

## بسمه تعالی



دانشگاه علوم پزشکی خدمات بهداشتی درمانی خراسان شمالی

### مرکز مطالعات و توسعه آموزش علوم پزشکی دانشگاه علوم پزشکی خراسان شمالی

### دستورالعمل استاندارد سنجی سوالات و آنالیز آزمون ها

#### مقدمه:

تحلیل دقیق ارزیابی آزمون، نه تنها به عنوان مهمترین بخش روند آموزشی و 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 تاثیر گذار را ارائه می نماید. لذا تحلیل دقیق آزمون بی تردید در زمینه ارتقاء کیفیت سیستم آموزشی بسیار مؤثر خواهد بود.

#### تحلیل سؤال:

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

## شاخص های تحلیل و ارزیابی

### شاخص دشواری – شاخص تمیز – روایی آزمون – پایایی آزمون

\* در انتهای هر نیمسال تحصیلی لازم است شاخص های بالا جهت هر آزمون سنجیده شوند و نتیجه نهایی از هر دانشکده به مرکز توسعه و مطالعات دانشگاه (EDC) گزارش گردد.

۱- شاخص دشواری: این شاخص نشان می دهد که سوالات با ضریب دشواری بین ۳۰-۷۰ درصد قابل قبول هستند و خارج از این بازه برای ترم های بعد مورد استفاده قرار نگیرد و قابلیت تکرار ندارد.

شاخص دشواری و نحوه محاسبه آن: 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. هرچه 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، سؤالهای مناسب و قابل قبول سؤالهای هستند که سطح دشواری آنها در حد متوسط باشد. سؤال های با سطح دشواری متوسط امکان پراکندگی 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:

+ ----- ۲۵٪ ----- ۴۵٪ ----- ۶۵٪ ----- ۸۵٪ ----- ۱۰۰٪

فوق العاده دشوار = + تا ۲۵٪

خیلی دشوار = ۲۵٪ تا ۴۵٪

دشوار = ۴۵٪ تا ۶۵٪

متوسط = ۶۵٪ تا ۸۵٪

ساده = ۸۵٪ تا ۱۰۰٪

محاسبه ضریب دشواری:  $P = 100 \times \frac{\text{انتخابهای درست گروه پایین} + \text{انتخابهای درست گروه بالا}}{\text{تعداد افراد گروه بالا} + \text{تعداد افراد گروه پایین}}$  = ضریب دشواری سؤال

تعداد افراد گروه بالا + تعداد افراد گروه پایین

۲- قدرت تمیز ( افتراق ) : این شاخص نشان می دهد که سوالات با قدرت تمیز بالای ۳۰ درصد قابل قبول هستند و زیر ۳۰ درصد فاقد ارزش می باشد.

قدرت تمیز و نحوه محاسبه آن: قدرت تمیز یا 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.

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 یا حذف یک سؤال باید با توجه به این رابطه صورت گیرد.

\* اگر شاخص دشواری حدود صفر باشد، سؤال فاقد قدرت تفکیک است. در چنین حالتی یا مطلب بطور کامل تدریس نشده است، یا متن سؤال مبهم و ناروا 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.

\* 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 و باید حذف شوند.

\* 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.

محاسبه ضریب تمیز سوالات:  $d = \frac{\text{انتخابهای درست گروه پایین} - \text{انتخابهای درست گروه بالا}}{\text{ضریب تمیز سوال}}$

تعداد افراد یک گروه (بالا یا پایین)

۳- روایی آزمون (روایی محتوایی - روایی صوری): آزمونی دارای روایی است که برای اندازه گیری آنچه مورد نظر است، کافی و مناسب باشد. در این بخش باید روایی صوری و روایی محتوایی برای آزمونهای چهارگزینه ای و تشریحی سنجیده شود. بهتر است برای سنجش روایی صوری آزمونهای چهارگزینه ای از چک لیست میلمن استفاده گردد و در بخش روایی محتوایی پوشش تمام مباحث سرفصل درس، مطابق با اهداف آموزشی ذکر شده در کوریکولوم رشته با تایید کمیته آزمون در EDO دانشکده و خوداظهاری استاد مورد پذیرش می باشد.

روایی محتوایی: به این مطلب اشاره می کند که نمونه سؤال های موجود در یک آزمون تا چه حد معرف کل جامعه سؤال های ممکن است که می توان از محتوا یا موضوع مورد نظر تهیه کرد. هرچه آزمون از این لحاظ بهتر باشد، دارای روایی بیشتری است. برای تعیین روایی محتوایی یک آزمون از قضاوت متخصصان (مدیر گروه آموزشی یا نماینده وی) در این باره که سؤال های آزمون تا چه میزان معرف محتوا و هدف های برنامه یا حوزه محتوایی (content domain) هستند، استفاده می شود.

روایی صوری: به این مطلب اشاره می کند که سؤال های یک آزمون تا چه حد در ظاهر شبیه به موضوعی هستند که برای اندازه گیری آن تهیه شده اند. چک لیست میلمن نیز برای آزمونهای چهارگزینه ای در این زمینه بسیار کاربردی می باشد.

شماره سؤال	چک لیست میلمن															ارزیابی کیفی MCQ				
	۱۵	۱۴	۱۳	۱۲	۱۱	۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱					
																	۱	ساقه سؤال	سؤال یک موضوع مهم (یک هدف اختصاصی) را مورد ارزیابی قرار می دهد و مرتبط با محتوی مشخصی از بلو پریت آزمون می باشد.	
																	۲		مخور مناسب برای سؤال انتخاب شده است (بیشتر از یک مسئله یا یک مطلب در سؤال قرار ندارد و یک مورد یا موقعیت بالینی مبنای طرح سؤال می باشد) (تشخیص - درمان - .....)	
																	۳		محتوی سؤال مستقل از محتوی سؤالات دیگر می باشد.	
																	۴		ایده اصلی (بخش اعظم اطلاعات) بطور کامل در ساقه سؤال گنجانده شده است. (پایه سؤال حاوی اطلاعاتی که یک داوطلب توانمند برای پاسخ دهی به سؤال نیاز دارد می باشد)	
																	۵		بدنه اصلی سؤال واضح، بدون ابهام و روان بودن و از نظر لغوی و دستوری پیچیده نمی باشد و قواعد دستوری رعایت شده است.	
																	۶		از بکارگیری عبارات مبهم مثل همه موارد - هیچکدام در ساقه سؤال پرهیز شده است.	
																	۷		از بکارگیری قیود خاص مانند هیچوقت - همیشه - کاملاً - مطلقاً در ساقه سؤال پرهیز شده است.	
																	۸	سؤال هدایت کننده به روشنی مشخص کرده است که چگونه باید به سؤال پاسخ داده شود.		
																	۹	می توان بدون نگاه کردن به گزینه ها به سؤال پاسخ داد.		
																	۱۰	سؤال بجای ارزیابی محفوظات به ارزیابی توان به کارگیری اطلاعات می پردازد.		
																	۱۱	سؤال به صورت جمله مثبت طراحی شده است، در صورت منفی بودن سؤال هدایت کننده، لغات منفی مشخص شده اند.		
																	۱۲	گزینه ها با هم همپوشانی ندارند و مستقل از هم طراحی شده اند.		
																	۱۳	گزینه ها از نظر محتوی و دستوری هموزن طراحی شده اند و به موضوع واحدی مربوط می باشند.		
																	۱۴	طول جمله گزینه ها تقریباً یکسان هستند.		
																	۱۵	از دادن سرخ هایی که نشان می دهند گزینه صحیح کدام گزینه است خودداری شده است. مانند: کاربرد قیود خاص نظیر همیشه، هیچ وقت، کاملاً و مطلقاً گزینه هایی که از نظر لغوی دارای کلمات مشابه پایه سؤال باشند تلاوت دستوری گزینه ها که داوطلب را به سوی گزینه صحیح هدایت می کند. گزینه ای که به طور کاملاً واضح صحیح باشد.		
																			۱۶	از عبارات نظیر همه موارد فوق، همه موارد بجز، هیچکدام از موارد فوق، در گزینه استفاده نشده است
																			۱۷	از عبارات مبهمی (نظیر غالباً، معمولاً، گاهی، هرگز) استفاده نشده است.
																			۱۸	گزینه ها بصورت عمودی مرتب شده اند.
																	۱۹	جای گزینه صحیح نسبت به سؤالات دیگر تغییر کرده است.		
																	۲۰	تا حد امکان از کاربرد عبارت های تکراری در گزینه ها پرهیز شده است.		
																	۲۱	در میان گزینه ها یک گزینه کاملاً صحیح وجود دارد.		
																	۲۲	از کاربرد گزینه های منفی برای ساقه منفی (منفی مضاعف) خودداری شده است.		
																	۲۳	از کاربرد گزینه های متضاد یکدیگر، خودداری شده است.		

۴- پایایی آزمون: جهت سنجش پایایی باید ضریب آلفای کرونباخ آزمون بیشتر از  $0.7$  باشد. در مورد آزمون ها مجازی، داده های موجود در سیستم آزمون های مجازی دانشگاه جهت محاسبه ضریب پایایی به کار گرفته می شوند.

ضریب پایایی سؤال: پایایی یک ابزار اندازه گیری عبارت است از درجه ثبات، همسانی و قابلیت پیش بینی آن در اندازه گیری هر آنچه اندازه می گیرد.

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 می کند به این معنی که اگر نمرات بدست آمده از سؤال در موارد مختلف نتایج مختلفی به همراه داشته باشد آن سؤال، سؤال پایایی نخواهد بود. حداقل ضریب پایایی قابل قبول آزمون  $0.5$  است.

\* با انتخاب سؤالاتی با ضریب تمیز بالا و 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.