نامه به سردبیر

آزمون‌های پیشرفت تحصیلی را چگونه تحلیل کنیم؟

هادی عباسپور*، زهره گریوانی، حسین کریمی مونقی

مجله ایرانی آموزش در علوم پزشکی/ شهریور ۱۴۰۲ ۲۳ (۲۶): ۱۶۳ تا ۱۶۵

DOI: 10.48305/23.163

با مروری بر متون مشخص شد که اغلب آنها در تحلیل سوالات آزمون‌های پیشرفت تحصیلی (پایان‌ترم، پایان‌دوره) تفسیر نادرستی از موضوع دارند(۱تا۵). با تجارب موجود نیز در تحلیل آزمون‌های پیشرفت تحصیلی در مراکز آموزش عالی کاستی‌هایی وجود دارد. لذا ذکر چند نکته اصلاحی در این مورد قابل تامل است. در ابتدا به تفاوت در تحلیل آزمون‌های هنجاری با ملاکی می‌پردازیم. در ادامه به چگونگی تحلیل سوال‌های آزمون‌های وابسته به ملاک و در پایان به نتیجه‌گیری نهایی خواهیم پرداخت.

ارزشیابی آموزشی بر اساس دو رویکرد ملاکی و هنجاری انجام می‌پذیرد. در ارزشیابی ملاکی، معیار ارزشیابی از پیش تعیین می‌شود و عملکرد فراگیر در آزمون با توجه به آن ملاک مورد قضاوت قرار می‌گیرد. در این نوع ارزشیابی که بیشتر ویژه پیشرفت تحصیلی است، نوعی ملاک مطلق که همان هدف‌های آموزشی از پیش تعیین‌شده هستند مورد استفاده قرار می‌گیرند. در این شرایط سنجش برای بررسی دستیابی یا عدم دستیابی به اهداف آموزشی با ابزاری مانند آزمون صورت می‌گیرد. یعنی کسب موفقیت در آزمون‌هایی که مبتنی بر ملاک است مستلزم یادگیری هدف‌های آموزشی در سطحی است که برنامه آموزشی بر آن اساس طراحی شده است(۶). مانند آزمون‌های پایان ترم تحصیلی و آزمون‌های جامع علوم پایه و پیش‌کارورزی.

تفسیر نمره ملاک–محور به‌ویژه در آزمون شایستگی مفید است. به‌عنوان مثال، ارائه بازخورد در آزمون طراحی شده با هدف بهبود عملکرد باید نمرات از منظر میزان تحقق ملاک مورد نظر تفسیر شوند. به‌همین ترتیب، ارزیابی پایان دوره باید نمراتی را ایجاد کند که نشان دهد دانشجویان چقدر مطالب را یاد گرفته‌اند. با این حال، تفسیر نمره آزمون ملاکی زمانی که هدف مقایسه آنها و رتبه‌بندی دانشجویان باشد مفید نیست. نمرات ارزیابی‌های مبتنی بر ملاک می‌تواند به‌ویژه در برنامه‌های آموزشی مبتنی بر نتیجه یا شایستگی مفید باشند(۷). اما در ارزشیابی وابسته به هنجار به جای یک ملاک از پیش تعیین‌شده و مطلق، بر اساس نوعی ملاک نسبی طراحی می‌شود. به این صورت که عملکرد فراگیران را با یکدیگر مقایسه می‌کنیم. در این روش می‌توان تعیین کرد که پیشرفت فراگیر نسبت به سایر فراگیران چه وضعی دارد اما نمی توان تعیین کرد که پیشرفت یک دانشجو نسبت به هدف‌های آموزشی چگونه است. در پایان این نوع آزمون‌ها جایگاه فراگیر را برحسب رتبه‌ی درصدی یا نوعی هنجار سنی یا کلاسی و در مقایسه با گروه هنجار تعیین می‌شود. از این نوع می‌توان به آزمون‌های دستیاری اشاره کرد(۸و۹).

* **نویسنده مسؤول:** دکتر هادی عباسپور (استادیار)، گروه پرستاری، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی خراسان شمالی، بجنورد، خراسان شمالی، ایران.
abbaspourh53@gmail.com

زهره گریوانی (مربی)، گروه هوشبری، دانشکده پیراپزشکی، دانشگاه علوم پزشکی خراسان شمالی، بجنورد، خراسان شمالی، ایران. (gerivaniz70@gmail.com)؛

دکتر حسین کریمی مونقی)استاد(، مرکز تحقیقات مراقبتهای پرستاری و مامایی، گروه آموزشی، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی مشهد، مشهد، خراسان ، خراسان رضوی، ایران. (karimih@mumus.ac.ir)

تاریخ دریافت مقاله: ۲۶/۵/۱۴۰۲ تاریخ پذیرش: ۱۲/۶/۱۴۰۲

[Downloaded from ijme.mui.ac.ir on 2026-11-02]

[DOI: 10.48305/23.0.163]

یکی از فعالیت‌هایی که پس از اجرای آزمون برای تحلیل سؤالات آن می‌توان انجام داد محاسبه ضریب دشواری و ضریب تمیز هر سؤال می‌باشد. در تعریف ضریب دشواری باید گفت: تعیین درصد کل آزمون‌شوندگانی که به یک سؤال جواب درست می‌دهند. به‌عبارتی از محاسبه ضریب دشواری هر سؤال می‌توان نتیجه گرفت که چه تعدادی از داوطلبان توانسته‌اند به آن پاسخ صحیح بدهند. ضریب تمیز به توانایی آزمون در تعیین تفاوت بین گروه‌های قوی و ضعیف گفته می‌شود. به عبارتی ضریب تمیز به بررسی قدرت سؤال از حیث تشخیص و تفکیک داوطلب قوی از ضعیف می‌پردازد. لذا هر قدر ضریب تمیز سوال بزرگتر باشد قوه تمیز آن بیشتر و هر قدر این ضریب کوچک‌تر باشد قوه تمیز آن کمتر است. ضریب تمیز صفر حاکی از آن است که سوال بین گروه قوی و ضعیف تمایزی ایجاد ننموده است(۹).

اگر در تفسیر ضریب دشواری یا ضریب سهولت سوال با آزمون‌های وابسته به هنجار سروکار داریم، می‌توانیم این ضریب را از دیدگاه آماری مورد بررسی قرار دهیم. یکی از انتظارات ما در آزمون‌های وابسته هنجار این است که برای آزمون‌شوندگان مختلف نمراتی بدست دهد که در طول پیوستار پراکنده باشد و هرچه این پراکندگی بیشتر باشد بهتر است. به سخن دیگر هر چه واریانس نمرات حاصل از یک آزمون وابسته به هنجار بزرگ‌تر باشد آن آزمون بهتر است. اما نتیجه آرمانی آزمون‌های وابسته به ملاک این است که در آنها اکثر آزمون‌شوندگان اکثر سوال‌ها را درست جواب دهند. به عبارتی توزیع نمرات دارای کجی[1]منفی خواهد بود. در چنین حالتی ضریب دشواری سوالات بسیار بزرگ و ضریب تمیز آنها بسیار کوچک خواهد بود. یعنی برای گروه آزمودنی مورد نظر اکثر سوال‌ها آسان و قدرت تمیز آن‌ها بسیار کم خواهد بود. لذا در آزمون‌های وابسته به ملاک ضریب تمیز صفر (زمانی‌که کلیه فراگیران به سوال پاسخ صحیح داده‌اند) نشان‌دهنده کاستی در سوال نیست چرا که بر اساس اهداف آموزش مورد توافق اولیه (تعیین نقطعه شروع آموزش)، کلیه فراگیران باید به این سوال پاسخ صحیح بدهند چون درفرآیند یاددهی-یادگیری با آزمون‌های تشخیص، تکوینی، بسیاری از نقاط ضعف آموزشی با تلاش استاد، دانشجو، محیط آموزشی و... شناسایی و برطرف شده است. لذا استفاده از ضریب دشواری و تمییز به شرحی که گذشت صرفاً برای آزمون‌های وابسته به هنجار مناسب است. اما در آزمون‌های وابسته به ملاک این ضرایب مورد استفاده چندانی ندارند(۹). آزمون‌های وابسته به ملاک دارای یک نمره مرزی برای حد تسلط[2] هستند به زبانی دیگر هدف ما این است که فراگیران ما پس از اتمام دوره آموزشی به سطحی از تسلط در اهداف آموزشی رسیده باشند و وظیفه آزمون‌های وابسته به ملاک نیز دقیقاً تفکیک بین گروه فراگیران مسلط و غیرمسلط است. در همین راستا تحلیل این آزمون‌ها نیز باید بر همین اساس اجرا گردد(٦و١٠).

هر چند که روش معمول تحلیل سوال آزمون‌های وابسته به هنجار قابل استفاده برای آزمون‌های وابسته ملاک نیست، با این حال متخصصان روان‌سنجی معتقدند محاسبه ضرایب دشواری و تمیز برای اینگونه سوالات مفید است. از طریق مراجعه به این ضرایب می‌توان نحوه عمل سوالات را وارسی کرد. سوالاتی که ضرایب تمیز صفر یا منفی هستند اشکال اساسی دارند که باید برطرف شوند. همچنین گزینه‌های انحرافی این سوالات نیز باید تحلیل شوند تا معایب آنها اصلاح گردد. بنابراین در آزمون‌های وابسته به ملاک تفسیر ضریب دشواری و تمیز بایستی صرفاً توسط *مدرس* و با توجه به اهداف آموزشی و رسیدن به یادگیری در حد تسلط صورت گیرد.

منابع

[1] Skewness
[2] Mastery limit

[Downloaded from ijme.mui.ac.ir on 2026-11-02]
[DOI: 10.48305/23.0.163]

1. Shakurnia AH, Ghafourian M, Khodadadi A, Ghadiri A, Amari A. [Analytical Study of Quantitative Indices of Multiple-choice Questions of Immunology Department in Ahvaz Jundishapur University of Medical Sciences]. Educational Development of Jundishapur. 2017; 9(2): 72-83.[Persian]
2. Khajeali N, Aslami M, Araban M. [Analysis of medical and dentistry basic sciences examinations: A case study]. Payesh 2020; 19 (4): 383-389.[Persian]
3. Kazemi A, Ehsanpour S, Hassanzadeh A. [Investigating the Academic Achievement Evaluation of Specialized Theoretical Courses of Midwifery BS]. Iranian Journal of Medical Education. 2010; 9 (4): 346-355.[Persian]
4. Baharvand M,Hoseinzadeh M, Jaberiansari Z, Abbaszadeh E, Mortazavi H. [Evaluation of 30 Structural Validity and Content Quality Indices of Theoretical Dental Exams]. Journal of Mashhad Dental School. 2014; 38(4): 291-302.[Persian]
5. Ashraf Pour M, Beheshti Z, Molook Zadeh S. [Quality of final examination in students of Babol Medical University, 1999-2000]. Journal of Babol University of Medical Sciences. 2003; 5 (5): 42-47.[Persian]
6. Dent J, Harden RM, Hunt D. A practical guide for medical teachers. Amsterdam, Netherlands: Elsevier health sciences; 2021.
7. Mohan R, Measurement, evaluation and assessment in education. New Delhi: PHI Learning Pvt; 2023.
8. Montgomery PC, Connolly BH. Connolly, Norm-referenced and criterion-referenced tests: Use in pediatrics and application to task analysis of motor skill. Phys Ther. 1987; 67(12): 1873-1876.
9. Saif AA.[Andazehgiri, Sanjesh, Arzyabiye Amoozeshi]. Tehran: Duran publication.[Persian]
10. Turnbull JM. What is… normative versus criterion-referenced assessment. Med Teach. 1989; 11(2): 145-50.

[Downloaded from ijme.mui.ac.ir on 2026-11-02]

[DOI: 10.48305/23.0.163]