

بسمه تعالی

نحوه آنالیز سوالات آزمون

یکی از اقداماتی که با هدف ارتقاء روند ارزیابی فراگیران و دستیابی به آزمون های معتبر و با کیفیت، انجام می گیرد، آنالیز آزمون است. بنابراین، لازم است آنالیز آزمون ها انجام گرفته و با اصلاح و برطرف ساختن نقاط ضعف هر آزمون، زمینه افزایش کیفیت آن را فراهم آوریم. به منظور آنالیز آزمون، نیازمند شاخص هایی هستیم که اطلاعات خوبی در مورد تک تک آیتم های آزمون در اختیار ما بگذارند، منظور از آیتم، سوالات امتحان چند گزینه ای، شفاهی، تشریحی و حتی ایستگاه های یک آزمون OSCE است. دستیابی به این شاخص ها هم به صورت الکترونیکی و هم به صورت دستی امکان پذیر است. قاعدتا، محاسبه و تعیین این شاخص ها به صورت دستی کار نسبتا دشواری است. بدین ترتیب خوشبختانه در اکثر دانشگاه های کشور، امکان انجام آنالیز آزمون به روش الکترونیک در قالب سامانه های متنوعی که به این منظور طراحی و راه اندازی شده اند، فراهم شده است. در دانشگاه ما نیز امکان آنالیز سوالات آزمون ها، از طریق سامانه فرادید فراهم شده است. در واقع، در این سامانه، آنالیز سوالات آزمون یا به عبارت دقیق تر، تحلیل آیتم بر مبنای اطلاعات وارد شده از آزمون های برگزار شده انجام می گیرد. در این فرایند، اطلاعات حاصل از پاسخنامه های شرکت کنندگان هر آزمون، وارد و ثبت گردیده و بر مبنای داده های حاصل از آن، آنالیز سوالات آزمون انجام می گیرد. سه شاخص اصلی و مهم در تحلیل سوالات آزمون به ترتیب ضریب جذب گزینه ها، ضریب تمیز و ضریب دشواری سوال می باشد که از طریق این سامانه برای تک تک سوالات یا به عبارت اختصاصی تر برای تک تک آیتم ها محاسبه می شود و ما بر مبنای آنها می توانیم در مورد مناسب و یا نامناسب بودن هر یک از سوالات (آیتم های) آزمون تصمیم گیری نماییم. تصمیم گیری برای هر یک از سوالات آزمون به سه حالت انجام می گیرد، بدین معنا که ما بر مبنای شاخص های محاسبه شده با سه تصمیم برای سوال مواجه می شویم که به ترتیب ذیل خواهد بود:

- اول؛ سوال در هر سه شاخص عنوان شده عملکرد ضعیفی داشته و قابل اصلاح نمی باشد که بدین ترتیب از بانک سوالات آزمون حذف می گردد.
- دوم؛ سوال در هر سه شاخص عنوان شده عملکرد مناسبی داشته که در این صورت در بانک سوالات آزمون باقی مانده و حفظ می گردد.
- سوم؛ سوال در برخی شاخص ها عملکرد ضعیفی دارد و بدین ترتیب اصلاح می شود.

شاخص هایی که در این سامانه برای تک تک سوالات آزمون محاسبه و ارائه می شوند و همچنین نحوه تفسیر آنها به شرح ذیل است:

الف) ضریب جذب گزینه ها

در آزمون های چند گزینه ای یکی از مواردی که می تواند اطلاعات خوبی در مورد سؤال در اختیار طراح قرار دهد، محاسبه در صد جذب هر گزینه است. گزینه ای که توسط هیچ کدام از دانشجویان انتخاب نشود، مناسب نیست. هنگام طراحی سؤال چنانچه استادان گزینه ای مناسبی پیدا نکنند، به مواردی روی می آورند که گزینه های انحرافی ضعیفی هستند و دانشجویان به سرعت نادرستی آنها را تشخیص می دهند. به جدول زیر توجه کنید.

تعداد دانشجویانی که گزینه را انتخاب کرده اند.

شماره سوال	پاسخ صحیح	الف	ب	ج	د	نرده
۱	د	۰	۱۸	۲۲	۸۰	۰
۲	الف	۲۰	۲۵	۳۰	۳۲	۱۳

در سوال اول، هیچ کس گزینه الف را انتخاب نکرده است. این نشان می دهد که اشتباه بودن آن به طرز واضحی برای همه دانشجویان مشخص بوده است. بهتر است برای استفاده در آزمون های بعدی گزینه الف با گزینه مناسب تری جایگزین شود. چنانچه تعداد کمی از دانشجویان (معمولاً کمتر از ۵ درصد) جذب گزینه ای شوند، به این معنا است که گزینه انحرافی جذاب نبوده است که به آن گزینه غیر عملکردی می گویند. این ضریب از طریق بررسی درصد پاسخگویی آزمون شوندگان به هر گزینه به آسانی قابل بررسی است.

ب) ضریب دشواری آیتم

ضریب دشواری که به صورت p نشان داده می شود و به آن p value نیز گفته می شود. برای یک آیتم (به عنوان مثال یک سوال در آزمون چند گزینه ای یا یک ایستگاه در OSCE) تعریف می شود. ضریب دشواری درصد شرکت کنندگانی است که به آیتم مورد نظر جواب درست داده اند و می تواند مقداری بین صفر تا یک داشته باشد. به این ترتیب هر چقدر تعداد دانشجویانی که به آیتم پاسخ داده باشند، بیشتر باشد، منطبقاً آن آیتم آسان تر است در حالی که بنا به تعریف ضریب دشواری آن بزرگتر است.

در سامانه فرادید، محاسبه ضریب دشواری صورت گرفته و با عبارت (diff-index) و در چهار حالت آسان، دشوار، نسبتاً آسان و نسبتاً دشوار نشان داده می شود. در واقع اکثر نرم افزارهایی که برای تصحیح پاسخنامه ها به کار می روند. دارای افزونه ای به این منظور هستند و همزمان با اعلام نمرات دانشجویان قادر هستند، گزارشی از ضریب دشواری سوالات ارائه دهند. هر چند گاهی استاد قصد دارد خودش ضریب دشواری سوالات را برای یک آزمون کلاسی محاسبه کند. در هر دو حالت، روش محاسبه تفاوتی ندارد.

برای محاسبه ضریب دشواری سوالات در آزمون چند گزینه ای، باید تعداد کل افرادی که به هر سؤال پاسخ صحیح داده اند مشخص کنیم و بر تعداد کل دانشجویان تقسیم کنیم:

$$\text{ضریب دشواری سوال} = \frac{\text{دانشجویانی که به سوال پاسخ صحیح داده اند}}{\text{تعداد کل دانشجویان}}$$

به مثال زیر که در خصوص داده های جدول است، توجه کنید :

$$\frac{20}{120} = 0.16 \text{ ضریب دشواری سوال یک}$$

$$\frac{80}{120} = 0.66 \text{ ضریب دشواری سوال دو}$$

ضریب دشواری هر سؤال مطابق فرمول بالا محاسبه شده است.

در این میان ، سؤال شماره دو سخت تر از سوال شماره یک است.

در آزمون OSCE نیز می توان حد نصاب قبولی را در سطح ایستگاه تعیین کرد و به این ترتیب مشخص کرد که در هر ایستگاه چند نفر قبول شده اند. به عنوان مثال، در یک آزمون OSCE از بین ۱۰۰ دانشجو، ۷۵ نفر توانسته اند در ایستگاه سونداژ قبول شوند پس ضریب دشواری ایستگاه، ۷۵ درصد است. همان طور که مشخص است به سادگی می توان گفت درصد قبولی ایستگاه، معادل ضریب دشواری آن است .

برای محاسبه دستی ضریب دشواری، در صورتی که تعداد دانشجویان زیاد باشد (بیش از ۴۰ نفر) سخت است که پاسخنامه همه دانشجویان بررسی شود . به همین دلیل دانشجویان را به دو دسته ضعیف و قوی تقسیم می کنند. برای این کار دانشجویان بر اساس نمره کل ردیف می شوند. سپس، برگه های یک سوم بالایی (گروه قوی) و یک سوم پایینی (گروه ضعیف) جدا می شوند. ضریب دشواری به صورت فرمول زیر قابل محاسبه است.

$$\text{ضریب دشواری هر سؤال} = \frac{\text{تعداد دانشجویان قوی که به سوال پاسخ صحیح داده اند} + \text{تعداد دانشجویان ضعیف که به سوال پاسخ صحیح داده اند}}{\text{تعداد کل دانشجویان}}$$

به عنوان مثال در یک آزمون دانشجویان بر اساس نمره کل به دو گروه ۴۰ نفره قوی و ضعیف تقسیم شدند. در گروه قوی ۳۸ دانشجو به سؤال شماره ۲۰ پاسخ صحیح داده اند و در گروه ضعیف ۱۴ نفر به این سؤال شماره پاسخ صحیح داده اند . ضریب دشواری این سؤال این صورت زیر محاسبه می شود:

$$\frac{30 + 40}{80} = 0.65$$

در آزمون های معیار محور، به جای تقسیم بندی دانشجویان به دو گروه ضعیف و قوی، آنها را به دو گروه رد و قبول تقسیم می کنیم اما نحوه محاسبه ضریب دشواری برای این آزمون تغییری نمی کند.

$$\text{ضریب دشواری سوال} = \frac{\text{تعداد دانشجویان قبول که درست جواب اند} + \text{تعداد دانشجویان رد که درست جواب داده اند}}{\text{تعداد دانشجویان دو گروه}}$$

به مثال زیر توجه کنید:

در یک آزمون چند گزینه ای پایان ترم (معیار محور است) ۲۰ دانشجو قبول و ۵ دانشجو رد شده اند. تعداد کسانی که به سؤال شماره ۱۰ جواب درست داده اند به ترتیب در گروه اول قبول و رد ، ۹ و ۱ نفر بوده است . ضریب دشواری این سؤال به صورت زیر حساب می شود:

$$\frac{9 + 1}{25} = 0.4$$

به مثال زیر در خصوص آزمون OSCE توجه کنید:

در یک آزمون OSCE از بین ۱۰۰ دانشجو ۹۵ نفر توانسته اند در کل آزمون قبول شوند . میزان قبولی در ایستگاه تعبیه سونداژ در گروه قبول، ۷۳ و در گروه رد، ۲ نفر بوده است. ضریب دشواری ایستگاه به صورت زیر قابل محاسبه است:

$$\frac{73 + 2}{100} = 0.75$$

آنچه تا اینجا ذکر شد، برای آزمون هایی کاربرد دارد که جواب دو ارزشی دارند مانند سؤالات چند گزینه ای به عبارت دیگر، دانشجویان در مواجهه با هر سؤال یا به آن پاسخ صحیح داده اند و تمام نمره را دریافت کرده اند یا پاسخ اشتباه داده اند و صفر شده اند. بنابراین همین روش برای تعیین ضریب دشواری در یک آزمون با سؤالات درست-نادرست و جور کردنی گسترده نیز می تواند به کار رود.

در آزمون هایی که پاسخ به صورت صفر و یک نیست، قضیه کمی متفاوت است. آزمون تشریحی، سؤالات شفاهی و OSCE از این دست سؤالات هستند که نمره ای که می تواند به پاسخ آنها تعلق بگیرد. به صورت یک محدوده نمره است. محاسبه ضریب دشواری این سؤالات به صورت محاسبه میانگین نمرات دانشجویان در آن سؤال است که بر نمره کل سؤال (یعنی نمره ای که دانشجویان می توانند به صورت بالقوه کسب کنند) تقسیم می شود.

$$\text{ضریب دشواری سوال} = \frac{\text{میانگین نمرات دانشجویان}}{\text{نمره سوال}}$$

دو ایستگاه OSCE یا دو سؤال تشریحی را در نظر بگیرید که حداکثر نمره در آنها ۱۰ است میانگین نمرات آنها را محاسبه می کنیم که ۷/۵ و ۳ به دست می آید. بنابراین، ضریب دشواری دو ایستگاه /دو سؤال به ترتیب ۰/۷۵ و ۰/۳ است همان طور که مشخص است ایستگاه با ضریب دشواری ۰/۳ برای دانشجویان سخت تر بوده است.

در صورتی که نمرات دانشجویان را به صورت دسته بندی شده داشته باشیم، می توان از فرمول زیر که در واقع تغییر یافته فرمول بالاست، استفاده کرد:

$$\text{ضریب دشواری سوال} = \frac{\text{مجموع نمرات دانشجویان}}{\text{تعداد دانشجویان} \times \text{نمره سوال}} = \frac{\text{میانگین نمرات دانشجویان}}{\text{نمره سوال}}$$

به عنوان مثال: در یک آزمون تشریحی با شرکت ۵۰ دانشجو، یک سؤال تشریحی سه نمره دارد و مقادیری که برای پاسخ های مختلف دانشجویان در آن می توان در نظر گرفت، شامل ۰، ۱، ۲ و ۳ است. با بررسی برگه های دانشجویان، مشخص می شود که به ترتیب ۱۰، ۲۵، ۱۰ و ۵ دانشجو نمرات فوق را کسب کرده اند.

$$\text{ضریب دشواری سوال } ۰/۴۶ = \frac{(۵ \times ۳) + (۱۰ \times ۲) + (۲۵ \times ۱) + (۱۰ \times ۰)}{۵۰ \times ۳}$$

در این گونه سؤالات نیز اگر به علت تعداد زیاد دانشجویان آنان را به دو گروه قوی و ضعیف تقسیم می کنیم، می توان از فرمول زیر استفاده کرد:

$$\text{ضریب دشواری سوال} = \frac{\text{میانگین نمرات دانشجویان ضعیف} + \text{میانگین نمرات دانشجویان قوی}}{\text{نمره سوال} \times \text{تعداد دانشجویان دو گروه}}$$

توزیع نمرات دانشجویان در یک ایستگاه OSCE با چک لیست ۱۰ نمره ای

نمره	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
تعداد دانشجویان گروه قوی	۰	۰	۱	۰	۴	۵	۸	۸	۵	۵	۴
تعداد دانشجویان گروه ضعیف	۱	۳	۵	۷	۵	۸	۴	۵	۲	۰	۰

ضریب دشواری ایستگاه:

$$\frac{۱۶۹ + ۲۷۲}{۸۰ \times ۱۰۰} = ۰/۵۵$$

اولین مساله ای که در تفسیر ضریب دشواری باید مد نظر قرار بگیرد، نوع آزمون از لحاظ هدفی است که دنبال می کند باید مشخص شود که آیا آزمون مبتنی بر معیار بوده است یا نتایج آن به طور هنجاری تفسیر می شوند .

در آزمون هنجاری، بهتر است نمرات پراکندگی خوبی داشته باشند یعنی واریانس آنها بالا باشد. زیرا اگر نمرات در یک محدوده باریک تجمع پیدا کنند، یعنی نمرات دانشجویان نزدیک هم یا مشابه شود، نمی توان تمایز بین دانشجویان قائل شد و رتبه بندی درستی انجام داد. واریانس سؤال با ضریب دشواری (p) رابطه ای دارد که به صورت زیر است.

$$\text{واریانس سؤال} = p \times (1-p)$$

طبق فرمول فوق، اگر ضریب دشواری سؤال ۱ یا ۰ باشد یعنی سؤال بسیار آسان یا بسیار سخت باشد، واریانس به کمترین مقدار می رسد و صفر می شود. هنگامی واریانس به حداکثر مقدار خود می رسد که ضریب دشواری برابر ۰/۵ باشد در این صورت واریانس سؤال ۰/۲۵ می شود. بنابراین می توان گفت که در آزمون هنجاری بیشتر سوالاتی مورد نیاز هستند که واریانس بزرگتری داشته باشند یعنی میزان دشواری آنها متوسط باشد.

در مورد سوالات چند گزینه ای از آنجا که یکی از نگرانی های رایج، پاسخ دادن دانشجویان بر اساس حدس و گمان است تلاش شده است تا این دغدغه لحاظ شود و برای جلوگیری از آن گفته می شود بهتر است ضریب دشواری آزمون هنجاری چند گزینه ای کمی بالاتر از سطح متوسط باشد. برای به دست آوردن عدد دقیق تر، باید احتمال پاسخ حدسی را برآورد کرد. احتمال اینکه دانشجویان با حدس به پاسخ سؤال چهار گزینه ای برسند، ۲۵ درصد است. بنابراین میزان پایه به جای صفر، ۰/۲۵ در نظر گرفته می شود. در این حالت، حدوسط یعنی نقطه میانی بین ۰/۲۵ و ۱ محاسبه می شود که ۰/۳۷۵ است. به عبارت دیگر، ضریب دشواری مطلوب دشواری بین ۰/۳ تا ۰/۷. بیشترین اطلاعات را در مورد تفاوت دانشجویان به دست می دهد، هر چند که عوامل دیگری نیز تعیین کننده هستند.

در آزمون معیاری، مانند ارزیابی پایان نیمسال، به طور معمول انتظار می رود که اگر دانشجویان به صورت فعال در دوره مشارکت کرده باشند و فعالیت های یاددهی-یادگیری را انجام داده باشند، مطالب را فرا بگیرند و به اهداف دوره برسند از طرفی سوالات نیز به گونه ای طراحی می شوند که دانشجویانی که در دوره شرکت کرده اند و منابع را مطالعه نموده اند بتوانند به آنها پاسخ دهند. به همین دلیل، اکثر دانشجویان به اکثر سوالات جواب صحیح می دهند. به عبارت دیگر، عجیب نیست که در آزمون معیاری، ضریب دشواری سوالات بزرگ و نزدیک ۱ باشد.

پس این باور که ضریب دشواری باید متوسط و بین ۰/۳ تا ۰/۷ باشد. هر چند رایج است اما دقیق نیست. از آنجا که در ارزیابی فراگیران علوم پزشکی، جز چند نمونه محدود، اکثر با آزمون های معیاری سر و کار داریم، توجه به این نکته قبل از تفسیر ضریب دشواری ضروری است .

ج) ضریب تمیز آیتم

سومین شاخصی که در آنالیز سوالات آزمون حائز اهمیت است و محاسبه می شود، ضریب تمیز است که به صورت D نشان داده می شود، و مانند ضریب دشواری برای یک آیتم تعریف می شود. در خروجی سامانه فرادید ضریب تمیز با عبارت (disc-index) نشان داده می شود و در سه حالت کم، مناسب و نامناسب قرار می گیرد. ضریب تمیز مشخص می کند که یک آیتم تا چه میزان قادر به ایجاد تمایز و افتراق بین دانشجویان ضعیف و قوی است. هر چقدر ضریب تمیز بزرگتر باشد، به این معنا است که قدرت آیتم در جدا کردن دانشجویان قوی از ضعیف بیشتر بوده است. ضریب تمیز می تواند مقداری بین منفی یک تا مثبت یک داشته باشد.

به منظور محاسبه ضریب تمیز در سوالات بسته پاسخ، شرکت کنندگان باید بر اساس نمره ای که در کل آزمون کسب کرده اند، به دو دسته پایین (ضعیف) و بالا (قوی) تقسیم شوند. سپس باید مشخص شود که چه تعداد از هر دسته به آیتم مورد نظر جواب درست داده اند. تفاضل این دو مقدار، ضریب تمیز را مشخص می کند:

$$\text{ضریب تمیز} = \frac{\text{تعداد دانشجویان ضعیف که درست جواب داده اند} - \text{تعداد دانشجویان قوی که درست جواب داده اند}}{\text{تعداد دانشجویان یک گروه}}$$

به مثال زیر توجه کنید:

در یک آزمون چهار گزینه ای، دانشجویان بر اساس نمره کل به دو گروه ۴۰ نفره قوی و ضعیف تقسیم شدند که به ترتیب ۳۸ و ۱۴ نفر ایشان به سؤال ۲۰ پاسخ صحیح داده اند. ضریب تمیز این سؤال به صورت زیر محاسبه می شود:

$$\frac{38 - 14}{40} = \frac{24}{40} = 0.6$$

در خصوص تفسیر ضریب تمیز باید گفت که هر چه قدر ضریب تمیز سؤال به یک نزدیک تر باشد، به این معنا است که سؤال دانشجویان قوی و ضعیف را بهتر از یکدیگر جدا می کند. ضریب تمیز معادل صفر به این معنا است که سؤال اصلاً قادر نیست دانشجویان دو گروه را از هم تفکیک کند. ضریب تمیز می تواند مقادیر منفی نیز به خود بگیرد که این حالت به این معنا است که دانشجویان گروه پایین، بهتر به سؤال جواب داده اند. در آزمون های پیشرفت تحصیلی این وضعیت مطلوب نیست و باید به دنبال علت آن بود. یکی از اولین مواردی که در سؤال با ضریب تمیز منفی به ذهن می رسد، این احتمال است که کلید سؤال اشتباه اعلام شده باشد. احتمال دیگر، وجود نکته انحرافی یا خطای ساختاری در سؤال است که موجب شده دانشجویان طی دوره مبحث مورد نظر را به صورت اشتباه یاد گرفته اند.

در تفسیر ضریب تمیز همانند تفسیر ضریب دشواری مهم است که هدف آزمون در نظر گرفته شود. در یک آزمون هنجاری، سؤالات باید قادر باشند، بین دانشجویان تمایز ایجاد کنند. اگر سؤالات قادر با افتراق دانشجویان رقابت کننده نباشند، نمرات شرکت کنندگان مشابه و نزدیک هم خواهد شد و انتخاب نمرات برتر را با دشواری مواجه می سازد. پس

لازم است سوالاتی با ضریب تمیز بالا یعنی نزدیک به یک داشته باشیم. در حالی که در آزمون های معیاری، مورد انتظار است که اکثر فراگیران به اکثر سوالات به درستی جواب دهند، بنابراین دور از انتظار نیست که ضریب تمیز سوالات چندان بالا نباشد.

تهیه و تنظیم: واحد آموزش و ارزشیابی بالینی، مرکز مطالعات و توسعه آموزش علوم پزشکی دانشگاه علوم پزشکی شهرکرد