



معاونت آموزش، تحقیقات و دانشجویی فرهنگی

دفتر مطالعات و توسعه آموزش پزشکی (EDC)

# راهنمای استاندارد سازی سؤالات و تحلیل و ارزیابی آزمون

تحلیل دقیق ارزیابی آزمون، نه تنها به عنوان یکی از مهمترین بخش روند آموزشی و یکی از مهمترین ابزارهای تضمین کیفیت بهینه در نظر گرفته می شود، بلکه بررسی آن تصویری نقادانه و قابل استناد از اجزای سیستم آموزشی شامل فراگیران (انگیزه ها و تواناییها)، برنامه درسی (اهداف محتوا و روشهای آموزش)، مدرسین و سایر اجزای تأثیر گذار را ارائه می نماید. لذا تحلیل دقیق آزمون بی تردید در زمینه ارتقاء کیفیت سیستم آموزشی بسیار مؤثر خواهد بود.

## ۱- تحلیل سؤال:

تحلیل سؤال مجموعه ای از روش هایی است که از طریق آنها اثر بخشی و کارآیی هر یک از سؤالات آزمون و کل امتحان اندازه گیری می شود. شناسایی عوامل مؤثر در افزایش اثر بخشی و کارآمدی هر یک از سؤالات آزمون، ما را قادر می سازد تا آزمون ها با امتحانات استاندارد و مناسب تری طراحی کنیم، تحلیل سؤالات این امکان را فراهم می سازد تا ویژگی های هر یک از سؤالات آزمون بررسی می شود و بدین سان از قدرت سؤال جهت سنجش توانایی آزمون شونده اطمینان حاصل نمود.

## ۲- شاخص های تحلیل و ارزیابی:

الف- شاخص دشواری

ب- قدرت تمیز (افتراق)

ج- گزینه های انحرافی

۲-۱ شاخص دشواری یا سهولت سؤال:

تعریف شاخص: شاخص دشواری یا سهولت سؤال عبارت است از نسبت افرادی که به یک سؤال پاسخ صحیح داده اند. تفسیر شاخص: شاخص دشواری یک شاخص وارونه است. هر چه مقدار این شاخص کوچکتر باشد. (به صفر نزدیکتر باشد) سؤال مشکل تر و هر چه مقدار آن بزرگتر باشد (به ۱۰۰ نزدیکتر باشد) سؤال آسان تر است. در آزمونهای هنجاری (تعریف آزمونهای) (سؤالهای مناسب و قابل قبول سؤالهای هستند که سطح دشواری آنها در حد متوسط باشد. سؤال های با سطح دشواری متوسط امکان پراکنده در نمره آزمون شوندگان را افزایش می دهند، سطح دشواری مطلوب برای یک سؤال بدون توجه به نوع سؤال حدود ۵۰ درصد (بین ۳۰٪ تا ۷۰٪) در نظر گرفته می شود.

دامنه درجه دشواری: ۰ ----- ۲۵٪ ----- ۴۵٪ ----- ۶۵٪ ----- ۸۵٪ ----- ۱۰۰٪

فوق العاده دشوار = ۰ تا ۲۵٪

• خیلی دشوار = ۲۵٪ تا ۴۵٪

• دشوار = ۴۵٪ تا ۶۵٪

• متوسط = ۶۵٪ تا ۸۵٪

• ساده = ۸۵٪ تا ۱۰۰٪

۲-۳ قدرت تمیز:

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

تفسیر شاخص: هر چه عدد ضریب تمیز بزرگتر باشد سؤال 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 سؤال می باشد و چنانچه سوال دارای اشکال فنی نباشد. افراد گروه بالا یا آن سؤال را به طور کامل یاد نگرفته اند یا آن را به غلط آموخته اند.

+۱ \_\_\_\_\_ -۱

- خوب: بزرگتر از ۳۰٪
- متوسط: بین ۱۰٪ تا ۳۰٪
- ضعیف: بین ۰ تا ۱۰٪
- منفی: کم تر از صفر

بین سطح دشواری و قدرت تمیز یک سؤال یک رابطه علمی وجود دارد. انتخاب با حذف یک سؤال باید با توجه به این رابطه و بین سه شاخص زیر صورت گیرد.

۲-۲-۱: اگر شاخص دشواری حدود صفر باشد، سؤال فاقد قدرت تفکیک است. در چنین حالتی با مطلب بطور کامل تدریس نشده است، یا متن سؤال مبهم و نارسا می باشد. اگر مطلب تدریس شده است، احتمالاً مشکل در ارائه متن سؤال یا گزینه های پیشنهادی است. لذا باید متن سؤال و الگوی پاسخهای داده شده بررسی شود. اگر نسبت زیادی در 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.

۲-۲-۲: اگر شاخص قدرت تمیز منفی باشد، علت را باید در گزینه های پاسخ سؤال جستجو کرد. در هر صورت سؤالهای با قدرت تمیز منفی سؤال فاقد ارزش هستند باید حذف شوند.

۲-۲-۳: اگر شاخص قدرت تفکیک مثبت پایین باشد، به سطح دشواری سؤال توجه کنید. سؤالهای خیلی ساده یا خیلی دشوار از قدرت تفکیک کمی برخوردار می باشند. اگر سطح دشواری سؤال مطلوب باشد، به الگوی پاسخها و متن سؤال توجه کنید.

### ۳-۲: تحلیل گزینه های انحرافی:

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 به صحیح منحرف کنند.

تفسیر شاخص: در گزینه های انحرافی قاعده 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.

در سوال استاندارد افراد با دانش کافی گزینه صحیح را انتخاب می نمایند و افراد فاقد اطلاعات مناسب پاسخ های نادرست (گزینه های انحرافی) را به طور مساوی انتخاب می کنند.

### ۴-۲ تاکسونومی سؤال:

تعریف شاخص: تاکسونومی، طبقه بندی هدفهای آموزشی است که از طریق سؤالات مطرح شده در آزمون سنجیده می شوند.

تفسیر شاخص: تاکسونومی دارای سه حیطه دانش، درک و فهم و کاربرد سؤال می باشد.

مفهوم دانش در تاکسونومی سؤال: 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: 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: توانایی استفاده از امور انتزاعی، قواعد و قوانین، اصول، اندیشه‌ها و روشها در موقعیتهای عینی و عملی.

۲-۵ ضرورت سؤال:

ضرورت سؤال نیز دارای سه حیطه به شرح زیر 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.
- 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.
- 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 پایین تری نسبت به دو حیطه دیگر قرار دارد.

#### ۲-۶ همبستگی سؤال با کل آزمون:

تعریف شاخص: ضریب همبستگی سؤال با کل آزمون در واقع ضریب توافق با همخوانی سؤال با کل آزمون است که از طریق ضریب همبستگی در رشته‌های نقطه‌ای محاسبه می‌شود.

تفسیر شاخص: هر چه این ضریب همبستگی بیشتر باشد بدان معنی است که سؤال با مجموعه سؤالی دیگر برای اندازه‌گیری صفت با توانایی مورد نظر همخوانی بیشتری دارد. اگر این ضریب همبستگی منفی باشد بدان معنی است که سؤال با کل آزمون ناهمخوانی دارد یعنی چیزی را می‌سنجد که به آنچه باقی سؤالی آزمون می‌سنجد متفاوت است. در واقع این شاخص همبستگی بین عملکرد آزمون شونده‌گان در هر سؤال با عملکرد آنان در کل آزمون را می‌سنجد به عبارتی یکی دیگر از روشهای بررسی ضریب تشخیص با تمایز بین افراد قوی و ضعیف ضریب همبستگی سؤال با کل آزمون است برای اینکه مطمئن شویم که همبستگی بدست آمده بین سؤال با کل آزمون ناشی از شانس و عوامل تصادفی نیست می‌توان این ضریب را از لحاظ معنی داری آماری آزمون کرد. در صورتی که سؤال به لحاظ آماری معنادار نباشد، سوال مورد تحلیل با سایر سؤال‌های آزمون همسانی درونی ندارد و سوال مناسبی نیست.

#### ۲-۷: ضریب پایایی سؤال:

تعریف شاخص: پایایی یک ابزار اندازه‌گیری عبارت است از درجه ثبات، همسانی و قابلیت پیش بینی آن در اندازه‌گیری هر آنچه اندازه می‌گیرد. تفسیر شاخص: یک آزمون در صورتی پایایی دارد که اگر آن را در یک فاصله زمانی کوتاه چندین بار به گروه واحدی از افراد بدهیم نتایج حاصل نزدیک به هم خواهد بود. اگر آزمونی در هر بار اجرا بر روی تعدادی آزمودنی نتایج مختلفی به دست بدهد آن آزمون یک آزمون پایا نخواهد بود و در واقع هیچ چیزی را به درستی اندازه نخواهد گرفت. در مورد پایایی سؤال نیز همین تفسیر صدق می‌کند به این معنی که اگر نمرات بدست آمده از سؤال در موارد مختلف نتایج مختلفی به همراه داشته باشد آن سؤال پایایی نخواهد بود حداکثر ضریب پایایی سؤال ۰/۵ است.

رابطه بین شاخص های تحلیل آزمون:

- هر چقدر ضریب دشواری به  $0/5$  و ضریب تمیز به  $1$  نزدیکتر باشد آن سؤال ، سؤال مناسب تری است .
- هر چه سوال دارای واریانس بزرگتری باشد و با سایر سوالات همبستگی بیشتری داشته باشد به واریانس کل آزمون کمک بیشتری می کند. اگر در تحلیل آزمون به انتخاب سؤال های پیردازیم 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  $0/5$  باشد واریانس سوال به حداکثر خود یعنی  $0/25$  خواهد رسید. در این صورت حداکثر اطلاعات در باره تفاوت بین آزمودنیها به دست خواهد آمد.
- با انتخاب سئوالاتی با ضریب تمیز بالا، ضریب دشواری متوسط ، ضریب پایایی آزمون افزایش پیدا می کند
- هر چقدر ضریب پایایی آزمون به  $1$  نزدیکتر باشد از میزان واریانس خطا کاسته و نشان دهنده این است که واریانس مشاهده شده واریانس حقیقی است.