



A review of the basics of determining the cut-off point in academic achievement tests and the introduction of Angoff scientific method

Hossein Kareshki ¹, Fatemeh Hajiabadi ^{2*}, Maryam Bagheri ³, Akram Ghanbari Moghaddam ⁴

1 Associate Professor, School of Education & psychology, Ferdowsi University, Mashhad, Iran.

2 Assistant Professor, Department of Medical-Surgical Nursing, School of Nursing and Midwifery, Mashhad University of Medical Sciences, Mashhad, Iran.

3 Instructor, Department of medical surgical, School of Nursing and Midwifery, Mashhad University of Medical sciences, Mashhad, Iran

4 Instructor, Department of medical surgical, School of Nursing and Midwifery, Mashhad University of Medical sciences, Mashhad, Iran

ARTICLE INFO

Article type

Review Article

Article history

Received: 26 Dec 2020

Accepted: 14 Apr 2021

Keywords

Academic achievement tests
Test standardization
Determine the cutting point



ABSTRACT

Introduction: The cut-off point in a test is the score that divides test takers into different classes (such as rejection-acceptance). The process of determining the cut-off point is frequently used in educational tests. The cut-off point should be based on a generally accepted methodology and reflect the judgment of competent and qualified individuals. The aim of this study was to determine the basics of determining the cut of point in academic achievement tests.

Materials & Methods: The present study is a review study in which all articles related to the purpose, using keywords of standardization of the test and determination of the cutting point in reputable Persian and English scientific sites were searched.

Results: In the search, it was found that there was no reliable Persian scientific source in determining the cutting point. However, English articles and books in this field were carefully studied.

Conclusion: Despite the importance of determining the correct cut point, especially in academic achievement tests, but Persian texts contain few resources in the field of principles and foundations for doing this and probably the lack of these resources makes researchers less familiar with the correct process of determining the point. Their inclination and tendency to use unscientific methods to do it in their studies. This article can be a useful tool as a guide to the basic principles of determining the cut-off point in academic achievement tests, which puts more emphasis on using the cut-off point for important decisions.



[10.22038/HMED.2021.54426.1116](https://doi.org/10.22038/HMED.2021.54426.1116)

► Cite this paper as:

Kareshki H, Hajiabadi F, Bagheri M, Ghanbari Moghaddam A, A review of the basics of determining the cut-off point in academic achievement tests and the introduction of Angoff scientific method. *Horizon of Medical Education Development*. 2021;12(4):85-96

*Corresponding author: Dr. Fatemeh Hajiabadi;

Department of Medical-Surgical Nursing, School of Nursing and Midwifery, Mashhad University of Medical Sciences, Mashhad, Iran .

Tel: +989359568726

Email: ghanbarima@mums.ac.ir

مروری بر مبانی تعیین نقطه برش در آزمون های پیشرفت تحصیلی و معرفی روش علمی انگوف

حسین کارشکی^۱ ID، فاطمه حاجی آبادی^{۲*} ID، مریم باقری^۳ ID، اکرم قنبری مقدم^۴ ID

۱ دانشیار گروه آموزشی علوم تربیتی، دانشکده علوم تربیتی و روانشناسی، دانشگاه فردوسی مشهد، مشهد، ایران.
 ۲ استادیار، گروه داخلی جراحی، دانشکده پرستاری مامایی، دانشگاه علوم پزشکی مشهد، مشهد، ایران.
 ۳ مربی، گروه فوریتهای پزشکی، دانشکده پرستاری مامایی، دانشگاه علوم پزشکی مشهد، مشهد، ایران.
 ۴ مربی پرستاری، گروه داخلی جراحی، دانشکده پرستاری مامایی، دانشگاه علوم پزشکی مشهد، مشهد، ایران.

مشخصات مقاله	چکیده
نوع مقاله مقاله مروری	مقدمه: نقطه برش در یک آزمون، نمره ای است که آزمون دهندگان را به طبقات مختلف (نظیر رد - قبول) تقسیم می کند. فرایند تعیین نقطه برش در آزمون های حوزه آموزشی به کرات مورد استفاده قرار می گیرد. نقطه برش باید بر اساس یک متدولوژی مورد قبول عموم بوده و منعکس کننده قضاوت افراد با کفایت و واجد شرایط باشد. مطالعه حاضر با هدف تعیین مبانی اولیه تعیین نقطه برش در آزمون های پیشرفت تحصیلی انجام شد.
پیشینه پژوهش تاریخ دریافت: ۹۹/۱۰/۰۶ تاریخ پذیرش: ۱۴۰۰/۱/۲۵	روش کار: مطالعه حاضر یک مطالعه مروری است که طی آن کلیه مقالات مرتبط با هدف، با استفاده از کلید واژه های فارسی استاندارد سازی آزمون و تعیین نقطه برش در سایت های معتبر علمی فارسی و انگلیسی جستجو شد.
کلمات کلیدی آزمون های پیشرفت تحصیلی استاندارد سازی آزمون تعیین نقطه برش	نتایج: در جستجوی انجام شده، هیچ منبع علمی معتبر فارسی در زمینه تعیین نقطه برش یافت نشد. با این حال مقالات و کتب انگلیسی موجود در این زمینه به دقت مورد مطالعه و بررسی قرار گرفتند. نتیجه گیری: با وجود اهمیت تعیین صحیح نقطه برش علی الخصوص در آزمونهای پیشرفت تحصیلی، و اینکه متون فارسی حاوی منابع اندکی در زمینه اصول و مبانی انجام این کار می باشد و احتمالاً اندک بودن این منابع موجب آشنایی کمتر پژوهشگران با روند صحیح تعیین نقطه برش و تمایل آنها به استفاده از روش های غیر علمی جهت انجام این مهم در مطالعات خود می باشد، این مقاله می تواند به عنوان راهنمای مبانی اولیه تعیین نقطه برش در آزمون های پیشرفت تحصیلی، کمک کننده باشد



0.22038/HMED.2021.54426.1116

نحوه ارجاع به این مقاله

Karshki H, Hajiabadi F, Bagheri M, Ghanbari Moghaddam A, A review of the basics of determining the cut-off point in academic achievement tests and the introduction of Angoff scientific method. Horizon of Medical Education Development. 2021;12(4):85-96

ایمیل: ghanbarima@mums.ac.ir

تماس: ۰۹۳۵۹۵۶۸۷۲۶

*نویسنده مسئول: دکتر فاطمه حاجی آبادی

گروه داخلی جراحی، دانشگاه علوم پزشکی مشهد، مشهد، ایران

مقدمه

نقطه برش در یک آزمون، نمره ای است که آزمون دهندگان را به طبقات مختلف (نظیر رد - قبول) تقسیم می کند. به عنوان مثال نقطه برش در اکثر آزمون های رانندگی ۷۰٪ است و این به این معنی است که هر نمره زیر این درصد به منزله مردود شدن در آزمون و هر نمره بالاتر از آن به معنی قبولی در آزمون است. زیکی بیان می کند که نقاط برش، نقاط انتخاب شده در مقیاس نمره ای یک آزمون است [۱]. این نقاط برای تعیین اینکه آیا یک نمره آزمون خاص برای بعضی اهداف کافی است یا خیر، مورد استفاده قرار می گیرد. به عنوان مثال عملکرد دانشجویان در یک آزمون ممکن است در یکی از طبقات نظیر ابتدایی، با کفایت یا پیشرفته بر اساس نقطه برش آن آزمون تقسیم بندی شود.

پیر و زیکی متذکر شده اند به این موضوع که با وجودی که راه کاملی برای تعیین نقطه برش وجود ندارد، ولیکن مراحل مشخصی وجود دارد که طی این مراحل منجر به تولید نقاط برش معقول و مفید خواهد شد. نویسنده همچنین تاکید می کند که نقطه برش باید بعد از تعیین، اعتبار یابی شود و اینکه در عمل در صورتیکه تجربه نشان دهد که نقطه برش تعیین شده اهداف مورد نظر را پوشش نمی دهد، باید تغییراتی در آن صورت گیرد. تعیین نقطه برش در آزمونهای پر استفاده در زمینه آموزشی نیاز به مشارکت سیاستگذاران، مدرسین، متخصصین اندازه گیری و دیگران در مراحل چندگانه فرایند قضاوت دارد. نقاط برش باید بر پایه متدولوژی ای باشد که به طور کلی مورد قبول باشد و منعکس کننده قضاوت افراد شایسته و با صلاحیت باشد. برای تعیین منطقی و معقول نقطه برش باید چندین مرحله طی شود که عبارتند از: تعیین اینکه آیا نقطه برش مفید خواهد بود، انتصاب پرسنل برای وظایفی که باید

انجام شود، انتخاب سطوح عملکردی که باید گزارش شوند (به عنوان مثال: ابتدایی، با کفایت و پیشرفته)، توضیح در مورد اینکه برای رسیدن به هر کدام از سطوح عملکردی دانشجویان به چه چیزی نیاز دارند، تعیین نقاط برش موقت، تعیین نقاط برش عملکردی، ثبت مراحل فرایند و ارزیابی نتایج استفاده از نقطه برش تعیین شده است.

موضوعات مهمی که در تعیین و استفاده از نقطه برش باید به آن توجه داشت و مد نظر قرار داد عبارتند از: ضرورت نیاز به قضاوت در هنگام تعیین نقطه برش، شایستگی داورانی که در مراحل مختلف فرایند تعیین نقطه برش درگیرند، مفهوم عملکرد مرزی، احتمال خطای طبقه بندی در هنگام استفاده از نقطه برش، نیاز به هماهنگی با نقطه برش در سراسر نمره دهی، انتخاب بین نمره دهی جبرانی و عطفی و اهمیت اطلاعات هنجاری.

یکی از مهمترین اصول در زمینه تعیین نقطه برش، مشخص کردن این موضوع است که آیا تعیین نقطه برش مفید است یا خیر. سیاستگذاران باید فواید آنچه از تعیین نقطه برش مورد انتظار است را شرح دهند. چه تصمیماتی بر پایه نقاط برش گرفته می شود؟ چگونه این تصمیمات هم اکنون در غیاب نقطه برش گرفته می شوند؟ چه دلایلی وجود دارد که نقاط برش منجر به تصمیم گیری بهتر خواهد شد. فواید مورد انتظار از این تصمیمات چه هستند؟ [۱]

این مهم است که لیست فواید ناشی از تعیین نقطه برش برای کسب اجماع در میان بهره گیران از برنامه مبنی بر مناسب بودن دلایل تهیه شود. واضح است که نقطه برش به خودی خود قدرت بسیار کمی برای ارتقاء آموزش دارد. اندازه گیری قد و وزن یک کودک و طبقه بندی رشد وی، خواه متناسب یا غیر متناسب، کمکی به رشد کودک نخواهد کرد [۲].

طرح پیشنهادی خود را ارائه دهند. مدیران باید افراد متخصص و خبره را به این کار بگمارند. این افراد ممکن است از بین مدرسین منطقه ای باشند اما اغلب مشاورین خارجی می باشند یا فردی از پرسنل آژانس های خصوصی آزمون سازی که برای کمک به تعیین نقطه برش، قرار داد بسته است. مدیر با وجود محدودیت های تعیین شده توسط سیاستگذاران، وظایفی که باید انجام شوند را تعیین میکند و برنامه زمانبندی و بودجه بندی کار را طراحی کرده، مشارکت کنندگان مراحل مختلف فرایند را تعیین می کند، روش مورد استفاده برای تعیین نقطه برش را مشخص می کند، داوران را آموزش می دهد، فرایند کار را پایش می کند و از پیشرفت کار اطمینان حاصل می کند.

سیاستگذاران با کمک مدرسین، باید در مورد سطوح عملکردی مورد استفاده برای گزارش نتایج و نیز تعریف کلی هر سطح تصمیم بگیرند. سطوح عملکردی نشاندهنده طبقاتی است که عملکرد دانشجویان بر اساس آنها تقسیم بندی می شود. سطوح عملکردی در اصطلاحات کلی بیان می شوند. به عنوان مثال ممکن است سیاستگذاران تصمیم بگیرند که سطح عملکردی نظیر قبول و مردود، با کفایت، و پیشرفته را مورد استفاده قرار دهند [۱].

انتخاب این نام ها به روش خاص مورد استفاده برای تعیین نقطه برش بستگی ندارد و ممکن است قبل از تعیین نوع روش مورد استفاده، مشخص شوند. این فرایند برای رسیدن به اجماع برای شمار طبقاتی است که قرار است دانش آموزان به آنها تقسیم بندی شوند، ضرورت دارد و نیز توافقی در زمینه عنوانی است که برای هر طبقه استفاده می شود و یک تعریف کلی در مورد این عنوان است. داشتن کمتر از سه تا چهار طبقه از همه بهتر است چرا که در صورت تعداد بیشتر از این، ممکن است که افتراق بین آنها مشکل باشد.

به عنوان مثال، اگر یکی از دلایل تعیین نقطه برش ارتقاء کیفیت تدریس در مدارس است، بلید بدانیم که نقطه برش به خودی خود اثرات مطلوب را در بر نخواهد داشت. استفاده از نقطه برش ممکن است نشان دهد که مدارس خاص، گروههای خاص و کوریکولوم های خاص بیشتر از دیگران نیازمند ارتقاء می باشند. ولی نقطه برش به تنهایی آموزش را بهبود نمی بخشد. تا زمانیکه نیازهای زیربنایی برای ارتقاء آموزش سر جای خود قرار دارد، تعیین نقطه برش برای این اهداف، کاری بیهوده است [۱].

در نظر داشتن اثرات منفی بالقوه نقطه برش نیز ضروری است. چه اتفاقی برای دانش آموزان مردود شده می افتد؟ آیا به آنها انگ زده شده و فراموش می شوند و یا اینکه به آنها کمک می شود؟ چه اتفاقی برای مدارس که نسبت دانشجویان مردود شده آنها زیاد است می افتد؟ آیا این موسسات تنبیه می شوند یا اینکه مورد کمک واقع می شوند؟ چه اتفاقی برای معلمینی می افتد که تعداد زیادی از دانشجویان آنها مردود شده اند؟ آیا این معلمین تنبیه می شوند یا کمک های اضافی و بیشتری را دریافت می دارند؟ با اینکه استفاده از نقطه برش منجر به نتایج مثبت می شود با این حال بعضی افراد ممکن است استفاده از نقطه برش را ناعادلانه تصور کنند. مهم است که دلایل استفاده از نقطه برش به مدرسین و جامعه گفته شود. مردم باید بدانند که چرا آزمون گرفته می شود و چرا دانش آموزان به سطوح مختلف شایستگی تقسیم بندی می شوند.

اگر توافقی در زمینه دلایل تعیین نقطه برش، حاصل شده و فواید مورد انتظار از پیامدهای منفی آن بیشتر باشد، در این صورت عاقلانه است که به مرحله بعد برویم. مرحله بعد اختصاص به سیاستگذاران دارد که افرادی را برای وظیفه تعیین نقطه برش استخدام کرده یا بگمارند. یک کار معمول این سیاستگذاران این است که از این افراد می خواهند تا

educational achievement و cut score در منابع لاتین بین سال های ۲۰۰۰ تا ۲۰۲۰ مورد بررسی قرار گرفتند .

اولین روش برای تعیین نقطه برش در آزمون های تحصیلی معیار- مرجع بوسیله ندل اسکي در سال ۱۹۵۴ ایجاد شد. سال بعد شمار تعداد این روش ها به ۳۸ عدد رسید(berk,1986) و هنوز هم رو به افزایش است. به عنوان مثال در فرایند انتخاب روش مناسب برای تعیین نقطه برش برای بررسی ملی پیشرفت آموزشی در امریکا، ریکاس ، ۱۴ مدلی که به تازگی تدوین شده بودند را تشریح کرد. در طی همین سالها، هامبلتون و همکارانش ، ده روش دیگر را برای تعیین نقطه برش که خصوصا برای مهارت های تولیدی (productive) مناسب بودند را ارائه دادند. امروزه بیش از ۶۰ روش ، ثبت شده است[۲,۳].

طرح های طبقه بندی روش های تعیین نقطه برش (classification schemes):

یکی از ابتدایی ترین طبقه بندی های روش های تعیین استاندارد (standard setting) به مسکاساس در سال ۱۹۷۶ نسبت داده شده است که روش ها را به ۲ گروه اصلی تقسیم کرده است که عبارتند از: state models (یک متغیر مجرد و جدا و مجزا با سطوح کاملا تعریف شده و روشن (clearly defined levels) و continuum model (یک کمیت مستمر و پیوسته). تفاوت بین این دو مدل در نوع متغیرهایی (مجرد و مجزا یا پیوسته) است که شایستگی و صلاحیت در آن بیان شده است. در مفاهیم جدید توسعه صلاحیت های متفاوت ، آنها به صورت متغیر های پیوسته تصور می شوند. به همین دلیل است که جاگر (۱۹۸۹) توصیه کرده است که انتخاب روش تعیین نقطه برش در موقعیت واقعی باید محدود به روش های متعلق به گروه مدل های پیوسته باشد. در عمل، همه ۶۲ مدل لیست شده متعلق به طبقه مدل های پیوسته است. که در عوض نشان می

باید قبل از شروع کار، در مورد اینکه دانش آموزان برای رسیدن به هر سطح عملکردی باید چه چیزی بدانند و یا قادر به انجام چه کاری باشند توضیح داده شود. و باید قبل از اجراء، در مورد این موضوع، بین مدیران و سیاستگذاران توافق به عمل آید. نوشتن توضیحات هر سطح عملکردی کار بزرگ و وقتگیری است. و برای این کار ابتدا لازم است که به طور مقدماتی یک نقطه برش تعیین شود. نقطه برش مقدماتی، انتخاب اولیه نمرات مورد نیاز برای یک سطح عملکرد در یک آزمون خاص است. نقطه برش در آزمونهای پیشرفت تحصیلی توسط افرادی که از این موضوع آگاهی دارند که چه چیزی دانشجویان باید بر پایه آموزشهای دریافت شده خود در مورد موضوع، بدانند و قادر به انجام باشند تعیین می شود. افرادی که نقطه برش اولیه را تعیین می کنند ممکن است و نه ضرورتا مشابه افرادی باشند که توضیحات سطوح عملکردی را ارائه می دهند[۱]. با توجه به اهمیت تعیین درست نقطه ی برش در تصمیم گیری های مرتبط با پیشرفت تحصیلی و از طرفی عدم وجود منابع معتبر و تخصصی در متون فارسی جهت استفاده ی مدیران و سیاستگذاران و مدرسان پژوهش حاضر با هدف تعیین هدف تعیین مبانی اولیه تعیین نقطه برش در آزمون های پیشرفت تحصیلی و معرفی یک روش کاربردی در این رابطه انجام شد.

روش کار

برای یافتن مطالعات مرتبط با تعیین نقطه برش به پایگاه های اطلاعاتی و مجلات علمی مربوطه (, pubmed , Erice و CrossRef, Science direct Magiran, Google scholar, SID) مراجعه شد و کلیه متون مرتبط با موضوع پژوهش با کلید واژه های فارسی : پیشرفت تحصیلی و متد های تعیین نقطه برش و نیز کلید واژه های انگلیسی :

مرزی و متضاد.

روش طبقه بندی غالبی که برای تعیین نقطه برش مورد استفاده قرار می گیرد، توسط جاگر ارائه شده است. بر طبق این نوع روش طبقه بندی، روش ها به ۲ گروه اصلی بسته به اینکه قضاوت متمرکز بر چه چیزی است، تقسیم می شوند: Test-centered: به طوریکه موضوع مورد قضاوت، سوال های آزمون می باشند و Examinee-centered که در آن، موضوع مورد قضاوت، فرد آزمون دهنده است. تنها ۵٪ کل روش ها از این معیار استفاده کرده اند و مابقی همه معیار اول را مد نظر قرار داده اند. چندین دلیل برای این عدم تعادل وجود دارد. از جمله اینکه روش های متمرکز بر آزمون از یک نقطه نظر عملی استفاده می کنند که در نتیجه نیاز به منابع کمتر داشته و سازماندهی آنها آسانتر خواهد بود، همچنین علی رغم فقدان مدارک و مستندات، به طور سنتی، روش های متمرکز بر آزمون بیشتر معتبر و قابل اعتماد در نظر گرفته می شوند.

استفاده رو به افزایش از فرمتی که برای اندازه گیری عملکرد (performance assessment) مناسب است در ۱۵ الی ۲۰ سال گذشته، منجر به خلق روش های جدید برای تعیین استاندارد های عملکردی شده است. در این روش ها، تمرکز قضاوت خود آزمون نیست بلکه پاسخ های آزمون گیرنده به ایتام های آزمون مورد قضاوت قرار می گیرد. لذا این روش ها به طور جداگانه به عنوان performance-centered گروه بندی شده اند.

علی رغم استفاده زیاد از روش جاگر، ولی این روش قادر به در بر گرفتن همه روش های موجود برای تعیین نقطه برش نیست برای همین، طرح های طبقه بندی بعدی ارائه شده است. به عنوان مثال ریکاس (۲۰۰۰) یک طرح طبقه بندی شامل سه سطح را مورد استفاده قرار داد که بر پایه: سطح پیچیدگی وظیفه ای که برای قضاوت تعیین

دهد که این طرح طبقه بندی برای موقعیت حاضر نامناسب بوده و صرفا دارای ارزش تاریخی است.

چند سال بعد هامبلتون (۱۹۸۰) یک طرح طبقه بندی متفاوت را پیشنهاد کرد که روش ها را به ۳ گروه مهم تقسیم می کرد که عبارت بود از: قضاوتی (judgmental)، تجربی (empirical) و ترکیبی (combination). مهمترین نقد این طرح دسته بندی این است که این اسامی و نامها این حقیقت را که همه روش ها شامل یک جزء ذهنی می باشند را کتمان می کنند و این بر طبق نظر جاگر (۱۹۸۹) می تواند منجر به سردرگم شدن و تفسیر اشتباه گردد.

برگ (۱۹۸۶) یک طرح طبقه بندی پیچیده تر را پیشنهاد داد که شامل ۳ سطح می باشد. سطح اول مربوط به طبیعت و ماهیت شایستگی مورد اندازه گیری است و در عمل همزمان با طبقه بندی مسکاساس است که روش ها را به مدل های مجزا/منفرد و پیوسته تقسیم بندی کرده است. سطح دوم طرح طبقه بندی، روش ها را بر اساس میزان نسبی ذهنیت در تعیین نقطه برش گروه بندی میکند. برک این روش ها را به ۳ گروه اصلی تقسیم می کند: judgmental که نقطه برش فقط بر پایه قضاوت تعیین می شود. Judgmental-empirical که در آن تعیین نقطه برش اساسا بر پایه قضاوت است ولی با در نظر گرفتن داده های تجربی و Empirical-judgmental که داده های تجربی عامل تصمیم گیری هستند اما قضاوت نیز به حساب آورده می شود. همانطور که دیده شد، همیشه در این نوع طبقه بندی، یک جزء ذهنی وجود دارد ولی میزان مشارکت متغیر است. مثال های تیبیک از روش های قضاوتی، روش کلاسیک ندل اسکی، انگوف و ایبل است. مثال روش Judgmental-empirical روشی است که توسط جاگر تدوین شده است. و دو روش مشهور از گروه Empirical-judgmental عبارتند از روش گروه های

شده است، کمیت و نوع اطلاعات و یا داده های تجربی فراهم شده برای قضاوت و سطح پیچیدگی روش های مورد استفاده برای خلاصه کردن نتایج قضاوت و تعیین نهایی نقطه برش می باشد.

مقبول ترین و مدرن ترین طرح طبقه بندی، طرحی است که توسط هامبلتون، جاگر و پلیک ارائه شده است. این طرح شامل ۶ سطح مربوط به تمرکز، فرایند قضاوت و تخصیص، انتخاب قضات، روایی نقطه برش و ویژگی های ابزار است. از همه ۶ سطح، تنها دو سطح اول عملیاتی شده اند که عبارتند از موردی که مربوط به تمرکز قضاوت و نیز مورد مربوط به تخصیصی که به قضات داده شده است می باشد. سطح اول به طور واقعی، بسط یافته مدل طبقه بندی جاگر است و شامل ۴ زیر طبقه است که عبارتند از:

۱- Test-centered: موضوع مورد قضاوت سوالهای آزمون می باشد (task-material): دو سوم روش های تعیین نقطه برش از همین نوع می باشند. این روش نیاز به آمادگی در سطح بالا و نیز زمان و منابع زیاد ندارد. قسمت زیادی از این روش ها (۶۸٪) می تواند قبل از اینکه تست اجرا شود، انجام شود که این یکی از مهمترین فوائد آن است. دیگر مزیت این روش این است که در کاربرد آن، در مقایسه با آزمون های آزمون شونده - محور، داوران بیشتری می توانند مورد استفاده قرار گیرند که این، باعث دستیابی به سطوح بالاتر پایایی در وظیفه قضاوت می شود. به عبارت دیگر همه روش های آزمون محور نیاز به داور جهت داوری سختی آیتم های آزمون دارد و این با قضاوت در مورد احتمال پاسخ صحیح در گروه هدف آزمون دهنده تعریف شده یا بوسیله طبقه بندی سوالهای آزمون به چندین گروه از پیش تعیین شده بر حسب سطح شایستگی مورد نیاز جهت دادن پاسخ صحیح

به آنها انجام می شود. توانایی داوران برای تقسیم بندی صحیح آیتم ها بر حسب دشواریشان موضوع بسیاری از مطالعات است. در کل مهمترین نتیجه ای که از این مطالعات استخراج شده است این است که داوران نمی توانند به طور کافی در مورد سختی آیتم ها قضاوت کنند. خصوصا اگر از ابتدا جهت انجام این کار تربیت نشده باشند. این مشکل مهمترین عیب روش های آزمون محور در نظر گرفته می شود.

۲- examinee-centered: موضوع مورد قضاوت آزمون شونده است (کسی که به آیتم های آزمون پاسخ می دهد): روشهای آزمون شونده محور در مراحل اولیه توسعه روشهای تعیین استاندارد ایجاد شدند. مهمترین مزیت این روش ها این است که نقطه برشی که از این روش ها به دست می آید معمولا دارای درجه پایداری و کفایت بالایی است. ولی از نقطه نظر عملی کاربرد این روش ها با مشکلات متعددی روبرو است. دیگر عیب این روش ها این است که تعداد داوران به یک نفر برای هر آزمون شونده محدود می شود. و این با توجه به سطح شایستگی هر آزمون شونده اجتناب ناپذیر است. و این دلیل آن است که این روش به طور روز افزون با روشهای عملکرد محور جایگزین می شود گرچه هنوز به طور خاص روش گروههای متضاد به کرات مورد استفاده قرار می گیرد.

۳- Performance-centered: موضوع مورد قضاوت پاسخ های آزمون شوندگان به سوالهای آزمون به طور مجزا است. روش های کارایی محور نیز در اساس آزمون شونده محور هستند اما با یک تمرکز محدودتر و باریک تر. در حالیکه در روش های آزمون شونده محور، موضوع مورد قضاوت کل رفتار های آزمون شونده قبل از آزمون است شاخص مورد ارزیابی در روش های کارایی محور فقط محصول فعالیت آزمون شونده در طی آزمون

همانطور که در ابتدا ذکر شد تعداد روش های تعیین نقطه برش به بالغ بر ۶۰ روش می رسد که مشهورترین آنها عبارتند از روش ندل اسکی که اولین روش ارائه شده برای تعیین نقطه برش می باشد. روش ایبل، روش گروه های مرزی، روش گروه های متضاد، بسط یافته روش انگوف، روش بوک مارک و روش بدنه کار. در اینجا به شرح یکی از این روش ها به نام روش انگوف می پردازیم [۱، ۴]. روش انگوف برای تعیین نقطه برش (Angoff Method):

روش انگوف فرایندی است که تعیین می کند که چقدر یک فرد واجد حداقل شرایط عملکرد (minimally qualified performer) یک سوال آزمون را به درستی پاسخ می دهد. در روند تعیین این میزان، گروهی از متخصصان انتخاب می شوند تا سوالات آزمون را مرور کنند و احتمال اینکه فرد واجد حداقل شرایط عملکرد به سوال مورد نظر پاسخ صحیح بدهد را برآورد کرده و تخمین بزنند. میانگین این تخمین ها به دست خواهد آمد و از این میانگین به دست آمده برای تعیین نقطه برش استفاده می شود. برای فهم بهتر مطلب یک مثال در این جا آورده می شود و مراحل کار به ترتیب توضیح داده می شود. این فرایند شامل ۵ مرحله است. که عبارتند از: ۱- انتخاب داوران Select the raters، ۲- برگزاری آزمون، ۳- نرخ دادن به هر سوال Rate the items، ۴- مرور نرخ های تخمین زده شده برای هر سوال توسط داوران Review the ratings، ۵- تعیین نقطه برش [۷].

مرحله اول: انتخاب داوران (Select the raters)
حداقل ۵ داور متخصص کارشناس موضوع ماده Subject Matter Expert (SME) انتخاب کنید و آنها را در یک محل جمع کنید. در این محل باید امکان کار گروهی

است. این نوع تمرکز خود باعث رفع عیب اصلی روش های آزمون شونده محور می شوند که همان محدودیت تعداد داوران است.

۴- Result-centered: موضوع مورد قضاوت نمرات آزمون است (و یا توزیع فراوانی آنها) یا پروفایل آزمون (در موردی که آزمون از چندین زیرآزمون تشکیل شده است). روش های نتیجه محور در اندازه گیری مهارت های پیچیده مورد استفاده قرار می گیرند جایی که آزمون از چندین قسمت مجزا تشکیل شده است یا به عنوان یک روش همراه و تابع برای اهداف مقایسه ای و برای روایی نقطه برش تعیین شده با کمک دیگر روش های مقدماتی مورد استفاده قرار می گیرد.

نکاتی در مورد وظیفه قضاوت در تعیین نقطه برش:

تقریباً در همه روش های تعیین استاندارد از قضاوت متخصصین برای تعیین نقطه برش استفاده می شود. و بعد از ارائه قضاوت جداگانه توسط هر داور، نتایج داوری ها خلاصه شده تا نمره قبولی اولیه به دست آید. به طور معمول، اجراهای متعدد قضاوت بوسیله متخصصان متعدد، اطلاعات کمی را ارائه می دهد که با استفاده از آنها می توان نقطه برش اولیه را تعیین کرد. در مورد تعداد داوران و حداقل تعداد آنها برای انجام قضاوت باید گفت که هر چه تعداد بیشتر باشد بهتر است چرا که تعداد بیشتر داوران منجر به کاستن از خطای استاندارد می گردد. لیوینگستون و زیکی توصیه می کنند که تعداد داوران نباید کمتر از ۵ تا باشد. بیدل (۱۹۹۳) پیشنهاد می کند که ۷ الی ۱۰ داور کافی خواهد بود و بر پایه تجزیه و تحلیل انجام شده با استفاده از تئوری تعمیم پذیری، هاتز و هرتر (۱۹۹۹) ۱۰ الی ۱۵ داور را توصیه کرده اند [۲].

روش های تعیین نقطه برش (Method for setting cut score):

آزمون نظر خواهی کنید. اگر سوال نیاز به تجدید نظر داشت، قبل از اینکه فرایند تخمین زدن شروع شود، این تجدید نظر را اعمال کنید.

مرحله سوم: تخمین سوالات یا نرخ دادن به انها (Rate the Items)

قبل از اینکه فرایند واقعی تخمین شروع شود ابتدا داوران را با بعضی مفاهیم آشنا کنید به عنوان مثال تعریفی از فرد با حداقل شرایط عملکرد ارائه دهید. راهنمایی های را در مورد چگونگی تخمین زدن سوالات آزمون به داوران بدهید و فرایند تخمین را شرح دهید. فرد واجد حداقل شرایط عملکرد کسی است که اولاً وظیفه را در محیط انجام می دهد نه اینکه دانش آموز است (مثلاً یک پرستار و نه یک دانشجوی پرستاری)، کسی که دارای حداقل دانش و تجربه مورد نیاز برای انجام آن وظیفه هست، کسی که در انجام وظیفه به سطح استاندارد می رسد البته به سختی و بالاخره کسی که وظیفه را در حد لب مرز ولی قابل قبول انجام می دهد. سپس از داوران بخواهید تا تخمین خود را به صورت درصد از ۱۰۰ در مورد هر سوال ارائه دهند و در فرم مربوطه ثبت کنند (مطابق جدول ۱)

و انفرادی برای داوران وجود داشته باشد. به طور ایده ال ده نفر یا بیشتر داور مورد نیاز است ولی ممکن است از نظر در دسترس بودن داوران و از نظر تدارکات امکان انتخاب این تعداد داور ممکن نباشد. برای انتخاب داوران شایستگی های زیر را در نظر بگیرید:

-آشنایی با وظیفه و عملکردی که آزمون آن را اندازه گیری می کند.

-توانایی قبولی در آزمون مورد بررسی در نقطه برش فعلی
-توانایی ویرایش سوالات آزمون از نظر شفافیت، صحت، جمله بندی و گرامر

-دانش مهارت مجموعه افرادی که آن وظایف را انجام خواهند داد. و علاوه براین برای تخمین زدن آزمون های پیشرفت (RAT)، از داورانی استفاده کنید که یک درجه بالاتر از آزمون شونده ها باشند. و سعی شود تا از داوران با نژاد، جنس، سن و سطح تحصیلی و زمینه آموزشی متفاوت و متنوع استفاده شود.

مرحله دوم: اجرای آزمون (Take the test)

از داوران بخواهید تا آزمون را با همین نقطه برش اجرا کنند و از داوران در مورد جمله بندی، طرح آزمون و سوالات

جدول ۱: نمونه فرم تخمین سوالات آزمون

تخمین سوالات آزمون	
نام آزمون:	نام دوره:
تاریخ:	نام ارزیاب:
تذکر: هر سوال را بخوانید و سپس احتمال اینکه فرد با حداقل شایستگی عملکردی بتواند به سوال پاسخ صحیح بدهد را تعیین کنید. میزان احتمال از ۹۵ درصد بیشتر و از ۲۵ درصد کمتر نباشد.	
فرد با حداقل شایستگی عملکردی: کسی است که دارای حداقل آموزش و تجربه لازم برای اجرای وظیفه مورد نظر است.	
سوالات آزمون	درصد

سوالی، حداقل ۲ ساعت است لذا باید زمان کافی به داوران داده شود. بعد از دادن تخمین از داوران بخواهید برای طی مراحل بعدی دور هم بنشینند.
مرحله چهارم: review the rating
فرم های مربوط به تخمین را از داوران جمع آوری کنید و نتایج را در برنامه excel مطابق جدول ۲ وارد کنید.

در نظر داشته باشید که داوران نباید کلید سؤالات را در اختیار داشته باشند چراکه در اختیار داشتن کلید سؤالات باعث می شود که میزان برداشت آنها از درجه دشواری سوال تحت تاثیر قرار بگیرد. داوران در هنگام دادن تخمین با یکدیگر مشورت نکنند و تخمین به صورت انفرادی داده شود. مدت زمان لازم برای یک آزمون ۱۰۰

جدول ۲: نمونه نتایج تخمین سؤالات آزمون

سوالات آزمون	میانگین در صد های تخمین زده شده	متخصص ۱	متخصص ۲	متخصص ۳	متخصص ۴	متخصص ۵	انحراف معیار
۱	۷۶	۹۰	۸۰	۵۰	۷۰	۹۰	۱۶,۷۳۳
۲	۶۸	۶۵	۷۰	۶۰	۷۵	۷۰	۵,۷۰۰۸
۳	۷۴	۸۵	۸۵	۶۰	۶۰	۸۰	۱۲,۹۴۲۱
۴	۸۵	۸۵	۹۰	۸۰	۸۵	۸۵	۳,۵۳۵۵
۵	۷۹	۹۵	۸۵	۶۰	۸۰	۷۵	۱۲,۹۴۲۱
میانگین	۷۶,۴	۸۴	۸۲	۶۲	۷۴	۸۰	

را مجددا تخمین دهی کنند و مجددا فرم ها را جمع آوری کرده و در برنامه excel وارد کنید.
مرحله ۵: تعیین نقطه برش: determining the cut score
بعد از اینکه سوالات آزمون مورد تجدید نظر مجدد قرار گرفت و تخمین ها در برنامه اکسل وارد شدند، داده های اکسل را از نظر موارد زیر مورد بررسی قرار دهید:
- اگر یک داور تخمین های مشابهی را برای همه سوالات در نظر گرفته بود، آن داوری را حذف کنید.
- اگر نتایج داوری یک داور بسیار با داوری دیگران تفاوت داشت، حذف داوری او را مد نظر قرار دهید.
- اگر باز هم در مورد یک سوال انحراف استاندارد بالای ۱۰ وجود داشت، جلسه بحث و بررسی دیگری برای

تخمین های متفاوت توسط داوران برای یک سوال پدیده ای مورد انتظار است. این تفاوت ها می تواند ناشی از درک متفاوت نسبت به سطح شایستگی وظیفه مورد نظر و یا دیگر عوامل باشد. انحراف استاندارد منعکس کننده میزان توافق یا عدم توافق بین داوران در مورد هر سوال است. انحراف استاندارد پایین تر نشان دهنده توافق بیشتر بین داوران است و در صورتی که انحراف استاندارد خیلی بالا باشد، نیاز است که آن سوال مورد تجدید نظر قرار گیرد. در مورد هر سوالی که انحراف معیار آن بالاتر از ۱۰ است، در مورد علت این انحراف استاندارد بالا با داوران بحث کنید تا در نتیجه این مباحثه احتمالا داوران در مورد میزان تخمین های ارائه شده خود تجدید نظر کنند. بعد از انجام مباحثه، داوران را از یکدیگر جدا کنید و از آنها بخواهید تا هر کدام از سوالات آزمون که انحراف استاندارد بالای ۱۰ داشته است

آن برگزار کنید.

- در هر حال اگر باز هم این عدم توافق های زیاد وجود داشت، نباید نگران آن بود. چرا که میانگین تخمین ها (عدد ۷۶,۴ در جدول شماره ۲) در تعیین نقطه برش دخالت خواهد داشت.

تقدیر و تشکر:

بدینوسیله از مسئولان دانشگاه علوم پزشکی مشهد تشکر و قدردانی می شود

تضاد منافع:

نتایج این مطالعه هیچ گونه تضادی با منافع ارگان یا فرد خاصی ندارد.

بحث و نتیجه گیری

باتوجه به وجود اهمیت تعیین صحیح نقطه برش علی الخصوص در آزمونهای پیشرفت تحصیلی، و اینکه متون فارسی حاوی منابع اندکی در زمینه اصول و مبانی انجام این کار می باشد و احتمالاً اندک بودن این منابع موجب آشنایی کمتر پژوهشگران با روند صحیح تعیین نقطه برش و تمایل آنها به استفاده از روش های غیر علمی جهت انجام این مهم در مطالعات خود می باشد، این مقاله می تواند به عنوان راهنمای مبانی اولیه تعیین نقطه برش در آزمون های پیشرفت تحصیلی، کمک کننده باشد.

References

1. Zieky, M., M. Perie, and S. Livingston, A primer on setting cut scores on tests of educational achievement. Princeton, NJ: Educational Testing Service, 2006. 320.
2. Kaftandjieva, F., Methods for setting cut scores in criterion-referenced achievement tests. Cito, Arnhem: EALTA, 2010. 170.
3. <http://www.nocti.org/pdf/CRCutScores.pdf>.
4. Tavakol, M. and R. Dennick, Standard setting: the application of the receiver operating characteristic method. 2012, IJME.
5. Wheaton, A. and J. Parry. Using the Angoff method to set cut scores. in Users Conference. 2012.
6. https://www.questionmark.com/us/seminars/Documents/webinar_angoff_handout_may_2012.pdf.
7. Haertel, E.H. and W.A. Lorie, Validating standards-based test score interpretations. Measurement, 2004. 2(2): p. 61-103.