌دهد که واریانس نمونه و طول تست مناسب بوده و از انواع سؤالات حوزه‌های مختلف به اندازه‌ی کافی نمونه سؤال موجود است. شاخص جداسازی شخص^۱ محاسبه شده در این آزمون نیز برابر ۳/۳۶ برآورد شد که نشان‌دهنده توانایی آزمون در جداسازی و تمیز آزمودنی‌ها است. بدین ترتیب این آزمون در تمیز سؤالات خرده آزمون‌های نوزده گانه توانمند عمل کرده است. همچنین ضریب آلفای آزمون جامع برابر ۰/۹۲ است که نشان‌دهنده پایایی مطلوب آزمون است.

جدول ۲. آماره‌های MNSQ و ZSTD آزمون جامع علوم پزشکی

آماره	I		Outfit	
	ZSTD	MNSQ	ZSTD ^۲	MNSQ ^۳
میانگین	۰	۱	۰	۱
انحراف استاندارد	۱/۵	۱,	۱/۶	۰/۱۷
پیشینه	۴/۵	۱/۴۲	۴/۷	۱/۸۲
کمینه	-۳/۷	۰/۷۹	-۳/۴	۰/۶۴

در بررسی برازش سؤالات آزمون با استفاده از نرم‌افزار WINSTEPS دو آماره MNSQ و ZSTD محاسبه می‌گردد. آماره MNSQ در دامنه‌ای از صفر تا بی‌نهایت است که جهت

1. Person Separation Index.
2. Z standard
3. Mean square

برازش سؤالات آزمون این آماره باید یک باشد. همان‌طور که در جدول ۲ مشاهده می‌شود مقدار مشاهده‌شده برای این آماره در INFIT و OUTFIT برابر یک است. همچنین آماره ZSTD نیز آماره معناداری آماری انحراف داده‌ها از مدل راش را آزمون می‌کند (مینائی، ۱۳۹۳)، که با توجه به جدول ۲ مقدار مشاهده‌شده این آماره در هر دو حالت (INFIT و Outfit) برابر صفر است. مقادیر به‌دست‌آمده نشان می‌دهد که سؤالات آزمون با مدل تک پارامتری راش (نظریه پرسش - پاسخ) برازش دارند.

پس از بررسی شاخص‌های برازش آزمون و محاسبه پارامترهای دشواری و توانایی سؤالات، با استفاده از روش‌های کیفی و بهره‌گیری از نظر متخصصان، نمره برش سؤالات مشخص شد. بدین منظور با توجه به اینکه اجرای آزمون جامع علوم پایه در شهریورماه ۱۳۹۵ بر عهده دانشگاه علوم پزشکی تهران بوده است، لذا برای بررسی نمره برش، فهرستی شامل ۱۸ نفر از اعضای هیئت‌علمی دانشگاه علوم پزشکی تهران، شناسایی، انتخاب و دعوت به همکاری شدند. در انتخاب ایشان تلاش شد تا بیشترین تنوع از نظر تخصص، سابقه آموزشی، سن در نظر گرفته شود. جلسه تعیین استاندارد به‌صورت کارگاهی نیم‌روزه (۶ ساعت) برگزار شد. از این تعداد ۱۴ نفر از داوران تا پایان جلسه تعیین استاندارد و نمره برش شرکت داشتند. در ابتدای جلسه داوران با مفاهیم تعیین استاندارد و نمره برش و اهمیت آن در آزمون‌های پزشکی آشنا شدند، سپس به بررسی توانایی‌های داوطلب با حداقل شایستگی پرداختند، آنگاه دفترچه سؤالات آزمون جامع علوم پایه پزشکی که برحسب پارامتر دشواری از آسان‌ترین تا دشوارترین سؤال تنظیم شده بود در اختیار داوران قرار گرفت. داوران به بررسی سؤالات پرداخته و به این سؤال پاسخ دادند که «کدام صفحه و کدام سؤال است که داوطلب با حداقل صلاحیت می‌تواند به آن پاسخ دهد و به سؤالات بعد از آن نمی‌تواند پاسخ دهد؟». در ذیل نمونه‌ای از سؤالات آزمون دفترچه جدید ارائه شده است.

شکل ۱. نمونه‌ای از سؤالات دفترچه آزمون جامع علوم پایه برحسب پارامتر دشواری

سوال شماره: ۱۱۷
شماره صفحه: ۹۰

- کدام یک از هسته‌های زیر، در عمق آنکوس قرار دارد؟

- الف) هسته سیاه
- ب) هسته دم‌دار
- ج) هسته آمیگدال
- د) هسته عدسی

کلید: ج

همان‌طور که در شکل ۱ مشاهده می‌شود، در دفترچه جدید، شماره سؤال، شماره صفحه سؤال در دفترچه آزمون جامع علوم پایه (قطب ۱۰ کشوری)، متن سؤال، گزینه‌ها و کلید آن ارائه شده است و داوران هر یک از سؤالات را از ابتدای دفترچه مطالعه کردند و نقطه برش آزمون را سؤالی قرار دادند که از نظر آن‌ها داوطلب با حداقل صلاحیت قادر به پاسخگویی به سؤالات بعد از آن نیست.

در مرحله بعدی (دور دوم داوری)، سؤالات آزمون جامع علوم پایه پزشکی برحسب پارامتر دشواری از ساده به دشوار مرتب شدند و داوران علاوه بر سؤال، ویژگی پارامتر دشواری سؤالات را بر اساس مدل تک پارامتری راش ملاحظه نمودند و مجدداً سؤالی را انتخاب نمودند که داوطلب با حداقل صلاحیت به سؤالات تا قبل از آن به درستی پاسخ داده و به سؤالات بعد از آن نمی‌تواند پاسخ دهد.

در مرحله سوم (دور سوم داوری)، نمودار هیستوگرام سؤالات (نقشه سؤال)، و جدول ۴ در اختیار داوران جهت تعیین نقطه برش و استاندارد قرار گرفت.

یافته‌ها

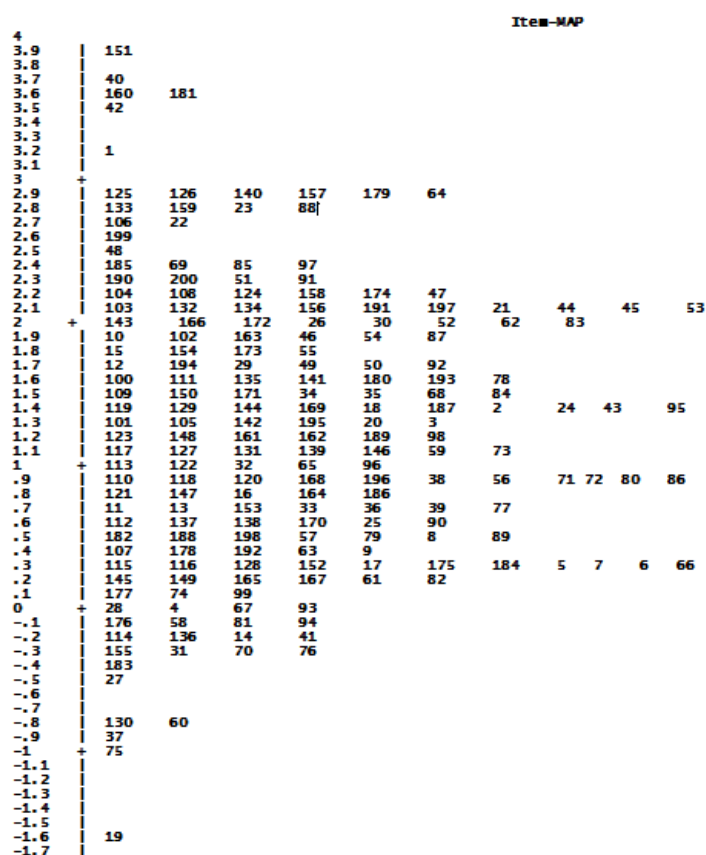
نتایج حاصل از مرحله اول و دوم داوری در جدول زیر آمده است.

جدول ۳. نقطه برش یا استاندارد از منظر داوران در مرحله اول و دوم تعیین نمره برش

داور	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
مرحله اول	۱۶۸	۱۰۰	۱۱۷	۱۰۵	۵۲	۶۶	۳۴	۸۹	۲	۷۷	۱۰۹	۸۳	۸۷	۱۱۹
مرحله دوم	۷۳	۱۶۲	۹۸	۱۱۹	۱۰۵	۱۱	۷۱	۸۴	۱۲۹	۹۶	۱۶۱	۱۲	۸۷	۱۶۹

همان‌طور که در جدول ۳ مشاهده می‌شود، هر یک از ۱۴ داور، در مرحله اول داوری و پس از مشاهده دفترچه، سؤالات متفاوتی را به‌عنوان نقطه برش (استاندارد) مشخص کردند. در مرحله دوم داوری و پس‌ازاین که داوران پارامتر دشواری مدل راش را ملاحظه کردند، نمره برش داوران تغییر یافت. همان‌گونه که در جدول ۳ مشخص است به‌جز داور ۱۳؛ داوران سؤال متفاوتی را نسبت به مرحله اول انتخاب نمودند. آسان‌ترین سؤال آزمون، سؤال شماره ۱۹ با پارامتر دشواری ۵۸/۱- و دشوارترین سؤال، سؤال شماره ۱۵۱ با پارامتر دشواری ۸۷/۳ بود. در مرحله سوم، داوران بر اساس نقشه سؤال و جدول برآورد میانگین پارامتر توانایی، درصد قبولی دانشجویان و نمره خام معادل با درجه دشواری سؤال به تعیین نمره برش مشترک پرداختند.

نمودار ۱. نمودار نقشه سؤال آزمون علوم پایه بر اساس مدل تک پارامتری راش



در نمودار نقشه سؤال ۱، توزیع سؤالات آزمون برحسب پارامتر دشواری نمایش داده شده است. سؤالاتی که از نظر پارامتر دشواری در یک سطح هستند، بر روی یک محور قرار گرفته‌اند. همان‌طور که مشاهده می‌شود، سؤال ۱۵۱ با پارامتر دشواری ۳/۸۷، دشوارترین سؤال و سؤال ۱۹ با پارامتر دشواری ۱/۶ ساده‌ترین سؤال در این آزمون است. پارامترهای دشواری ۰/۳، ۰/۹، ۱/۴ و ۲/۱ نیز دارای بیشترین فراوانی سؤالات هستند.

در مرحله‌ی سوم داوری علاوه بر نمودار نقشه سؤال ۱، جدول شماره ۴ نیز در اختیار داوران قرار گرفت. در این جدول میانگین درجه دشواری در هر ستون به همراه میانگین توانایی برآورد شده مدل راش، نمره خام معادل پارامتر دشواری و همچنین درصد قبولی داوطلبان در هر ستون؛ را در اختیار داوران قرار داده شد.

جدول ۴. درجه دشواری، میانگین پارامتر توانایی، نمره خام معادل با درجه دشواری، درصد قبولی دانشجویان در آزمون جامع علوم پایه شهریور ۱۳۹۵

ردیف	پارامتر دشواری	میانگین پارامتر توانایی	نمره خام معادل	درصد قبولی دانشجویان
۱	۳/۹	۳/۸۷	----	۰
۲	۳/۸	-----	----	۰
۳	۳/۷	۳/۶۷	----	۰
۴	۳/۶	۳/۶۱	----	۰
۵	۳/۵	۳/۴۷	----	۰
۶	۳/۴	----	----	۰
۷	۳/۳	----	----	۰
۸	۳/۲	۳/۱۸	----	۰
۹	۳/۱	----	----	۰
۱۰	۳	----	----	۰
۱۱	۲/۹	۲/۸۹	۱۶۰	۰/۶۱
۱۲	۲/۸	۲/۸۰	۱۵۸	۱/۲۳
۱۳	۲/۷	۲/۷۳	۱۵۵	۳/۳۸
۱۴	۲/۶	۲/۶۲	۱۵۲	۴
۱۵	۲/۵	۲/۴۹	۱۴۸	۵/۸۴
۱۶	۲/۴	۲/۴۱	۱۴۵	۸/۶۱
۱۷	۲/۳	۲/۲۸	۱۴۱	۱۰/۷۶

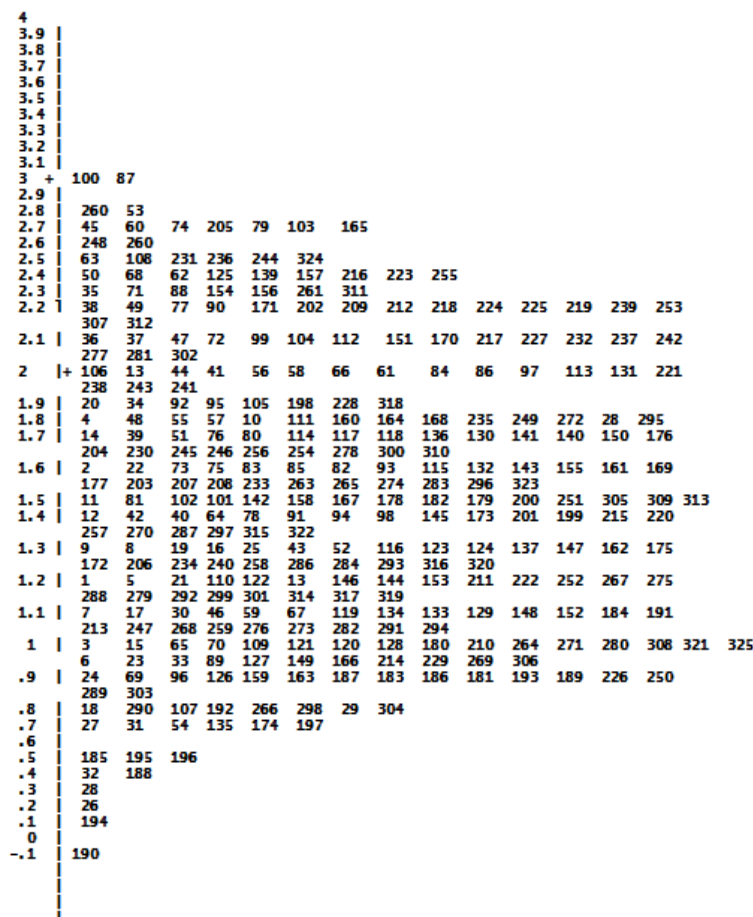
ردیف	پارامتر دشواری	میانگین پارامتر توانایی	نمره خام معادل	درصد قبولی دانشجویان
۱۸	۲/۲	۲/۱۸	۱۳۸	۱۵/۶۹
۱۹	۲/۱	۲/۱۰	۱۳۴	۲۰/۹۲
۲۰	۲	۱/۹۹	۱۳۰	۲۶/۱۵
۲۱	۱/۹	۱/۸۸	۱۲۶	۲۸/۶۱
۲۲	۱/۸	۱/۸۱	۱۲۴	۳۲/۹۲
۲۳	۱/۷	۱/۷۲	۱۲۰	۴۰
۲۴	۱/۶	۱/۵۸	۱۱۴	۴۷/۶۹
۲۵	۱/۵	۱/۴۸	۱۱۰	۵۲/۳۰
۲۶	۱/۴	۱/۴۰	۱۰۷	۵۸/۴۶
۲۷	۱/۳	۱/۳۱	۱۰۳	۶۵/۸۴
۲۸	۱/۲	۱/۱۸	۹۸	۷۲/۶۱
۲۹	۱/۱	۱/۰۹	۹۴	۷۹/۶۹
۳۰	۱	۱/۰۲	۹۱	۸۸
۳۱	۰/۹	۰/۸۹	۸۶	۹۲/۹۲
۳۲	۰/۸	۰/۸۲	۸۳	۹۵/۳۸
۳۳	۰/۷	۰/۶۹	۷۸	۹۷/۲۳
۳۴	۰/۶	۰/۵۹	۷۴	۹۷/۲۳
۳۵	۰/۵	۰/۴۹	۷۰	۹۸/۱۵
۳۶	۰/۴	۰/۴۰	۶۷	۹۸/۷۶
۳۷	۰/۳	۰/۳۰	۶۳	۹۹/۰۷
۳۸	۰/۲	۰/۱۹	۶۰	۹۹/۳۸
۳۹	۰/۱	۰/۱۱	۵۷	۹۹/۶۹
۴۰	۰	۰/۱	۵۳	۹۹/۶۹
۴۱	-۰/۱	-۰/۱۱	۴۹	۱۰۰

در جدول ۴ برای هر یک از ضرایب دشواری مستخرج از نمودار نقشه سؤال، شاخص میانگین پارامتر توانایی، نمره خام معادل با درجه دشواری سؤال و درصد قبولی داوطلبین گزارش شده است. داوران با اکثریت آراء ستونی را بر روی نمودار انتخاب نمودند که شامل سؤالات ۱۰۱، ۱۰۵، ۱۴۲، ۱۹۵، ۲۰ و ۳ می شد و میانگین پارامتر دشواری سؤالات برابر ۱/۳ بود.

همان‌گونه که در جدول ۴ مشاهده می‌شود ضریب دشواری موردتوافق داوران (عدد ۱/۳)، معادل نمره خام ۱۰۳ بود که ۶۵/۸۴ از داوطلبین این نمره را کسب کردند و میانگین توانایی این سؤال برابر ۱/۳۱ است؛ بنابراین می‌توان گفت که با استفاده از روش نقشه سؤال، ضریب دشواری آزمون که به‌خوبی بین داوطلبین با حداقل صلاحیت و عدم صلاحیت تفکیک‌فائل می‌شود، ۱/۳ است و نمره ۱۰۳ حداقل نمره قبولی در آزمون است و داوطلبینی که نمراتی کمتر از ۱۰۳ کسب کنند، در آزمون جامع مردود می‌شوند.

نمودار ۲. نمودار نقشه شخص بر اساس مدل تک پارامتری راش

Person-MAP



نمودار نقشه شخص بر اساس مدل تک پارامتری راش، داوطلبینی را که از نظر پارامتر توانایی در یک سطح هستند، شناسایی می‌کند. همان‌طور که در نمودار ۲ مشاهده می‌شود، پارامتر توانایی داوطلبان از ۰/۱- تا ۳+ متغیر است و داوطلب شماره ۱۰۰ و ۸۷ با پارامتر توانایی ۳ به‌عنوان توانمندترین دانشجویان در یک محور و داوطلب شماره ۱۹۰ با پارامتر توانایی ۱- ضعیف‌ترین داوطلب شرکت‌کننده در آزمون جامع است. همچنین بیشترین تعداد داوطلبین در دامنه پارامتر توانایی ۰/۹ تا ۲/۲ قرار دارند؛ بنابراین در آزمون جامع با ضریب دشواری ۱/۳، داوطلب با پارامتر توانایی ۱/۳۱ به بالا، با حداقل نمره ۱۰۳ قبول می‌شود و داوطلبین با ضرایب دشواری کمتر، پارامتر توانایی کمتر و نمره خام پایین‌تر از ۱۰۳ مردود می‌شوند.

بحث و نتیجه‌گیری

نتایج این پژوهش نشان داد که روش نقشه سؤال توانسته است به داوران در زمینه تعیین داوطلب با حداقل صلاحیت کمک نماید. نتایج پژوهش حاضر با نتایج پژوهش‌های Wang (2003) و هم‌چنین Buckendahl و همکاران (2002) همخوانی است. با توجه به اینکه در روش نقشه سؤال از مدل راش برای تعیین پارامتر دشواری و پارامتر توانایی استفاده می‌شود، پارامتر دشواری سؤال و پارامتر توانایی بر مقیاس یکسانی قرار دارند. هم‌چنین در روش نقشه سؤال به بررسی تک‌تک سؤالات پرداخته نمی‌شود و کل سؤالات با یکدیگر موردبررسی قرار می‌گیرند. در این روش نمره قبولی با شناسایی نقطه‌ای بر روی هیستوگرام پارامتر دشواری سؤال تعیین گردید که نشان می‌داد داوطلب با حداقل صلاحیت دارای فرصت ۵۰ درصد جهت ارائه پاسخ صحیح به سؤال بود و با توجه به اینکه در تعیین استاندارد و نمره برش، تعیین سطح توانایی داوطلب با حداقل صلاحیت است معیار ۵۰٪. بهترین نقطه برای داوران جهت تعیین نمره برش است که روش نقشه سؤال با تأکید بر مدل راش این برآورد را در اختیار داوران قرار می‌دهد و به تعیین نمره برش دقیق‌تر کمک می‌کند. نتایج این پژوهش نیز با نتایج پژوهش Schnabel (2018) همخوانی دارد. مسائلی که عمدتاً داوران در جریان نمره برش با آن مواجه هستند شامل: درک شناختی صحیح داوطلب با حداقل صلاحیت و هم‌چنین شناخت دقیق روش تعیین نمره برش است که روش تعیین نمره برش نقشه سؤال، داوران به درک شناختی مفهوم داوطلب با حداقل شایستگی کمک می‌نماید.

در فرایند تعیین استاندارد برای آزمون جامع علوم پایه، در طی سه مرحله (راند)، سؤالات آزمون داورى و ارزیابی شد. در مرحله اول و دوم داورى، بین داوران توافقی در مورد نقطه برش و استاندارد (سؤالی که داوطلب با حداقل صلاحیت قادر به پاسخگوى به سؤالات بعد از آن نیست) وجود نداشت اما در مرحله (راند سوم)، با در اختیار قرار دادن نقشه سؤال، داوران بر پارامتر دشواری ۱/۳ اتفاق نظر بالایی داشتند و این نقطه را به عنوان نقطه برش و استاندارد آزمون مشخص کردند. در این نقطه برش، پارامتر توانایی ۱/۳۱، و نمره آزمودنی با حداقل صلاحیت، ۱۰۳ به دست آمد که درصد قبولی داوطلبین ۶۵/۸۴ برآورد شد. در واقع بر اساس این یافته، داوطلبین با پارامتر دشواری کمتر از ۱/۳، پارامتر توانایی کمتر از ۱/۳۱ و نمره خام پایین تر از ۱۰۳ مردود می‌شوند. مطالعه Buckendahl و همکاران (2002) و Schnabel (2018) نیز نشان داد که گرچه روش آنگوف به صورت گسترده مورد استفاده قرار گرفته است اما روش نقشه سؤال می‌تواند در نمره برش و انتظارات عملکرد داوطلبین در آزمون‌های اعطای گواهینامه به دست‌اندرکاران امر آموزش کمک نماید. از محدودیت‌های این پژوهش می‌توان به این نکته اشاره نمود که به دلیل آنکه برای اولین بار روش نقشه سؤال برای تعیین استاندارد در علوم پزشکی مورد استفاده قرار گرفت، داوران هیچ آشنایی قبلی با این روش نداشتند.

پیشنهاد می‌شود جهت تعیین نمره برش بهینه در پژوهش‌های آینده به بررسی مقایسه‌ای روش تعیین استاندارد و نمره برش نقشه سؤال با روش‌های همانند آنگوف و روش سنتی پرداخت. همچنین پژوهش‌های آینده می‌تواند به بررسی تأثیر مقادیر جایگزین احتمال ۵۰ درصد با مقادیر رایج دیگر همچون ۰/۶۷ در تعیین نمره برش بپردازد. پیشنهاد می‌گردد در پژوهش‌های آتی جلسه داورى با دو گروه از داوران متخصص یا بیشتر صورت پذیرد و نتایج دو گروه با یکدیگر مورد مقایسه قرار گیرد.

تعارض منافع

بین نویسندگان تعارض منافع وجود ندارد.

منابع

عباسی، هادی. (۱۳۹۳). *ارزیابی جامع و تعیین استاندارد علمی چیرگی در آزمون‌های تخصصی ورود به دوره‌های اترنی رشته پزشکی با استفاده از مدل‌های کلاسیک و خصیصه مکنون*. رساله دکتری رشته سنجش و اندازه‌گیری دانشگاه علامه طباطبائی.

مرتاض هجری، سارا، جلیلی، محمد، و لباف، علی. (۱۳۹۰). تعیین نمره حدنصاب قبولی آزمون عینی ساختارمند بالینی به روش انگوف و ارزیابی تأثیر بحث و بررسی نمرات واقعی. *مجله ایرانی آموزش در علوم پزشکی*، ۱۱(۸)، ۸۸۵-۸۹۴.

مینائی، اصغر (۱۳۹۳). استفاده از مدل اندازه‌گیری راش در ارزیابی ویژگی‌های اندازه‌گیری تست مهارت‌های حرکتی (TVMS-R). *فصلنامه آموزش و پرورش*، ۵ (۱۸)، ۷۷-۱۱۴.

وزارت بهداشت، درمان و آموزش پزشکی، معاونت آموزش. (۱۳۹۶). آیین‌نامه آموزشی دوره دکتری عمومی پزشکی مصوب شصت و هفتمین جلسه شورای عالی برنامه‌ریزی علوم پزشکی.

References

- Abbasi, H. (2014). *Comprehensive evaluation and determination of scientific mastery standard in specialized entrance exams for medical internship programs using classical and latent trait models* [Unpublished doctoral dissertation]. Allameh Tabataba'i University. [In Persian]
- Alshawwa, L. (2023). Standard Setting: A Review of Methods. *Asian Journal of Education and Social Studies*, 42(2), 1-7.
- Baron, P., Sireci, S. G., & Slater, S. C. (2021). Evaluating Panelists' Understanding of Standard Setting Data. *Educational Measurement: Issues and Practice*, 40(2), 16-25.
- Berk, R. A. (1996). Standard setting: The next generation (where few psychometricians have gone before!). *Applied measurement in education*, 9(3), 215-225.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*. Springer Science & Business Media.
- Buckendahl, C. W., Smith, R. W., Impara, J. C., & Plake, B. S. (2002). A comparison of Angoff and Bookmark standard setting methods. *Journal of Educational measurement*, 39(3), 253-263.
- Cetin, S., & Gelbal, S. (2013). A Comparison of Bookmark and Angoff Standard Setting Methods. *Educational Sciences: Theory and Practice*, 13(4), 2169-2175.
- Cizek, G. J. (Ed.). (2012). *Setting Performance standards: Foundations, methods, and innovations*. Routledge.
- Duncan, P. W., Bode, R. K., Lai, S. M., Perera, S., & Glycine Antagonist in Neuroprotection Americas Investigators. (2003). Rasch analysis of a new stroke-specific outcome scale: Stroke Impact Scale. *Archives of physical medicine and rehabilitation*, 84(7), 950-963.
- Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Jama*, 287(2), 226-235.
- Glaser, B. E., Baldwin, P., Margolis, M. J., Mee, J., & Winward, M. (2017). An experimental study of the internal consistency of judgments made in bookmark standard setting. *Journal of Educational Measurement*, 54(4), 481-497.

- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. *Educational measurement*, 4, 433-470.
- Hein, S. F., & Skaggs, G. E. (2009). A qualitative investigation of panelists' experiences of standard setting using two variations of the bookmark method. *Applied Measurement in Education*, 22(3), 207-228.
- Jiao, H., Lissitz, R. W., Macready, G., Wang, S., & Liang, S. (2011). Exploring levels of performance using the mixture Rasch model for standard setting. *Psychological Test and Assessment Modeling*, 53(4), 499.
- Kane, M. T. (1995). Examinee-centered vs. task-centered standard setting. In *the Proceedings of the Joint Conference on Standard Setting for Large-scale Assessment* (Vol. 2, pp. 119-141).
- Kang, Y. (2022). Evaluating the cutoff score of the advanced practice nurse certification examination in Korea. *Nurse Education in Practice*, 63, 103407.
- Linacre, J. M. (2016). Winsteps® Rasch measurement computer program. (No Title).
- MacCann, R. G., & Stanley, G. (2019). The use of Rasch modeling to improve standard setting. *Practical Assessment, Research, and Evaluation*, 11(1), 2.
- McKinley, D. W., Newman, L. S., & Wisner, R. F. (1996). Using the Rasch model in the standard setting process. In *annual meeting of the National Council of Measurement in Education*, New York, NY.
- Minaii, A. (2014). Using the Rasch measurement model in evaluating the measurement characteristics of the Test of Gross Motor Skills (TGMS-R). *Quarterly Journal of Education*, 5(18), 77-114. [In Persian]
- Ministry of Health and Medical Education, Deputy of Education. (2017). *Educational regulations for the general medicine doctoral program approved by the 67th session of the Supreme Council of Medical Education Planning*. [In Persian]
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2013). The bookmark procedure: Psychological perspectives. In *Setting Performance standards* (pp. 263-296). Routledge.
- Mortaz Hajri, S., Jalili, M., & Laffaf, A. (2011). Determining the cut-off score for the objective structured clinical examination using the Angoff method and evaluating the impact of real score discussion. *Iranian Journal of Medical Education*, 11(8), 885-894. [In Persian]
- Peterson, C. H., Schulz, E. M., & Engelhard Jr, G. (2011). Reliability and validity of bookmark-based methods for standard setting: comparisons to Angoff-based methods in the National Assessment of Educational Progress. *Educational Measurement: Issues and Practice*, 30(2), 3-14.
- Schnabel, S. D. (2018). *A comparison of the Angoff and item mapping standard setting methods for a certification examination* (Doctoral dissertation, University of Illinois at Chicago).
- Shepard, L. A. (1995). Implications for standard setting of the National Academy of Education evaluation of the National Assessment of Educational Progress achievement levels. In *Joint conference on standard setting for large-scale assessments. Vol. 2. Proceedings* (pp. 143-160).
- Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*, 57(8), 1358-1362.
- Wagner, P., Hendrich, J., Moseley, G., & Hudson, V. (2007). Defining medical professionalism: a qualitative study. *Medical education*, 41(3), 288-294.
- Wang, N. (2003). Use of the Rasch IRT model in standard setting: An item-mapping method. *Journal of Educational Measurement*, 40(3), 231-253.