

بسمه تعالی

به منظور ارتقاء هر چه بیشتر سطح کیفی روشهای ارزیابی دانشجویان، لازم است هر سوال دارای روایی مناسبی باشد یعنی بتواند حیطه های مورد نظر دانش فراگیر را به درستی بسنجد. روایی هرآزمون را می توان از نحوه پاسخ شرکت کنندگان در آزمون ها مورد ارزیابی قرار داد و هر گروه آموزشی بهتر است این کار را با توجه به نتایج آزمونهای سالهای گذشته و یا با توجه به بازخوردهای دانشجویان، قبل از برگزاری آزمون انجام داد و در صورت نیاز سوال یا سوالات آن درس بازنگری و اصلاح گردد. روایی سوالات آزمونها از جنبه های زیر قابل بررسی است:

روشهای کمی تحلیل آزمون

جهت تحلیل و ارزیابی سوالات آزمون های تراکمی در هر نیمسال از شاخصهای زیر استفاده خواهد شد:

۱- شاخص دشواری یا سهولت سوال^۱ :

تعریف شاخص: شاخص دشواری یا سهولت سوال عبارت است از نسبت افرادی که به یک سوال پاسخ صحیح داده اند.

تفسیر شاخص: شاخص دشواری یک شاخص وارونه است. هر چه مقدار این شاخص کوچکتر باشد (به صفر نزدیکتر باشد) سوال مشکل تر و هر چه مقدار آن بزرگتر باشد (به ۱۰۰ نزدیکتر باشد) سوال آسان تر است. در آزمونهای هنجاری سوالهای مناسب و قابل قبول سوالهای هستند که سطح دشواری آنها در حد متوسط باشد. سوال های با سطح دشواری متوسط امکان پراکندگی درنمره آزمون شوندگان را افزایش می دهند.

$$\text{شاخص دشواری یا سهولت سوال} = \frac{\text{تعداد پاسخهای درست}}{\text{تعداد شرکت کنندگان در هر گروه}} \times 100$$

دامنه درجه دشواری :

$$0 \text{ --- } 25\% \text{ --- } 45\% \text{ --- } 65\% \text{ --- } 85\% \text{ --- } 100$$

- فوق العاده دشوار = ۰ تا ۲۵٪
- خیلی دشوار = ۲۵٪ تا ۴۵٪
- دشوار = ۴۵٪ تا ۶۵٪
- متوسط = ۶۵٪ تا ۸۵٪
- ساده = ۸۵٪ تا ۱۰۰٪

۲- شاخص تمیز^۲ :

تعریف شاخص: قدرت تمیز یا تفکیک سوال عبارت است از میزان قدرت سوال در تفکیک آزمون شوندگان دارای نمره بالا، از آزمون شوندگان دارای نمره پائین. به عبارت دیگر این شاخص مشخص می نماید که آزمون چقدر توانسته است بین گروه قوی و ضعیف افتراق بگذارد.

^۱ Item Facility

^۲ Item Discrimination

تفسیر شاخص: هرچه عدد ضریب تمیز بزرگتر باشد سوال از توان بیشتری در جدا کردن و تفاوت قائل شدن میان گروه قوی و ضعیف برخوردار می باشد. منفی بودن ضریب تمیز به معنی آن است که گروه قوی از گروه ضعیف عملکرد پایین تری داشته اند. این مشکل به طور معمول ناشی از ایراد خود سوال می باشد و چنانچه سوال دارای اشکال ساختاری نباشد، افراد گروه بالا یا آن سوال را به طور کامل یاد نگرفته اند یا آن را به غلط آموخته اند، دامنه قدرت تمیز سوال بر اساس نمودار ذیل تفسیر می شود.



- خوب : بزرگتر از ۰.۳۰٪
- متوسط: بین ۰.۱۰٪ تا ۰.۳۰٪
- ضعیف : بین ۰ تا ۰.۱۰٪
- منفی : کم تر از صفر

اگر شاخص دشواری حدود صفر باشد، سوال فاقد قدرت تفکیک است. در چنین حالتی یا مطلب بطور کامل تدریس نشده است، یا متن سوال مبهم و ناروا می باشد. اگر مطلب تدریس شده است، احتمالاً مشکل در ارائه متن سوال یا گزینه های پیشنهادی است. لذا باید متن سوال و الگوی پاسخهای داده شده بررسی شود. اگر نسبت زیادی از آزمون شوندگان یک گزینه انحرافی را به عنوان گزینه کلید انتخاب کرده باشند، متن گزینه انتخاب شده احتمالاً گمراه کننده است. و احتمال دیگر وجود اشتباه در کلید تصحیح گزینه ها می باشد.

اگر شاخص قدرت تمیز منفی باشد، علت را باید در گزینه های پاسخ سوال جستجو کرد. در هر صورت سوالهای با قدرت تمیز منفی سوال فاقد ارزش هستند و باید حذف شوند.

اگر شاخص قدرت تفکیک مثبت اما پائین باشد، به سطح دشواری سوال توجه نمایید. سوالهای خیلی ساده یا خیلی دشوار از قدرت تفکیک کمی برخوردار می باشند. اگر سطح دشواری سوال مطلوب باشد، به الگوی پاسخها و متن سوال توجه شود.

رابطه بین شاخص های تحلیل آزمون:

- هر قدر ضریب دشواری به ۰/۵ و ضریب تمیز به ۱ نزدیکتر باشد آن سوال، سوال بهتری است .
- هر قدر سوال دارای واریانس بزرگتری باشد و با سایر سوالها همبستگی بیشتری داشته باشد به واریانس کل آزمون کمک بیشتری می کند. اگر در تحلیل آزمون به انتخاب سوال هایی بپردازیم که **همبستگی زیادی با یکدیگر** دارند و در ضمن دارای واریانس های بزرگتری هستند در مجموع آزمون بهتری را تدارک دیده ایم.
- زمانی که ضریب دشواری سوال ۰/۵ باشد، واریانس سوال به حداکثر خود یعنی ۰/۲۵ خواهد رسید. در این صورت **حداکثر اطلاعات درباره تفاوت بین آزمودنیها به دست خواهد آمد.**
- با انتخاب سوالاتی با ضریب تمیز بالا، ضریب دشواری متوسط، **ضریب پایایی آزمون افزایش** پیدا می کند.
- هر چقدر ضریب پایایی آزمون به ۱ نزدیکتر باشد از میزان واریانس خطا کاسته و نشان دهنده این است که **واریانس مشاهده شده واریانس حقیقی است.**

۳- تحلیل گزینه های انحرافی ۳:

تعریف شاخص: گزینه های انحرافی گزینه های اشتباهی می باشند که همراه با گزینه پاسخ یا صحیح آورده می شوند تا توجه آزمودنی های ناآگاه و کم اطلاع را از گزینه پاسخ یا صحیح منحرف کنند.

تفسیر شاخص: در گزینه های انحرافی قاعده کلی به شرح زیر است:

هر گزینه انحرافی باید حداقل یک نفر از افراد گروه ضعیف را به خود جلب کند و اگر هر دو گروه قوی و ضعیف را به خود جلب کند، تعداد افراد گروه ضعیف باید بیشتر از گروه قوی باشد. چنانچه گزینه ی انحرافی خاصی وجود داشته باشد که هیچ یک از افراد گروه قوی و ضعیف آن را انتخاب نکرده باشند، گزینه انحرافی مناسبی نیست و باید اصلاح شود.

روشهای کیفی تحلیل آزمون

۱-روایی محتوا(Content Validity):

۱-۱- تاکسونومی سوال:

تعریف شاخص: تاکسونومی، طبقه بندی هدفهای آموزشی است که از طریق سوالات مطرح شده در آزمون سنجیده می شوند. تفسیر شاخص: تاکسونومی حداقل دارای سه حیطه دانش، درک و فهم، و کاربرد سوال می باشد. گاه در سطوح بالاتر، مانند تحصیلات تکمیلی نیاز به طرح سوال در تاکسونومی بالاتری همچون تحلیل، سنتز یا قضاوت نیز وجود دارد.

مفهوم دانش در تاکسونومی سوال) تاکسونومی سطح ۱: دانش شامل به یادآوری (بازخوانی و بازشناسی) می باشد. در واقع دانش در تاکسونومی عبارت است از به یادآوری پاسخهای صحیحی که قبلاً در موقعیت یادگیری و تمرین آموخته شده اند.

مفهوم درک و فهم در تاکسونومی سطح ۲: هنگامی است که ذهن با انجام یک مرحله فرایند به پاسخ می رسد. به عبارتی، توانایی درک منظور یا مقصود یک مطلب، یعنی فهمیدن به عملکردهایی فراتر از آنچه قبلاً تمرین و آموخته شده اند نیاز دارد. این عملکردهای اضافی شامل: ترجمه یا برگردان، تفسیر و تعبیر، خلاصه کردن و کشف شباهتها و تفاوتها.

مفهوم کاربرد در تاکسونومی سطح ۳: هنگامی که ذهن با انجام دو مرحله فرایند به پاسخ می رسد.

لذا در پایان، باید تعداد سوالاتی که در هریک از تاکسونومی فوق قرار دارند به تفکیک سطح گزارش شود.

لذا برای هریک از سوالات لازم است نسبت به ایجا ستونی در جلوی هر سوال و تفکیک سطح سوال در حیطه شناختی اقدام شود.

۲-۲- ضرورت سوال:

از این شاخص می توان به خصوص در آزمونهای تحصیلات تکمیلی استفاده شود در زمانی که تاکسونومی سوالات عمدتاً بر مبنای سنجش سطوح اولیه دانش است.

ضرورت سوال دارای سه حیطه است:

(۱) **(Must learn)**: موضوعاتی که یادگیری آنها از سوی آزمون شونده برای رسیدن به تبحر و مهارت لازم در رشته مورد

نظر و حرفه آینده، لازم و ضروری است (تا ۸۰٪ سوالات را تشکیل می دهد).

(۲) **(better to learn)**: موضوعاتی که یادگیری آنها به لحاظ ضرورت و الزامی بودن در درجه پائین تری از حیطه اول قرار

دارند و بهتر است آزمون شونده برای بالاتر بردن دانش و مهارت در رشته و حرفه خود این موضوعات را بیاموزد (تا ۱۵٪

سوالات را تشکیل می دهد).

۳) (nice to learn): موضوعاتی که یادگیری آنها صرفاً جهت داشتن اطلاعات بیشتر در زمینه رشته و شغل مطرح است و به لحاظ اهمیت در درجه پائین تری نسبت به دو حیطه دیگر قرار دارد (تا ۵٪ سوالات را تشکیل می دهد).

۲- روایی ساختاری آزمون (construct validity):

رعایت قواعد ساختاری سوال یا اعتبار ساختار (Construct Validity) با استفاده از پرسشنامه میلمن

این چک لیست، قواعد تهیه سوال های چند گزینه ای را نمره دهی می کند. قواعدی که برای تهیه سوال های چند گزینه ای پیشنهاد می شوند براساس این طرز فکر تدوین شده اند که یک سوال چند گزینه ای باید آزمون شونده را با تکلیفی که هم مهم و کاملاً برای او قابل فهم است مواجه سازد. این چک لیست شامل مواردیست که باید در ساختار یک سوال چند گزینه ای (MCQ) رعایت گردد که در دو قسمت ساقه سوال (stem) و گزینه ها (options ارائه میگردد) به ترتیب گزینه های چک لیست. نحوه نمره دهی از ۰ تا ۲ (خیر، تاحدودی، ۱ و کاملاً، ۲) می باشد.

پرسشنامه میل من (فرم کوتاه شده)

ردیف	موضوع	کاملاً	تاحدودی	خیر
۱	آیا بخش اعظم اطلاعات در تنه سؤال گنجانده شده است ؟			
۲	آیا سؤال یک هدف اختصاصی یادگیری را مورد ارزیابی قرار می دهد؟			
۳	آیا لغات استفاده شده در تنه یا گزینه ها ،شفاف و مستقیم بیان شده اند؟			
۴	آیا از کاربرد گزینه منفی برای تنه منفی خود داری شده است ؟			
۵	آیا از کاربرد گزینه هایی نظیر " همه موارد " " هیچکدام " و گزینه های ترکیبی خوداری شده			
۶	آیا از کاربرد گزینه های متضاد یکدیگر، خوداری شده است؟			
۷	آیا از لغات مثبت در تنه سوال استفاده شده است یا در صورت منفی بودن تنه سوال، لغات منفی مشخص شده اند؟			
۸	آیا این سوال مستقل از سئوالات دیگر می باشد؟			
۹	آیا گزینه ها از نظر طول، ساختار لغوی و سبک نگارش همسنگ هستند؟			
۱۰	آیا تا حد امکان از کاربرد عبارتهای تکراری در گزینه ها خوداری شده است؟			
۱۱	آیا کلمات بکار رفته در تنه و یا گزینه ها از نظر املائی صحیح است ؟			
۱۲	آیا بطور مشخص یک گزینه صحیح می باشد (خوداری از طرح سئوالات با پاسخهای متناقض)؟			
۱۳	آیا گزینه ها به طور عمودی لیست شده است ؟			
۱۴	آیا گزینه های انحرافی، تنها دانشجویان غیر مطلع را گمراه می کند؟			
۱۵	آیا تنه سوال و گزینه ها مثبت اند؟			

روش آنالیز آسان آزمونهای عملکردی و تشریحی

در این روش ضریب دشواری یک سوال به صورت نمره ی میانگین گروه آزمون شوندگان برای آن سوال تعریف شده است .

فرمول آن به این صورت است :

۱- ۲۷٪ از دانشجویانی که بالاترین نمرات را داشتند و ۲۷٪ از دانشجویانی که کمترین نمرات را داشتند، جدا کرده که به ترتیب گروه بالا و پایین نام می گیرند

۲- تعداد پاسخ های صحیح به هر سوال را در گروه بالا و پایین حساب کرده و با هم جمع می کنیم

۳- حاصل جمع بالا را به مجموع دانشجویان در دو گروه بالا و پایین تقسیم می کنیم

مجموع تعداد پاسخ های صحیح به هر سوال در دو گروه = ضریب دشواری سوال

مجموع تعداد دانشجویان در گروه بالا و پایین

ضریب تمییز براساس این روش بدینگونه محاسبه می شود :

۱- تعداد پاسخ های صحیح گروه بالا و پایین را برای هر سوال تعیین می نمایم .

۲- تفاوت بین این تعداد پاسخ های صحیح را محاسبه می کنیم .

۳- حاصل را به تعداد دانشجویان در یک گروه تقسیم می نمایم .

تفاضل تعداد پاسخ های صحیح در دو گروه = ضریب تمییز

تعداد دانشجویان در یک گروه

مثال: تعداد پاسخ های صحیح به سوال ۱ در گروه بالا (۱۰ نفر) مساوی با ۹ است و تعداد پاسخ های صحیح در گروه پایین (۱۰ نفر) مساوی با ۲ است. ضریب دشواری و تمییز برای سوال ۱ برابر است با:

$$\text{ضریب دشواری سوال ۱} = \frac{0}{55} = (9 + 2) \div 20$$

تفسیر: این سوال با شاخص دشواری ۵۵ درصد سوال دشواری محسوب می شود

$$\text{ضریب تمییز} = \frac{0}{7} = (9 - 2) \div 10$$

تفسیر: این سوال به لحاظ شاخص تمییز سوال بسیار خوبی جهت متمایز کردن دانشجویان قوی از ضعیف می باشد

گاهی برای محاسبه میزان سختی سوالات از دانشجویان **گروه متوسط** نیز استفاده می شود که این کار در صورتی که تعداد دانشجویان در گروه متوسط (مابین گروه قوی و ضعیف)، بیش از $\frac{1}{5}$ برابر آنها باشد، ارزش بیشتری دارد. در این صورت، با محاسبه میانگین پاسخ صحیح بین ۳ گروه قوی، متوسط و ضعیف می توان وجود سوالات مخدوش کننده را بهتر ارزیابی کرد، به طوری که اگر میانگین پاسخ صحیح به سوالی در گروه قوی و ضعیف از میانگین آن در گروه متوسط، **کمتر یا بیشتر باشد**، با احتمال زیاد آن سوال، سوال خوبی نبوده و مخدوش گر است. البته تعیین گروه های متوسط در کلاسهای مناسب تر است که تعداد دانشجویان زیاد است (مانند کلاسهای کارشناسی).