



# دانشکده علوم پزشکی و خدمات بهداشتی درمانی ساوه

شیوه نامه تحلیل آزمون

## تحلیل سوالات آزمون ها

هدف از تحلیل سئوالهای آزمون و بررسی تک تک سئوالها و تعیین میزان دقت و نارسای های آنهاست. در تحلیل سئوال های آزمون، نقاط قوت و ضعف آزمون و کیفیت سئوال های آن تعیین می شود. لازم است پس از اجرای هر آزمون سئوال های آن را تحلیل کنند و با استفاده از نتایج حاصل به تجدید نظر در آزمون و بهبود کیفیت سئوال ها برای استفاده های بعدی اقدام نمایند. برای تهیه یک آزمون خوب تحلیل پاسخ های آزمون شوندگان و بررسی تجربی آنها ضروری است. اطلاعات مورد نیاز برای تحلیل سئوال های یک آزمون از پاسخهایی که آزمون شوندگان به هر سئوال داده اند بدست می آید مانند در هر سئوال چند نفر گزینه درست را انتخاب کرده اند، هریک از گزینه های انحرافی چند نفر را به خود جلب کرده است و چند نفر آن را بی جواب گذاشته اند. قبل از هرکاری، باید نوع آزمون را مشخص نمود تا بتوان تحلیل مربوط به آن را انجام داد.

۱. **آزمون های هنجاری:** نتیجه آرمانی یک آزمون هنجاری، تمیز دانشجویان قوی و ضعیف است.

۲. **آزمون های ملاکی:** معمولاً در آزمون ملاکی انتظار می رود همه شرکت کنندگان قبول شوند.

## روشهای کیفی تحلیل آزمون :

۱. **روایی محتوا<sup>۱</sup>**

**الف) 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

<sup>1</sup> Content Validity

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 هر یک از تاکسونومی فوق قرار دارند به تفکیک سطح گزارش شود. برای هر یک از سوالات لازم است نسبت به ایجاد ستونی در جلوی هر سوال و تفکیک سطح سوال در حیطه شناختی اقدام شود.

### ب) 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 :

I. **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 میدهد).

II. **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).

III. **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).

عنوان	تعداد	درصد
تعداد سوالات		
سوالات با تاکسونومی ۱		
سوالات با تاکسونومی ۲ و ۳		
سوالات بدون هیچ اشکال ساختاری		
سوالات با ۱ اشکال ساختاری		
سوالات با ۲ اشکال ساختاری		
سوالات با ۳ اشکال ساختاری		
سوالات با ۴ اشکال ساختاری		
سوالات با بیش از ۴ اشکال ساختاری		

## ۲. روایی ساختاری آزمون<sup>۲</sup>:

### رعایت قواعد ساختاری سوال یا اعتبار ساختار با استفاده از پرسشنامه میلمن

این چک لیست، قواعد تهیه سوال‌های چندگزینه‌ای را نمره دهی می‌کند. قواعدی که برای تهیه سوال‌های چندگزینه‌ای پیشنهاد می‌شوند براساس این طرز فکر تدوین شده‌اند که یک سوال چندگزینه‌ای باید آزمون‌شونده را با تکلیفی که هم مهم و کاملاً برای او قابل فهم است مواجه سازد. این چک لیست شامل مواردیست که بای در ساختار یک سوال چندگزینه‌ای (MCQ) رعایت گردد که در دو قسمت ساقه سوال (stem) و گزینه‌ها (options) ارائه می‌گردد (به ترتیب گزینه‌های چک لیست). نحوه نمره‌دهی از ۰ تا ۲ (خیر، تا حدودی و کاملاً) می‌باشد.

پرسشنامه میل من (فرم کوتاه شده)			
ردیف	موضوع	کاملاً	تا حدودی
۱	آیا بخش اعظم اطلاعات در تنه سؤال گنجانده شده است؟		
۲	آیا سؤال یک هدف اختصاصی یادگیری را مورد ارزیابی قرار می‌دهد؟		
۳	آیا لغات استفاده شده در تنه یا گزینه‌ها، شفاف و مستقیم بیان شده‌اند؟		
۴	آیا از کاربرد گزینه منفی برای تنه منفی خودداری شده است؟		
۵	آیا از کاربرد گزینه‌هایی نظیر "همه موارد" یا "هیچکدام" و گزینه‌های ترکیبی خودداری شده است؟		
۶	آیا از کاربرد گزینه‌های متضاد یکدیگر، خودداری شده است؟		
۷	آیا از لغات مثبت در تنه سوال استفاده شده است یا در صورت منفی بودن تنه سوال، لغات منفی مشخص شده‌اند؟		
۸	آیا این سوال مستقل از سئوالات دیگر می‌باشد؟		
۹	آیا گزینه‌ها از نظر طول، ساختار لغوی و سبک نگارش همسنگ هستند؟		
۱۰	آیا تا حد امکان از کاربرد عبارتهای تکراری در گزینه‌ها خودداری شده است؟		
۱۱	آیا کلمات بکار رفته در تنه و یا گزینه‌ها از نظر املائی صحیح است؟		
۱۲	آیا بطور مشخص یک گزینه صحیح می‌باشد (خودداری از طرح سئوالات با پاسخهای متناقض)؟		
۱۳	آیا گزینه‌ها به طور عمودی لیست شده است؟		
۱۴	آیا گزینه‌های انحرافی، تنها دانشجویان غیر مطلع را گمراه می‌کند؟		
۱۵	آیا تنه سوال و گزینه‌ها مثبت‌اند؟		

<sup>2</sup> construct validity

## مراحل تحلیل سوالات آزمون

محور تحلیل پرسش ها، پاسخ های آزمون شوندگان است. برای این منظور هر سوال به طور جداگانه به همراه جواب در یک کارت یا برگه نوشته، نتایج تحلیل را در یک سمت برگه و صورت اصلاح شده ی سوال را در سمت دیگر آن قرار می گیرد. برای تحلیل هر پرسش، دو گروه جدا می کنیم، گروه بالا (قوی) و گروه پایین (ضعیف). قبل از جدا کردن افراد بهتر است، برگه ها به ترتیب نمره دسته بندی شود. انتخاب دو گروه بالا و پایین به صورت زیر انجام می شود:

- ۲۰ نفر یا کمتر تمام برگه ها به دودسته ی قوی و ضعیف تقسیم می شود.
- بیشتر از ۲۰ تا ۴۰ نفر ۱۰ برگه ی بالا و ۱۰ برگه ی پایین انتخاب می شود.
- چنانچه تعداد بیشتر از ۴۰ باشد، معمولاً در آزمونهای ملاکی، ۲۵ تا ۳۳ درصد کل برگه ها را برای گروه بالا و همان اندازه را برای گروه پایین و در آزمون های هنجاری ۲۷ درصد در نظر می گیرند.

بعد از تعیین افراد دو گروه باید مشخص شود که چه تعداد افراد از هر گروه به گزینه های مختلف جلب شده و آن گزینه ها را انتخاب نموده اند، یعنی چگونگی توزیع پاسخ افراد به گزینه های مختلف، مشخص شود و نیز تعداد افراد از هر گروه که به گزینه های انحرافی و گزینه ی کلید، جلب شده اند.

## محاسبه ی ضریب دشواری سوال

بنا به تعریف، درصد کل آزمون شوندگانی که به یک سوال جواب درست می دهند، ضریب دشواری آن سوال می گویند و به طور کلی اگر  $R$  پاسخ های درست و  $T$  کل پاسخ ها باشند مطابق فرمول زیر محاسبه می شود:

$$P = \frac{R}{T} * 100$$

چون از آزمون شوندگان نمونه گیری شده و دو گروه نمونه در اختیار داریم، فرمول به این صورت تغییر می کند: در آزمون های هنجاری:

$$P = \frac{(\text{انتخاب درست گروه بالا} + \text{انتخاب درست گروه پایین})}{\text{تعداد افراد گروه بالا} + \text{تعداد افراد گروه پایین}} * 100$$

در آزمون های ملاکی:

$$P = \frac{(\text{تعداد دانشجویان مردود با پاسخ درست} + \text{تعداد دانشجویان قبول با پاسخ درست})}{\text{تعداد دانشجویان دو گروه}} * 100$$

شاخص دشواری اصلاح شده:

$$P_c = \frac{1 - \frac{\text{تعداد جواب های نادرست}}{\text{تعداد جواب های صحیح} + \text{تعداد جواب های نادرست}}}{\text{تعداد جواب های صحیح}}$$

$$P_h = \frac{\text{فراوانترین پاسخ غلط} - \text{پاسخ های صحیح به سوال}}{\text{مجموع پاسخ های غلط هر سوال} + \text{پاسخ های صحیح به سوال}}$$

این شاخص از دقت بالاتری برخوردار است.

### تفسیر ضریب دشواری

هرچه عدد حاصله به ۱۰۰ نزدیک تر باشد، ضریب دشواری آن سوال کمتر و سوال آسان تر است، زیرا تعداد بیشتری از افراد دوگروه به آن پاسخ داده اند. از این ضریب به ضریب آسانی یا سهولت سوال هم تعبیر می کنند، ولی بیشتر به ضریب دشواری معروف است. از لحاظ انتخاب برای گنجاندن در فرم نهایی آزمون، سوالهایی بهتر هستند که ضریب دشواری آنها از ۱ کمتر و از صفر بیشتر و به ۰/۵ نزدیک باشد.

از طرفی یکی از راههای قضاوت درباره مفید بودن سئوالهای یک آزمون آن است که بینیم سئوال تا چه اندازه به پراکندگی یا واریانس نمرات کمک می کند. واریانس نمرات گروهی پاسخ دهنده به یک آزمون از دو قسمت تشکیل می شود:

- واریانس سئوال ها

- همبستگیهای بین سئوال ها

در یک آزمون یا وضعیت هنجاری، هر سئوالی که دارای واریانس بزرگتری است و با سایر سئوالهای آزمون همبستگی بیشتری دارد به واریانس کل آزمون کمک بیشتری می کند. در مقابل سئوالی که با سایر سئوال ها همبستگی ندارد و دارای واریانس کوچکی است کمک زیادی به واریانس کل آزمون نمی کند. اگر در تحلیل آزمون به انتخاب سئوالهایی پردازیم که همبستگی زیادی با یکدیگر دارند و در ضمن دارای واریانس های بزرگتری هستند در مجموع آزمون بهتری تدارک دیده ایم زیرا سوال و افراد نمونه، عملکرد مناسب تری داشته اند. برای محاسبه ی واریانس یک سوال از فرمول زیر استفاده می شود

$$Var = P(1 - P)$$

P درصد پاسخ های درست یعنی همان ضریب دشواری است.

زمانی واریانس بیشترین مقدار را دارد که ۵۰ درصد به سوال پاسخ درست داده باشند و زمانی کمترین مقدار را، یعنی صفر که همه یا هیچ یک به آن پرسش، پاسخ درست نداده باشند.

با توجه به تحلیل یاد شده، سوالاتی انتخاب و در آینده از آن ها استفاده می شود که، واریانس آنها بیشتر باشد و ضریب دشواری آنها از یک کمتر و از صفر بیشتر است و به (۰/۵) نزدیک تر باشد. به عبارتی دیگر، حداکثر واریانس را دارا باشد. البته انتخاب سئوال هایی با ضرایب دشواری مناسب به نوع سئوالهای آزمون نیز مربوط است.

$$\text{حد مطلوب سطح دشواری هر سوال} = \frac{\text{احتمال پاسخ تصادفی} + 1}{2}$$

بطور مثال برای آزمونهای چهارگزینه ای سطح بهینه دشواری در حدود ۰.۶ یعنی اندکی کمتر از وسط فاصله بین ۱ سطح موفقیت با حدس زدن یعنی ۰.۲۵ باشد. به طور کلی ضریب دشواری بین ۰/۳ تا ۰/۷، حداکثر اطلاع را در مورد تفاوت میان آزمودنی ها به دست می دهد و ضریب مناسبی خواهد بود اما بسته به اهداف گوناگون، نیز می توان ضریب دشواری بالا یا پایین را منظور کرد. مثلا سوالهایی برای آزمون های هنجاری بهتر هستند که ضریب دشواری متوسط داشته باشند و یا در امتحانی که قرار است بالای ۱۰ درصد قبول شوند، باید سوالات مشکل با ضریب دشواری کمتر ( عدد کمتر) داشته باشیم.

زیر ۰.۳	بالای ۰.۷	۱ یا نزدیک آن
سختی بسیار زیاد سوال • بازنگری در کلمات • خارج کردن سوال از آزمون بعدی • تدریس مجدد هدف آموزشی • ...	سادگی بیش از حد • نباید در آزمون بعدی مورد استفاده قرار بگیرد	سوال کاملا آسان

### ضریب تمییز سوال

این شاخص، بر خلاف ضریب دشواری، توان سوال را در ایجاد تمییز و تشخیص بین گروه قوی و ضعیف آزمون شوندگان، نشان می دهد؛ اینکه یک سوال تا چه اندازه در جدا کردن دو گروه قوی و ضعیف، موفق بوده است. برای محاسبه ی ضریب تمییز از فرمول زیر استفاده می کنیم:

$$D = \frac{\text{انتخاب های درست گروه پایین} - \text{انتخاب های درست گروه بالا}}{\text{تعداد افراد گروه پایین یا بالا}} * 100$$

### محاسبه ضریب تمییز با استفاده از ضریب همبستگی نمره سوال با نمره کل

در آزمون های هنجاری همبستگی سوال با نمره کل نشان دهنده قدرت آن در تمایز بین دانشجویان ضعیف و قوی است. ضریب همبستگی در آزمون های چندگزینه ای، نمره هر سوال به صورت صفر یا یک، یعنی یک متغیر دو حالته است در حالی که نمره کل، یک متغیر پیوسته مثال ۱ تا ۲۰ است. روشی که برای محاسبه همبستگی بین متغیر دو حالتی و متغیر پیوسته بکار می رود همبستگی دو رشته ای - نقطه ای<sup>۳</sup> است که در واقع ضریب توافق سؤال با کل آزمون است. هر چه این ضریب بیشتر باشد، بدان معنی است که سؤال با مجموعه سؤال های دیگر همخوانی بیشتری دارد؛ و لذا بین گروهها بیشتر تمایز قائل می شود پس هدف نهایی آزمون را بهتر بر آورده می سازد. ضریب همبستگی بیشتر یعنی سوالی که دانشجویان سطح بالا می توانند جواب صحیح بدهند و نمره بگیرند اما دانشجویان ضعیف نمی توانند و نمره ای از آن نمی گیرند و بدینوسیله تفکیک خوب از بد حاصل می شود. ضریب تمییز از فرمول زیر محاسبه می شود:

$$M_T = \text{میانگین نمره کل همه دانشجویان}$$

<sup>3</sup> CORRELATION BISERIAL- POINT

$M_P$  = میانگین نمره کل دانشجویانی که به سوال پاسخ درست دادند.

$S_T$  = انحراف معیار نمره کل

$P$  = نسبت تعداد دانشجویانی که پاسخ درست دادند به کل آزمون شوندگان

$$r_{pbis} = \frac{M_P - M_T}{S_T} * \sqrt{\frac{P}{1 - P}}$$

بهترین شرایط برای سوالات یک آزمون این است که همه سوالات یک آزمون با یکدیگر همبستگی کامل داشته باشند. محاسبه همبستگی تک تک سوالات با یکدیگر چند فایده مهم دارد:

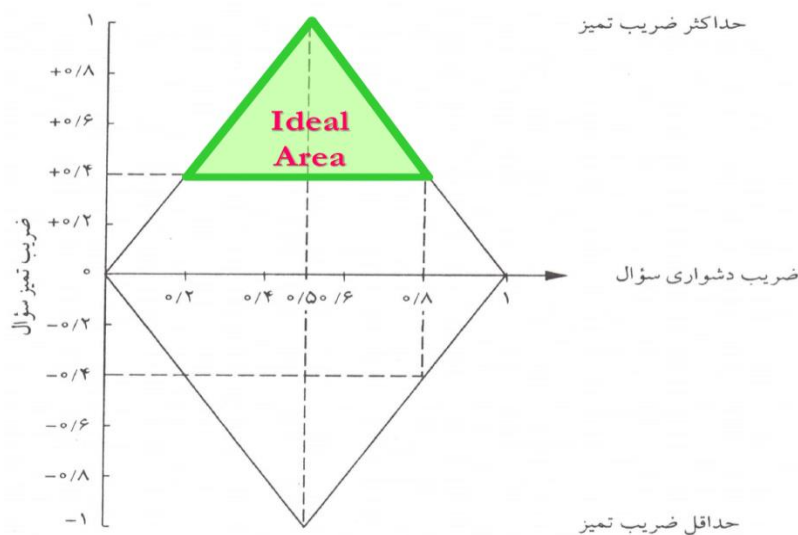
۱. با در دست داشتن تعداد سؤالهای آزمون و میانگین ضریب همبستگی بین آنها می توان ضریب پایایی آزمون را محاسبه کرد.

۲. با استفاده از این نوع همبستگی می توان به مفهوم قوه تمیز سؤال دست یافت.

### تفسیر ضریب تمیز

هرچه عدد ضریب تمیز بزرگتر باشد، سوال از توان بیشتری در جدا کردن و تفاوت قائل شدن میان گروه قوی و ضعیف، برخوردار بوده است و بر عکس. گاهی اوقات ضریب تمیز سوالی منفی است، یعنی گروه ضعیف به سوال پاسخ داده ولی گروه قوی قادر به پاسخ گویی نبوده است یعنی گروه قوی از گروه ضعیف، عملکرد پایین تری داشته است. این مشکل به طور معمول، به ایراد خود سوال مربوط بوده و باید اصلاح شود و چنانچه سوال دارای اشکال فنی نباشد، افراد گروه بالا یا آن سوال را به طور کامل یاد نگرفته اند یا آن را به غلط آموخته اند یا بدلیل نبودن گزینه درست در کلید یا تقلب در آزمون است. بنابراین تمامی ضرایب تمیز آزمون باید مثبت و بالا ۰/۳ قابل قبول است. به عنوان یک قاعده کلی باید بدانیم که سوالات بسیار مشکل و سوالات بسیار ساده هر دو دارای قدرت تمیز ضعیف هستند.

رابطه ضریب تمیز و ضریب دشواری



## گزینه های انحرافی

هدف از قرار دادن گزینه انحرافی در سؤال ها منحرف کردن آزمون شوندگانی است که جواب درست سؤال را نمی دانند. قاعده آن کلی به شرح زیر است: هر گزینه ی انحرافی باید دست کم یک نفر از افراد گروه ضعیف را به جلب کند و اگر هر دو گروه قوی و ضعیف را به جلب کند، تعداد افراد گروه ضعیف بیشتر از گروه قوی باشد عبارتی برای افرادی که توانایی و مهارت کافی دارند، گول زنده نباشند.

## تحلیل گزینه های انحرافی

حالت اول) در شرایط عادی در صورتی یک سوال به خوبی عمل می کند که افراد گروه ضعیف بیشتر از افراد گروه قوی گزینه های انحرافی آن سوال را انتخاب نمایند (ضریب دشواری متوسط و ضریب تمیز بالا).

حالت دوم) در حالت ایده آل برای یک سوال چهار گزینه ای مطلوب انتظار می رود که هر سه گزینه انحرافی به میزان نسبتاً یکسانی جلب توجه نموده و از سوی آزمون شوندگان بی اطلاع به یک میزان مورد انتخاب قرار گرفته باشند.

حالت سوم) اگر همه آزمون شوندگان (قوی و ضعیف) فقط یک گزینه مشخص (و مشابه) را انتخاب کرده اند که گزینه صحیح است و هیچیک از گزینه های انحرافی انتخاب نشده اند، موارد زیر مطرح می گردد:

۱. سوال فوق العاده آسان است با ضریب دشواری = ۱ و ضریب تمیز = ۰

۲. درست بودن گزینه پاسخ بطور بسیار آشکاری خودنمایی می کند.

۳. هیچیک از گزینه های انحرافی حتی برای آزمون شوندگان بی اطلاع از موضوع، مناسب و جذاب نیست.

۴. تقلب کرده اند.

در حقیقت این سوال یک گزینه ای است نه چند گزینه ای.

حالت چهارم) اگر همه آزمون شوندگان (قوی و ضعیف) فقط یک گزینه مشخص (و مشابه) را انتخاب کرده اند که گزینه ای نادرست است و هیچیک از آزمون شوندگان گزینه صحیح یا گزینه های انحرافی دیگر را انتخاب نکرده است (ضریب دشواری = ۰ و ضریب تمیز = ۰)، موارد زیر مطرح می گردد:

۱. تنه سوال غلط یا کاملاً مبهم است.

۲. گزینه ی انحرافی انتخاب شده برای همه آزمون شوندگان جذابیت کاذب و کاملی داشته است در حالیکه سایر گزینه های

انحرافی که انتخاب نشده اند و همچنین گزینه کلید هیچ جذابیتی نداشته اند.

۳. تقلب کرده اند.

در حقیقت این سوال نیز یک گزینه ای است نه چند گزینه ای.

حالت پنجم) اگر در مجموع انتخاب های همه آزمون شوندگان یک یا دو گزینه مشخص از چهار گزینه ارائه شده، اصلاً انتخاب نشده اند، یعنی آن یک یا دو گزینه ی مشخص توسط همه آزمون شوندگان کنار گذاشته شده اند:



۱. گزینه های انتخاب نشده خیلی دور از موضوع سوال بوده اند بطوریکه حتی برای آزمون شوندگان بی اطلاع از موضوع هم هیچ جذابیتی ایجاد نکرده اند.
۲. گزینه های انتخاب شده بطور بسیار آشکاری به پاسخ سوال نزدیک بوده اند بطوریکه توجه همه آزمون شوندگان را به خود جلب کرده اند.

بسته به تعداد گزینه های انتخاب نشده، این سوال سه یا دو گزینه ای است نه چهارگزینه ای

### جمع بندی :

پس از تعیین ضریب های دشواری و تمیز تمامی سوالات و بررسی گزینه های هر سوال، باید بر اساس نتایج به اصلاح یا حذف آنها اقدام کرد. سوالاتی که ضریب دشواری بالا یا بسیار پایین و ضریب تمیز آنها خیلی کم است باید مورد تجدید نظر قرار گیرند. همچنین گزینه های انحرافی معیوب نیز باید اصلاح یا عوض شوند. پس از این تغییرات سوالات در بانک سوال ذخیره گردد.

✓ تعیین عامل حدس و گمان:

$$1 - \frac{\text{تعداد پاسخ های غلط}}{\text{تعداد جواب درست}}$$

### تحلیل سوالات آزمون های تشریحی و عملکردی :

در تحلیل سوالات تشریحی و عملکردی روش های مختلفی وجود دارد. یکی از روش ها آن است که 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 :

$$P = \frac{\text{نمره میانگین سوال}}{\text{پایین ترین نمره ممکن - بالاترین نمره ممکن}} = \frac{\text{نمره میانگین سوال}}{\text{دامنه ی ممکن نمرات سوال}}$$

ضریب تمیز براساس این روش بدینگونه محاسبه می شود :

۱- میانگین نمرات گروه بالا و پایین را برای هر سوال تعیین می نمایم .

۲- تفاوت میان دو میانگین را محاسبه می کنیم .

۳- حاصل را به دامنه ی نمرات ممکن برای آن سوال تقسیم می نمایم.

$$D = \frac{\text{میانگین سوال گروه پایین} - \text{میانگین سوال گروه بالا}}{\text{پایینترین نمره ممکن سوال} - \text{بالاترین نمره ممکن سوال}} = \frac{\text{تفاوت میانگین دو گروه در هر سوال}}{\text{دامنه ی نمرات ممکن سوال}}$$

تحلیل سوالات آزمون های وابسته به ملاک ( ملاک مرجع )

اصولاً در آزمون های وابسته به ملاک مانند آزمون های پیشرفت تحصیلی، سوالات آزمون باید به گونه ای باشد که اکثر دانشجویان در حدود ۸۵ تا ۹۰ درصد، در صورت تسلط بر درس و آگاهی، بتوانند به تمام پرسش ها جواب درست بدهند. در واقع سوالات این نوع آزمون ها باید ضریب دشواری با مقدار زیاد و ضریب تمیز با عدد کم داشته باشند. یعنی برای دانشجویان اکثر سوالات آسان و قدرت تمیز آن ها نیز بسیار کم باشد.

با این حال متخصصان روان سنجی عقیده دارند، که در این موارد نیز محاسبه ی ضریب تمیز و دشواری سوالات می تواند مفید باشد و با استفاده از نمرات حاصل به اصلاح پرسش های آزمون و گزینه های انحرافی و کلید اقدام نمود. در این نوع آزمون ها معمولاً تفاوت نمرات و عملکرد دانشجویان پیش و پس از آموزش ملاک عمل قرار می گیرد.

### اعتبار آزمون<sup>۴</sup>:

اعتبار آزمون بیانگر میزان ثبات و پایایی آزمون در رابطه با نمرات آزمون است. که دامنه آن از صفر تا یک است. مقادیر بالا آن نشان دهنده اعتبار زیاد نمرات آزمون و همبستگی بالای سؤال های خوب با سایر سؤال ها (مفاهیم مشابه را مورد سنجش قرار می دهد) است. روش های متعدد محاسبه پایایی وجود دارد اما با توجه به امکانات بهترین روش برای اندازه گیری پایایی یک آزمون روش های ذیل است:

۱. روش گودریچاردسون: در این روش آزمون یک بار اجرا می شود اما در این روش متوسط ضریب پایایی آزمون بر اساس ضریب همبستگی در نیمه های آن از طریق همه روش های ممکن دو نیمه کردن بدست می آید. عدد بدست آمده پایایی و همسانی درونی محتوای آزمون را نشان می دهد.

۲. روش دو نیمه کردن آزمون (اسپلیت هاف): آزمون یک بار با گروه واحدی از آزمون شوندگان اجرا می شود و سپس آن را به دو نیمه تقسیم می کنیم. بهترین روش دو نیمه کردن آزمون بصورت جدا نمودن سوالات فرد و زوج است. برای محاسبه ضریب پایایی در این روش باید همبستگی بین نیمه های آزمون رابدهست آوریم.

۳. روش ضریب آلفا: این روش زمانی از روش گودریچاردسون متفاوت خواهد بود که آزمون از تعدادی خرده آزمون تشکیل شده باشد.

میدانیم اعتبار یک آزمون با افزایش تعداد سؤال ها بالا می رود و باید از سؤال هایی با قدرت تشخیص بالا نیز استفاده کنیم.

تفسیر	Reliability
اعتبار عالی (آزمون های استاندارد)	۰.۹ به بالا
بسیار خوب برای آزمون های پیشرفت تحصیلی	۰.۸-۰.۹
مناسب و خوب برای آزمون های پیشرفت تحصیلی (تعدادی از سؤال ها نیاز به بهبود دارند)	۰.۷-۰.۸
تقریباً پایین (باید بعضی سؤال ها حذف شوند)	۰.۶-۰.۷
پیشنهاد بازنگری کلی آزمون (احتمالاً سؤال های آزمون کم بوده است)	۰.۵-۰.۶
اعتبار آزمون زیر سؤال است (این آزمون نباید مورد استفاده قرار گیرد و نیاز به بازنگری کلی دارد)	زیر ۰.۵

<sup>4</sup> Reliability

منابع:

سیف علی اکبر. اندازه گیری، سنجش و ارزشیابی آموزشی. نشر دوران. ویرایش هفتم، سال ۹۶  
جلیلی محمد و دیگران. اصول و روش های ارزیابی فراگیران در علوم پزشکی. انتشارات فرهنگستان علوم پزشکی جمهوری اسلامی  
ایران - سال ۹۶

Amin Z, Seng C.Y, Eng K.H. Practical guide to medical student assessment. World Scientific Publishing. Last Edition

John A, Harden R.M. A Practical guide for medical teacher. Elsevier. Last Edition

Marilyn H. Oermann, Kathleen B.Gaberson .Evaluation and Testing in nursing

Education-Third Edition. 95-117