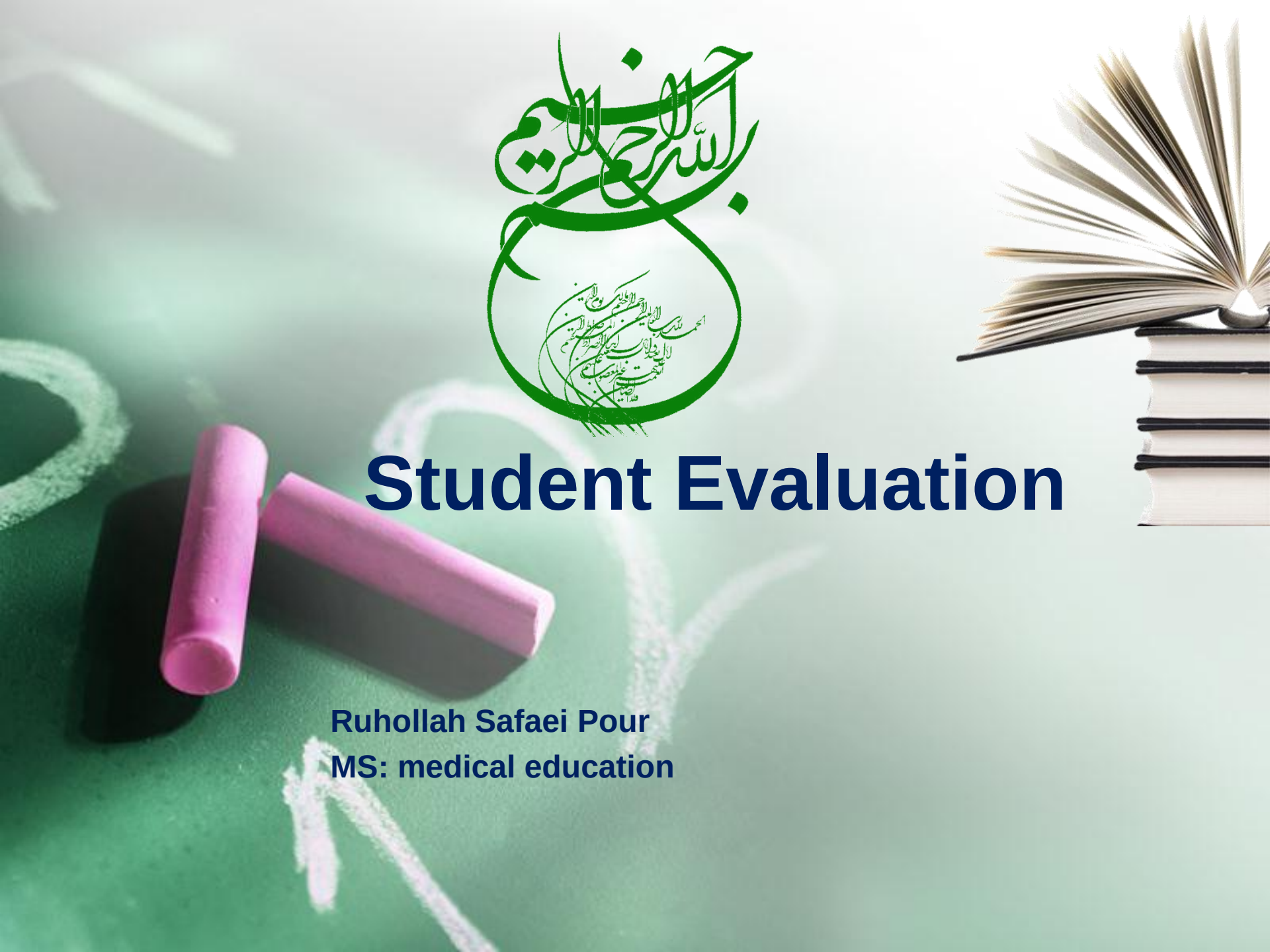


Student Evaluation

Ruhollah Safaei Pour
MS: medical education

مفاهیم پایه

- **اندازه گیری(Measurement):**

اطلاق یک برچسب یا label عددی به یک مفهوم را اندازه گیری می گویند.-مانند اندازه گیری قد- وزن

- **ارزیابی یا سنجش (Assessment):**

زمانی که اندازه گیری برای رسیدن به یک هدف خاص و یا در راستای یک هدف باشد این اندازه گیری ارزیابی نام دارد.

مثال :

- عدد حاصل از اندازه گیری فضای آموزش به نسبت دانشجویان یک اندازه گیری است.
- اما اگر این اندازه گیری جهت رسیدن به فضای مناسب آموزشی باشد ارزیابی صورت گرفته است چون در راستای هدف خاصی است.

ارزشیابی(Evaluation):

- ارزشیابی(Evaluation): فرایند نظاممند جمع آوری اطلاعات به منظور دستیابی به یک قضاوت ارزشی value judgment
- مثال: اگر با هدف خاصی سرانه فضا ی آموزشی را محاسبه می کنید و در واقع Assess می کنید و به عنوان مثال عدد ۲/۱ بدست می آید. تا زمانی که در مورد ۲/۱ متر به ازای هر دانشجو هیچگونه قضاوتی صورت نگیرد و نگوییم خوب است یا بد، کافی است یا ناکافی، این سنجش در سطح Assessment باقی مانده است. ولی اگر در مورد این عدد قضاوت صورت گیرد ارزشیابی صورت گرفته است.

اعتبار بخشی(Accreditation)

- تعریف اعتبار بخشی (Accreditation): نوعی خاص از قضاوت ارزشی است که منجر به نوع خاصی از تصمیم گیری اساساً (Yes/No) در باره ی یک موسسه یا یک برنامه می شود.
- مثال آزمونهای پایان ترم – اعتبار بخشی یک دانشگاه – آزمون استخدامی

امتیاز بندی (Scoring)

- مفهوم Scoring (وزن دهی- امتیاز دهی): فرایندی است که در آن به هریک از ملاکهای ارزیابی وزن یا امتیازی داده می شود. واین فرایند برای ایجاد رقابت بین ۲ موسسه است.

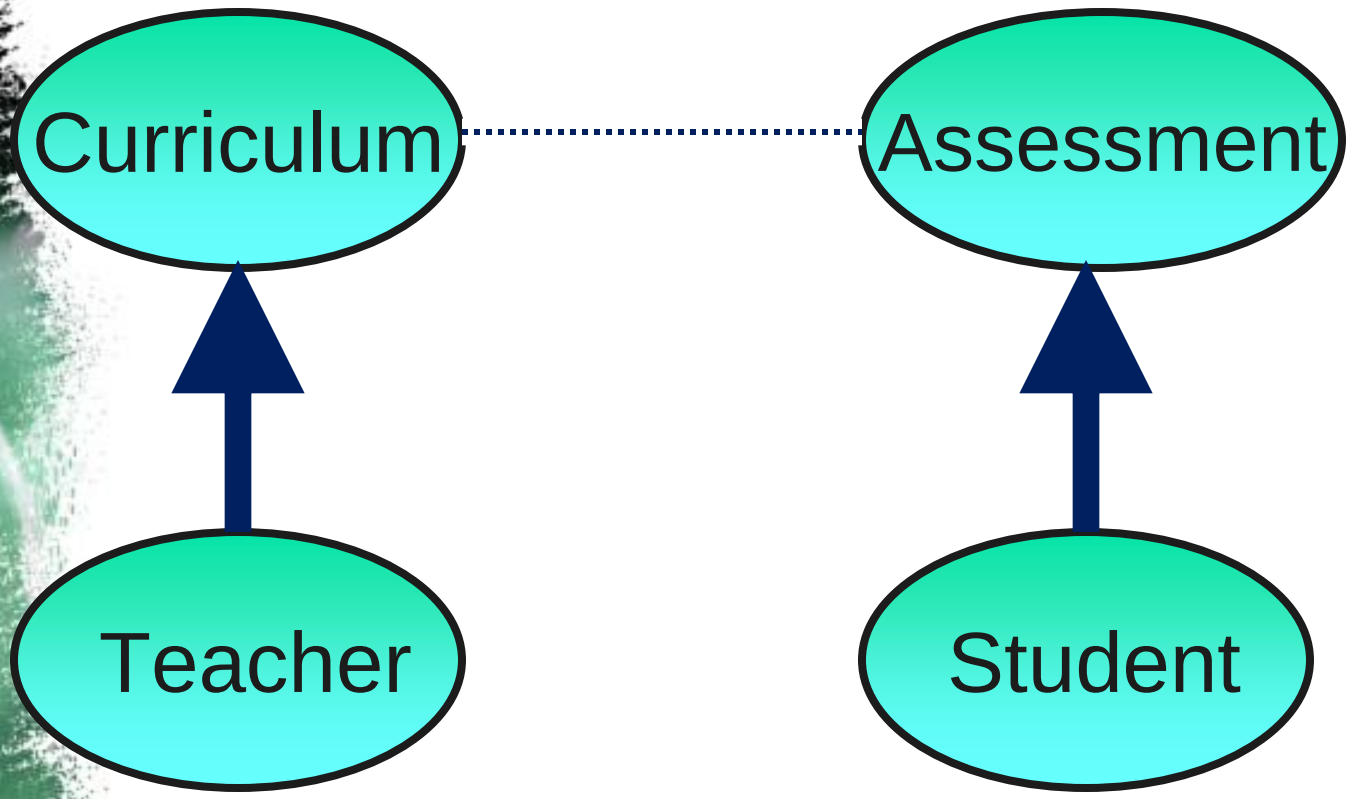
رتبه بندی (Ranking):

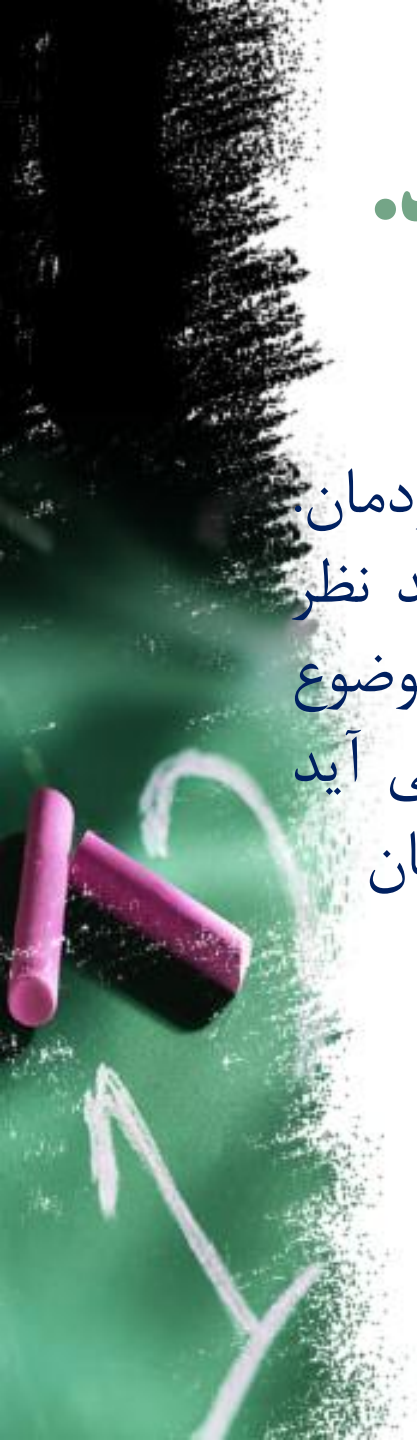
- مفهوم Ranking (رتبه بندی): مقایسه Score چند برنامه، چند موسسه و یا چند فرد را Ranking یا رتبه بندی می گویند مثلا می گوییم S1>S2>S3>S4

آزمودن (Testing)

- یک شکل خاص ازassessment است که محتویات آن آزمون است.
- مانند یک آزمون با ۱۰۰ سوال که هر تعداد از سوالات بخشی از توانایی های فرد را مورد سنجش قرار می دهد.

An alternative view

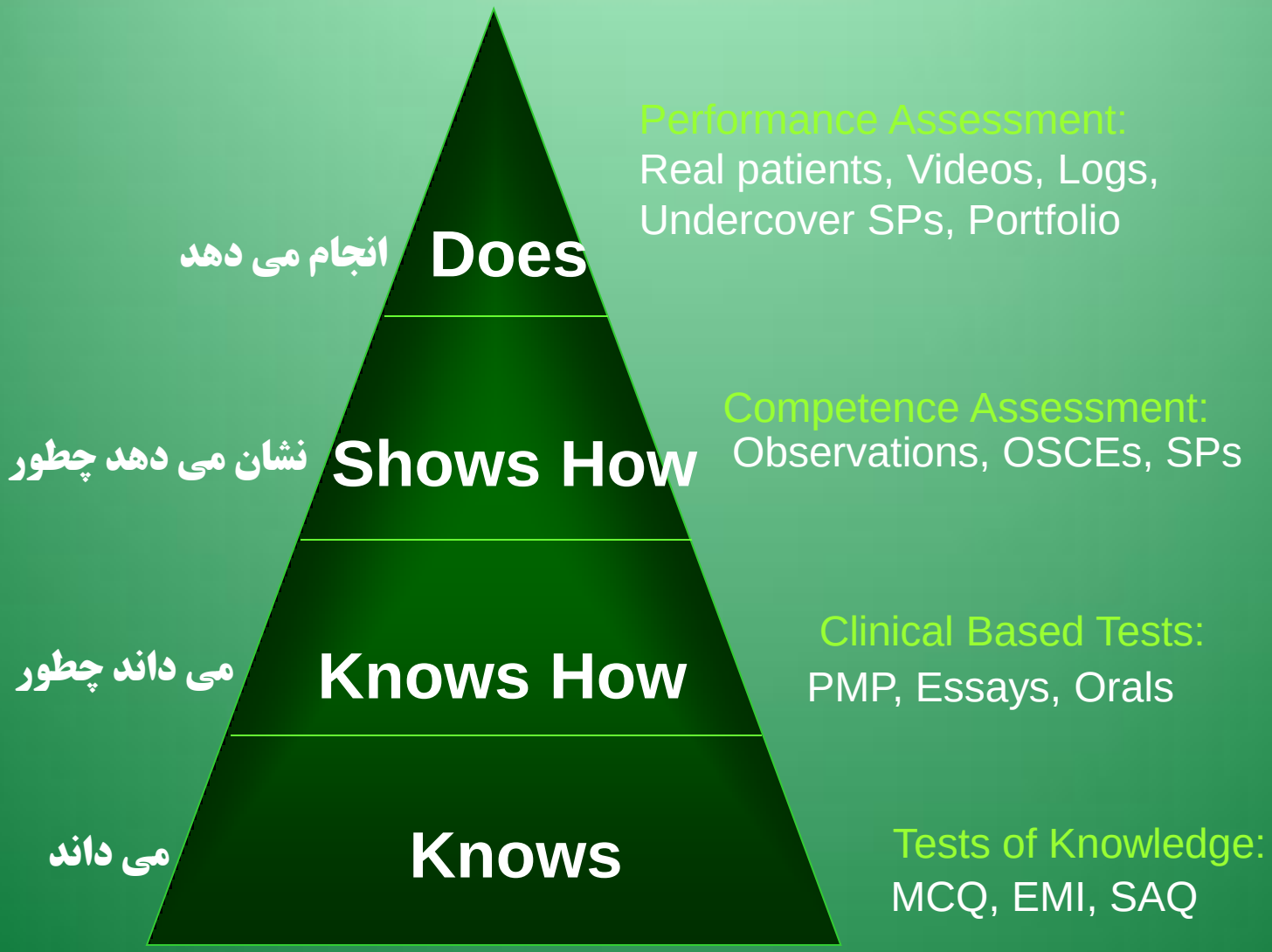




یک آزمون چه فیدبک هایی به ما می دهد.

۱. قضاوت و اندازه گیری سطح عملکرد دانشجو
۲. قضاوت در مورد اثر بخشی برنامه آموزشی و تدریس خودمان. بعنوان مثال اگر در امتحان اکثریت دانشجویان آنچه که مد نظر ومطلوب ما بوده بدست نیاورده اند علیرغم اینکه آن موضوع تدریس شده اثر بخشی آموزشی زیر سوال است. به نظر می آید در اینجا بیشتر ایراد از سوی برنامه آموزشی بوده نه دانشجویان

Miller - Van der Vleuten learning assessment pyramid



۱. Knows دانسته های فرد را می سنجد بعنوان مثال از فرد می پرسیم شیوع بیماری دیابت در ایران چقدر است.

۲. Knows How: آیا دانشجو می تواند آنچه را که در سطح کاربردی آموخته است را بیان کند؟ و آیا روش کار خود یا روش management خود را می تواند بیان کند. بعنوان مثال از دانشجو می پرسیم که اگر یک بیمار دیابت تیپ ۱ به شما مراجعه کرد برای کنترل قند خون بیمار چه اقداماتی را انجام می دهی . و دانشجو آنچه که قرار است در آینده انجام دهد شفاها بیان می کند و یا بصورت کتبی می نویسد.

۳. shows How : در این مرحله عملاً از دانشجو می خواهیم آنچه را که می داند به صورت عملی نشان دهد. بعنوان مثال یک بیمار دیابتی را در اختیار دانشجو قرار می دهیم و از او می خواهیم برای بیمار یک نسخه بنویسد و یا برای بیمار یک Order بنویسد و یا بیمار را معاینه کند.

- نکته : Shows How تحت نظارت و observation استاد انجام می گیرد. و این موضوع جزء محدودیت های این سطح است. زیرا ممکن است یک دانشجو در برابر استاد یک شرح حال کامل بگیرد اما در زمانی که استاد حضور ندارد وقت زیادی را صرف این کار نکند.
- Dose : آنچه که دانشجو در عالم واقع و بدون نظارت و حضور استاد انجام می دهد.

- سطح اول هرم میلر test of knowledge است و دانش فرد را می سنجد.
- سطح دوم clinical based test است و دانش کاربردی بالینی فرد را می سنجد .
- سطح سوم competence assessment است توانایی فرد را می سنجد.
- سطح چهارم performance assessment است و عملکرد فرد را می سنجد.

تفاوت performance با competence

- competence تحت observation است ولی performance تحت نظارت یا observation نیست.
- نکته competency شرط لازم برای performance هست ولی شرط کافی نیست.
- بعنوان مثال دانشجو می داند چگونه شرح حال کامل بگیرد اما مسئولیت پذیر نیست و انگیزه لازم را ندارد بنابراین شرح حال یک دقیقه ای می گیرد.

پاسخ به برخی از سوالات قبل از انتخاب ابزار ارزشیابی دانشجو

۱-چه چیزی را می خواهیم ارزیابی کنیم؟

۲-چرا می خواهیم اندازه گیری کنیم؟

۳-آیا ابزار اندازه گیری ما معتبر وvalid است؟

۴-آیا ابزار اندازه گیری ما پایا و Reliable است؟

۵-آیا روش ارزشیابی ما امکان پذیر یا Feasible است؟

سوال اول: چه چیزی را می خواهیم اندازه گیری کنیم؟

- سوال بسیار مهمی است زیرا یکی از ایرادهای اساسی که به آزمونها وارد است این است که عمدتاً حافظه فرد یا Recall را اندازه گیری می کنند.
- در حالیکه یک پزشک یا پرستار صرفاً با داشتن حافظه خوب و قوی پزشک یا پرستار خوبی نخواهد بود. و نیازمند توامندیهای دیگر از جمله مهارتهای ارتباطی – اخلاق پزشکی و حرفه ای و ... است .

یک اصل کلی:

- آزمونها جهت سنجش اهداف آموزشی یا Educational Objectives طراحی می شوند. به بیان دیگر ما قبل از شروع یک دوره، اهداف آموزشی دوره را تعیین می کنیم و در پایان دوره بررسی می کنیم که آیا اهداف آموزشی محقق شده یا خیر و روش بررسی ما همان آزمون است.

عناصر خروجی سنجش
Output measures

۱- عنصر دانش Knowledge component

۲- عنصر مهارت Skill Component

۳- عنصر نگرشی attitudinal Component

حیطه های دانش

۱. Cognitive domain(حیطه شناختی)

۲. Affective domain(حیطه نگرشی)

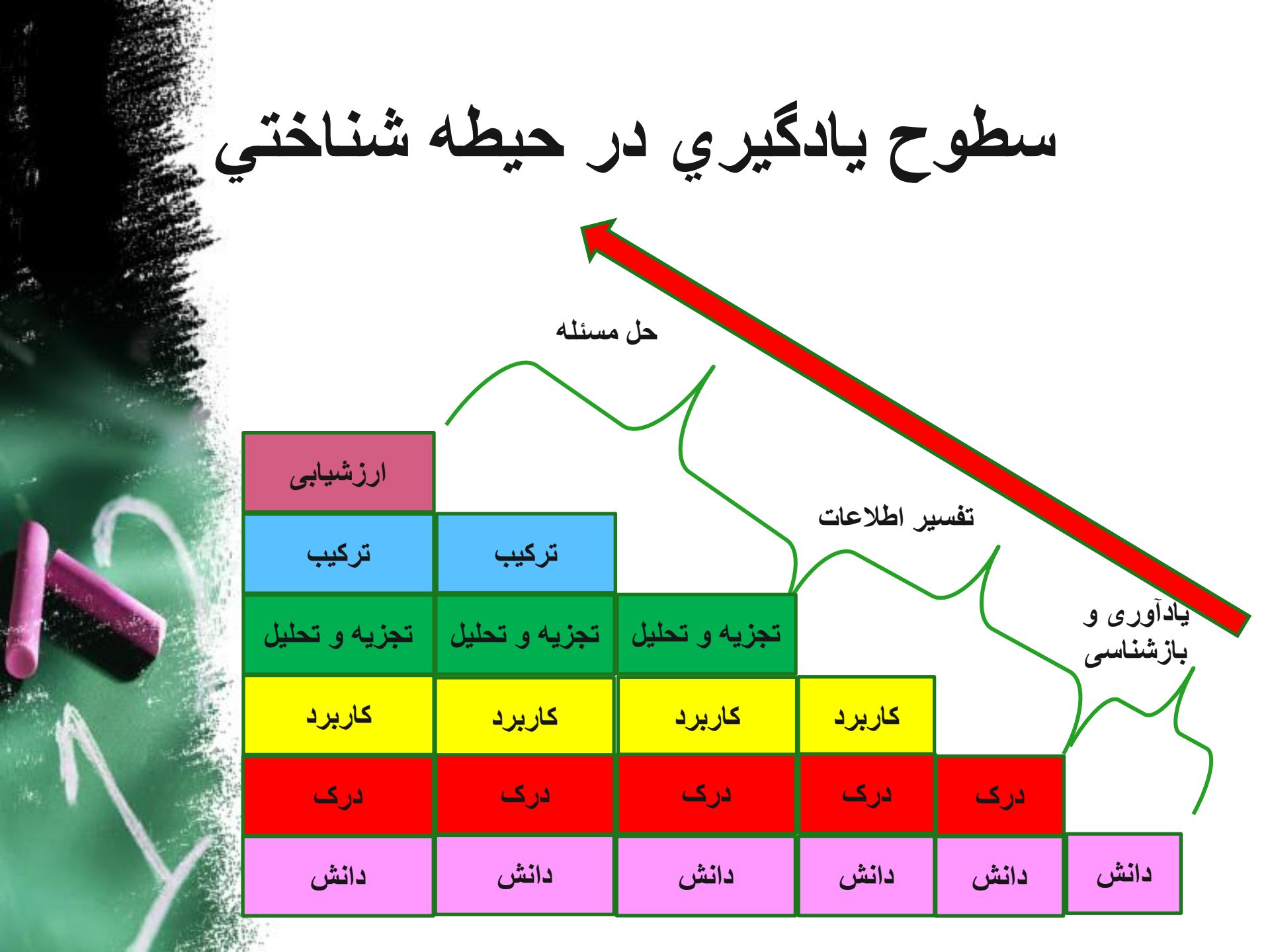
۳. Psychomotor domain(حیطه مهارتی)

- زمانی که ما اهداف آموزشی را برای یک course plan یا lesson plan می نویسیم باید این سه حیطه را مد نظر قرار دهیم.

حیطه شناختی

حیطه شناختی آن دسته از هدفهای آموزشی است که مستلزم یادآوری و بازشناسی آموخته هاست که از دانستنیهای ساده شروع می شود وبه سطوح پیچیده تر می رسد.

سطوح يادگيري در حيطه شناختي



۱- دانش

۱- دانش

عامل فهمیدن در اینجا دخالت کمتری دارد و شامل آموخته های حفظی پایدار است. که خود نیز ممکن است ساده یا مشکل باشد. مثل دانستن معنی یک لغت یا دانستن یک تئوری یا جریان حوادث

مصادیق یادگیری در سطح «یادآوری»:

نام بردن

لیست کردن

مثال : کدام هورمون، توسط غده پانکراس تولید و ترشح میشود؟

الف- کورتیزول

ب – انسولین

ج - سوماتوتروپین

د- گونادوتروپین

۲- درک و فهم

آن دسته از فعالیتهای آموزشی که به درک بیشتری از آگاهی سطحی احتیاج دارند در این طبقه قرار می گیرند.

فهمیدن سبب می شود یادگیری رضایت بخش و معنی دارشود. فهم مطالب سبب می شود دیرتر فراموش شود.

مصادیق یادگیری در این سطح:

- مثال زدن
- توضیح دادن
- خلاصه کردن

- دانشجو، کلیشهٔ رادیوگرافی قفسه سینه یک فرد سالم را با یک فرد مسلول مقایسه کرده و تفاوتهاي آنها را بیان میکند.

۳- کاربرد

آموخته ها در این حیطه نسبت به حیطه قبلی عمق بیشتری پیدا می کند.

مصادیق در این حیطه:

- محاسبه کردن
- تغییر دادن
- نمایش دادن

مثال : برای بیماری ۳۴ساله میزان ۳۰۰۰ سی سی سرم در ۲۴ ساعت تجویز شده است ، چند قطره در دقیقه باید دریافت نماید ؟

الف- ۲۰

ب – ۲۵

ج – ۳۰

د- ۳۵

۴- تجزیه و تحلیل

توانایی تقسیم مطلب به اجزای تشکیل دهنده آن

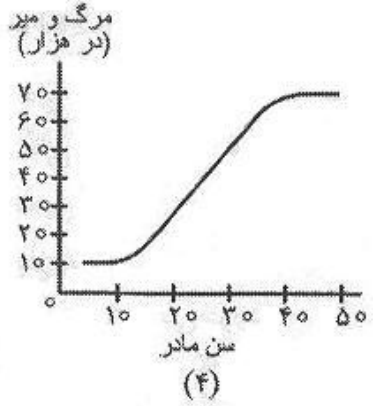
میزان دخالت فهمیدن در این حیطه زیادتر از طبقه قبلی است مثل تجزیه و تحلیل یک واقعه تاریخی

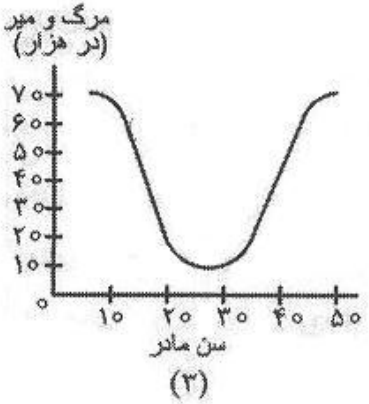
مصادیق یادگیری در این سطح :

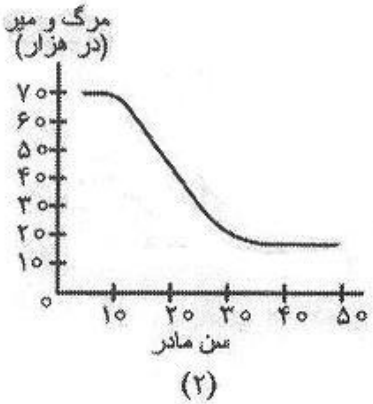
- با نمودار نشان دادن
- اثبات کردن
- به اجزا تقسیم کردن

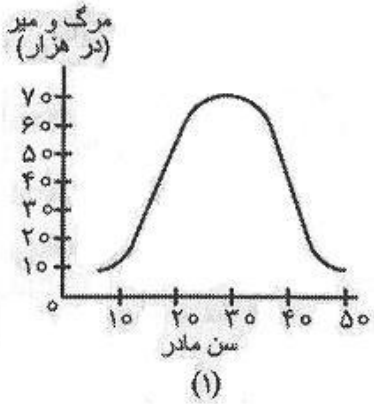
ساختار لیپیدها و هیدراتهای کربن را مقایسه نموده شباهت و تفاوتهای آنها را بنویسید.

منحنی رابطه سن مادر با خطر مرگ جنین یا نوزاد تقریبا به چه شکل است









۵- ترکیب

شامل توانایی در هم آمیختن اجزا، به نحوی که ساخت کلی جدیدی تشکیل می شود.

در این حیطه نتایج یادگیری به رفتارهای خلاق منتهی می شود. مثل ایراد کردن نطق، ارائه یک طرح بهداشتی

مصادیق رفتاری مورد استفاده در این حیطه شامل:

- طبقه بندی می کند
- تجدید نظر می کند
- سخنرانی می کند

مثل: یک مرد ۳۷ ساله به علت زخم معده تحت عمل گاسترکتومی ساب توتال قرار گرفت در دومین روز صبح بعد از عمل، درجه حرارت بیمار ۳۹ درجه همراه با تاکی کاردی ۱۳۰ و بطور جزئی تاکی پنه بود.

محتملترین تشخیص کدام است؟

الف – عفونت محل زخم

ب – عفونت ادراری

ج – ترومبوفلبیت

د- آتلکتازی

۶- ارزشیابی و قضاوت

اگر آموخته ها به کاملترین شکل خود برسد توانایی و قدرت داوری در خود پدید می آید و می تواند در مورد قضایا و پدیده ها به قضاوت بپردازد.

مثل نوشتن یک مقاله انتقادی

مصادیق یادگیری در این حیطه:

- مقایسه کردن
- انتقاد کردن- نقد کردن
- تفسیر کردن

مثال: کودک را از نظر مصرف مکملها و واکسیناسیون ارزیابی نمایید؟

سوال دوم

چرا می خواهیم ارزیابی کنیم؟

- ارزیابی دانشجو به منظورهای خاصی صورت می گیرد. و انتخاب نوع ارزشیابی هدف ما را از ارزشیابی مشخص میکند.

انواع ارزشیابی

summative evaluations
(ارزشیابی تجمعی)

- summative evaluations یک نوع قضاوت ارزشی است که منجر به تصمیم گیری بله /خیر می شود. (قبول –رد)
- این نوع ارزشیابی به این منظور صورت می گیرد که ببینیم آیا فرد حداقل هایی که مدنظر بوده است را کسب کرده یا نه. و ممکن است انتهای دوره، فاز و یا انتهای برنامه آموزشی باشد.
- این نوع ارزشیابی با رویکرد social accountability یا پاسخگویی اجتماعی معنا پیدا می کند. آیا جامعه می تواند به دانشجویی که از موسسه من فارغ التحصیل می شود اعتماد کند. با این رویکرد بایستی پل هایی ایجاد کنیم که هر کس به راحتی نتواند از آن عبور کند. و کسانی که قابلیت های لازم را دارند و یا حداقل استانداردهای تعیین شده را دارند بتوانند از آن عبور کنند.

formative evaluations -۲
(ارزشیابی سازنده)

- ارزشیابی است که در آن با دادن فیدبک به دانشجو عملکرد او را انعکاس می دهیم و چنانچه او نقطه ضعفی داشت در جهت بهبود آن تلاش کند.
- این ارزشیابی منجر به تصمیم گیری بله / خیری نمی شود .

آزمونهای ملاکی یا مطلق

criterion – references measurements absolute evaluations

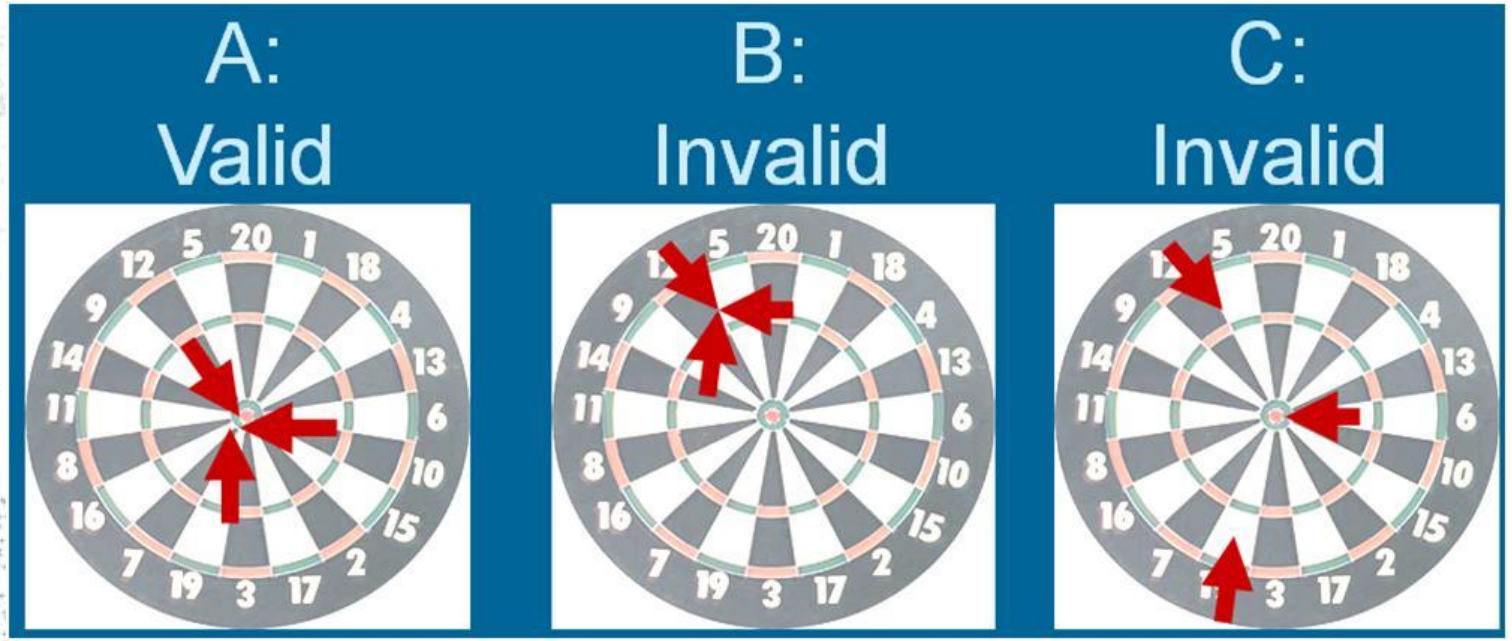
در این نوع ارزیابی یا آزمون ملاکهایی وجود دارد که قطعی است و از پیش تعیین شده است و غیر قابل تغییر است. و اگر دانشجو ملاکها را داشت قبول و در غیر اینصورت رد می شود.

آزمونهای هنجاری یا نسبی
relative criteria test
norm – referenced

در این نوع ارزیابی یا آزمون ملاکها از پیش تعیین نشده است. ابتدا آزمون گرفته می شود و سپس میانگین نمرات محاسبه می شود و افرادی بالا و تعدادی پایین این میانگین قرار می گیرند.(مثال آزمون کنکور)

سوال سوم

آیا ابزار اندازه گیری ما معتبر وvalid است؟



- validity به این معناست که آیا امتحانی را که طراحی کرده ایم واقعاً آنچه را که ما می خواستیم اندازه گیری می کند یا خیر
- به بیان دیگر امتحان مانند یک متر است، آیا این متر آنچه را که باید اندازه گیری می کند یا خیر؟

مثال

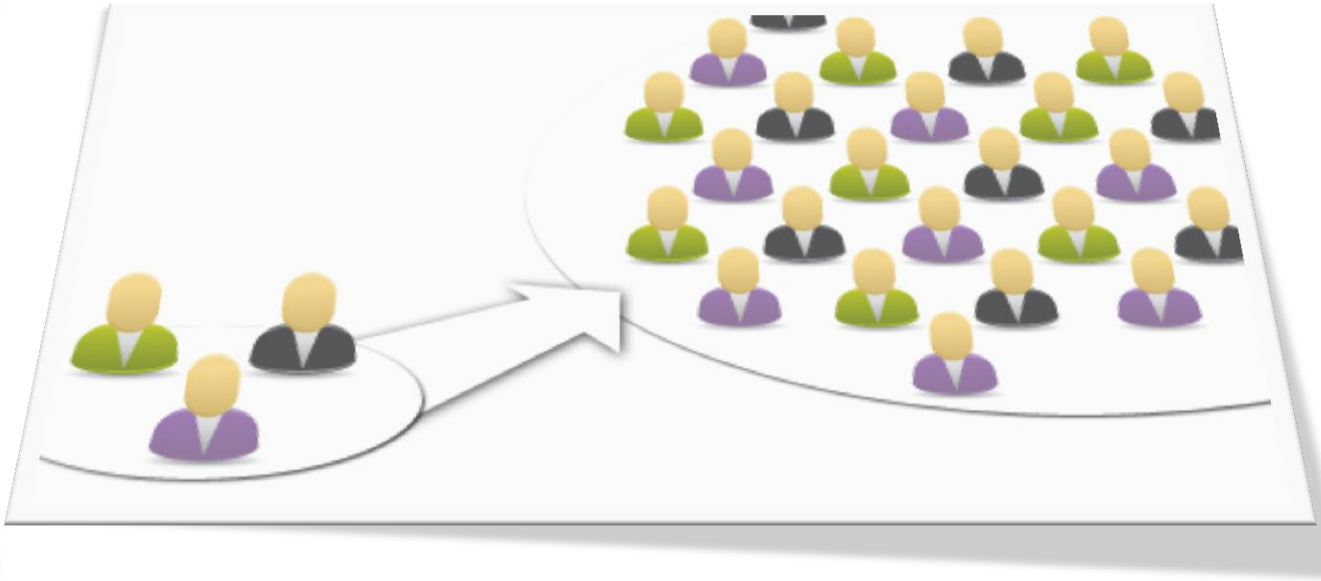
- برای مثال آزمونMCQ برای اندازه گیری recall یا knowledge میتواند valid ومناسب باشد اما برای اندازه گیری professionalism یا رفتار حرفه ای مناسب نمی باشد.
- بنابراین یک آزمون ممکن است برای اندازه گیری یک توانایی valid باشد و برای اندازه گیری توانایی دیگر valid نباشد.

انواع validity

- content validity: آیا ما همه آنچه مدنظر داشتیم را اندازه گرفته ایم.
- Criterion validity: آیا آزمون ما یا با عملکرد آتی فرد همخوانی دارد.
- Construct validity: آنچه که ما می سنجیم با آنچه که قرار است بسنجیم عیناً یکی است.

content validity

روایی محتوایی



content validity

- هر امتحانی مانند یک sample یا نمونه برداری است.
- در یک آزمون ما نمی توانیم همه ی محتوا را در سوالات بیاوریم و مورد سنجش قرار دهیم. همانگونه که در پژوهش قادر به سر شماری نیستیم و از نمونه گیری استفاده می کنیم.
- یعنی در آزمون نمونه ای از دانش و توانایی فرد را می سنجیم و نتیجه آن را به اهداف و محتوی وسیع تری تعمیم می دهیم . یعنی اگر دانشجو در آزمون به ۲۰ سوال پاسخ داد احتمالاً به ۲۰۰ سوال هم پاسخ درست خواهد داد.

روش Messik برای سنجش Content validity

- از نظر آقای Messik روایی محتوایی یا Content validity دو جنبه و دو وجه دارد.

۱. Content relevance (ارتباط محتوایی)
یعنی هر سوال امتحان باید در ارتباط با یک سری از اهداف آموزشی باشد.

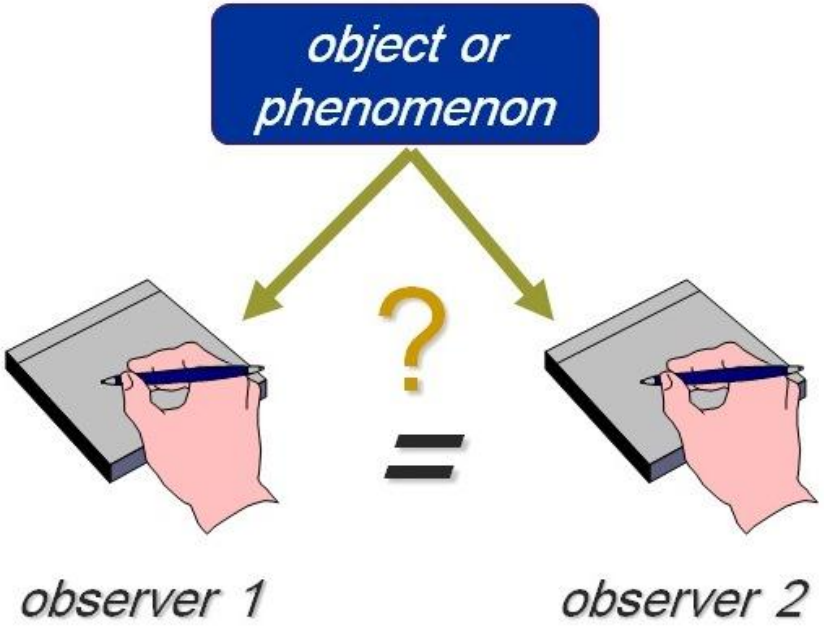
۲. Content coverage (پوشش محتوایی)
یعنی هر یک از اهداف آموزشی بایستی توسط سوالاتی ارزیابی شده باشد .

Checking content validity for a course in cardiology

Question	Content Area				
	Physiology	Semiology	Diagnosis	Anatomy	Treatment
1	○		○		
2					
3	○				
4					
5			○		
6					
7					○
8					
⋮					
20		○			

سوال سوم

آیا ابزار اندازه گیری ما پایا و Reliable است؟



تعریف reliability

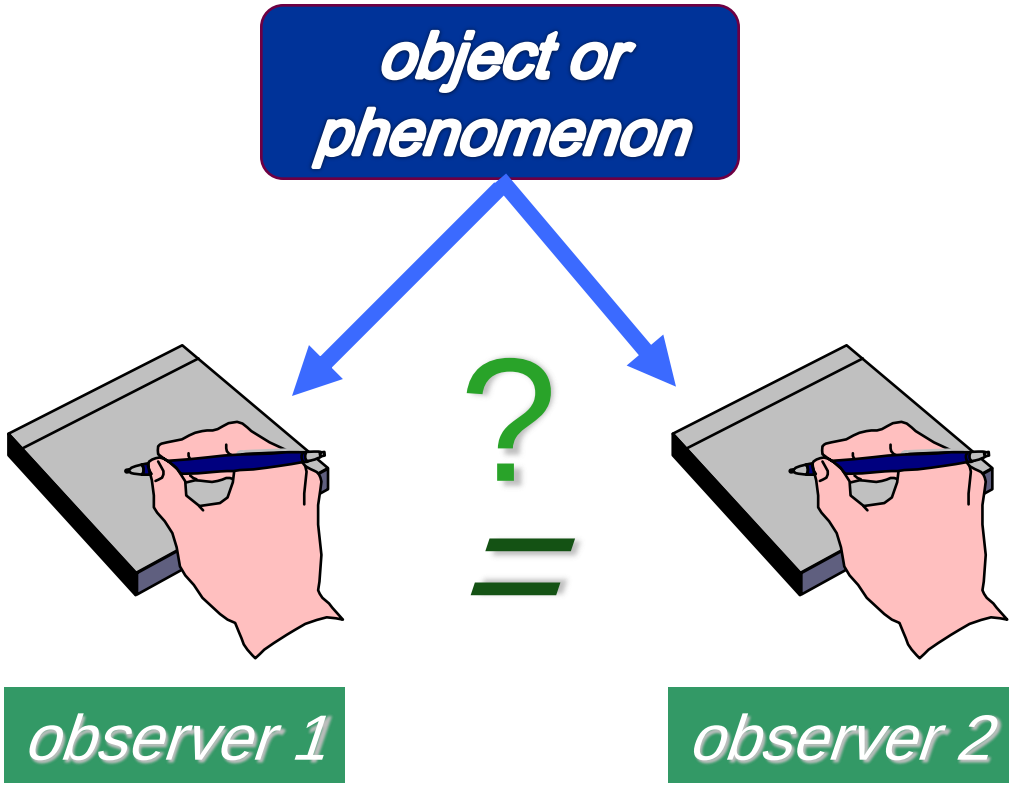
- میزان همخوانی یا ثبات نمراتی که توسط یک فرد، به دنبال تکرار آزمون یا توسط چند فرد بطور همزمان یا توسط چند ابزار اندازه گیری معادل صورت می گیرد.
- بنابراین reliability یعنی تکرار پذیری یا ثبات به این معنی که اگر اندازه گیری در طول زمان یا توسط چند نفر انجام شود یکسان باشد البته با این پیش فرض که آن اندازه تغییر نکرده باشد.

اهمیت reliability

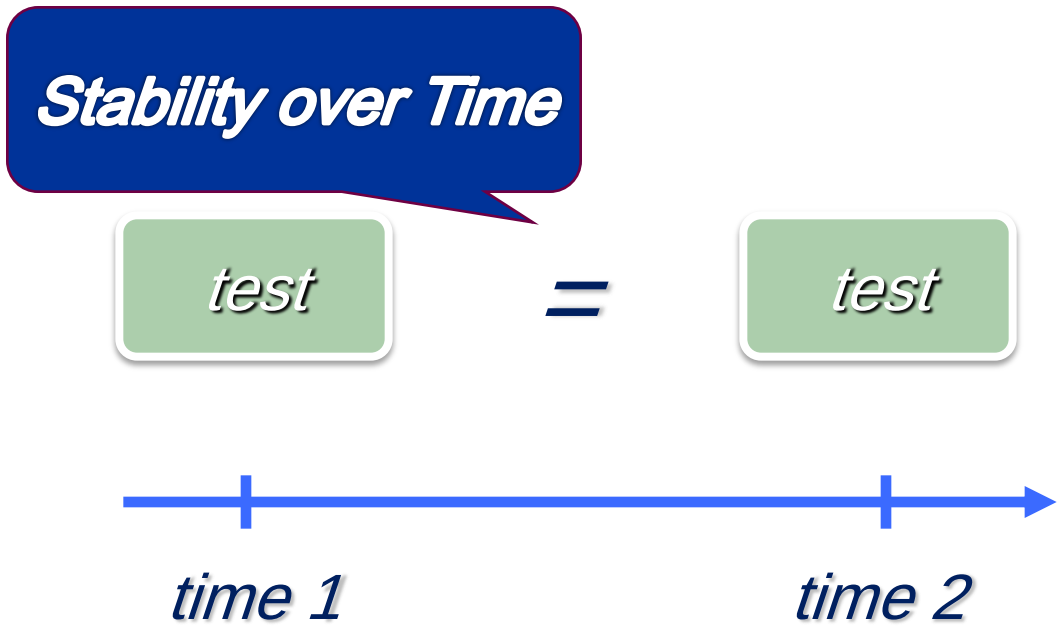
- چون نمرات مبنای استنباط، قضاوت ارزشی و تصمیم گیری ما قرار می گیرد و بر اساس آن فردی را رد یا قبول می کنیم اگر نمره ای که مبنای قضاوت ارزشی ما قرار می گیرد چند لحظه بعد تغییر کند قطعاً قضاوت ارزشی ما قضاوت درستی نخواهد بود.

انواع reliability

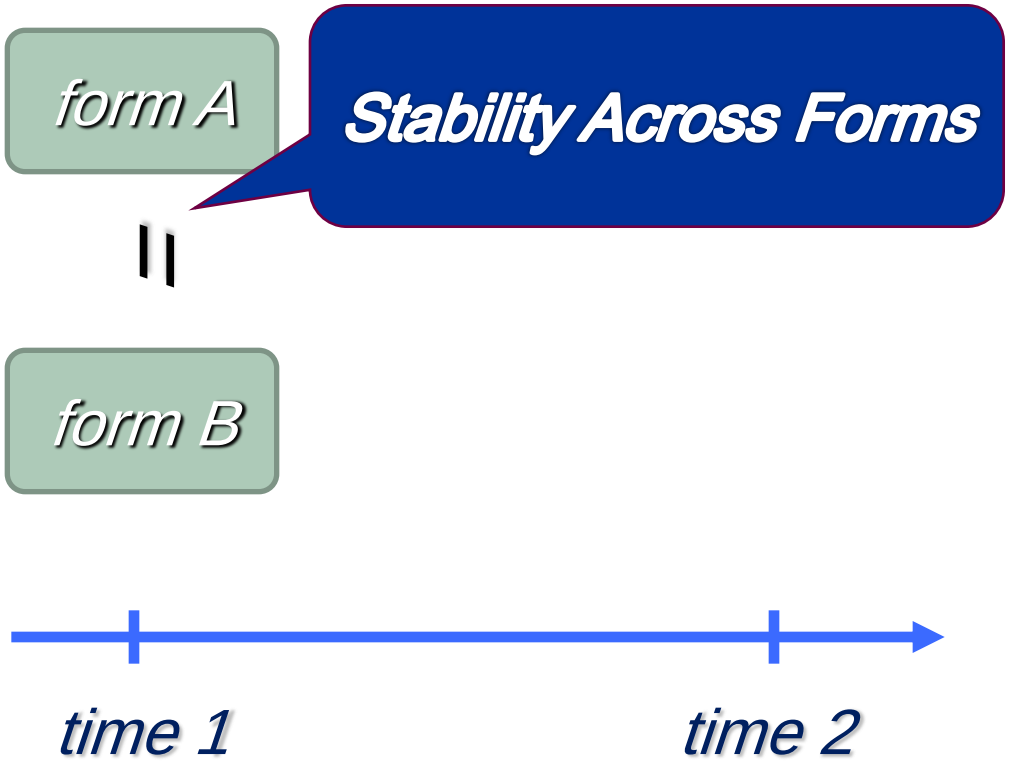
Interrater or Interobserver Reliability



Test-Retest Reliability



Parallel-Forms Reliability



سوال پنجم: آیا روش ارزیابی ما امکان پذیر یا feasible است.

- یک آزمون ممکن است از روایی و پایایی بالایی برخوردار باشد اما امکان پذیر نباشد . هزینه برگزاری آن زیاد است نیرو و پرسنل خاصی را نیاز دارد – فضای فیزیکی خاصی را می طلبد – زمان زیادی را نیاز دارد– دشوار است.

سوالاتی که در خصوص feasibility مطرح است.

۱. برای طراحی آزمون چقدر زمان مورد نیاز است؟

۲. برای اجرای آزمون چه میزان زمان مورد نیاز است؟

۳. چه میزان زمان برای تصحیح آن مورد نیاز است؟

۴. آیا کل فرایند آزمون هزینه – اثر بخش است ای خیر؟

۵. آیا نیروی انسانی – فضای فیزیکی و منابع لازم وجود دارد یا خیر؟

تاکسونومی سوالات چند گزینه ای

تاکسونومی یک (محفوظات)

سؤالی که تنها برمبنای حافظه باشد: شایعترین تظاهر بیماری x کدام است

شایع ترین تومور مغزی کدام است ؟

الف - منینجیوما

ب -استروسیتوما

ج -همانجیوما

د -اولیگودندروگلیوما

%۴۵-۳۵% از مجموع سوالات را تشکیل میدهند

زمان جواب دادن ۳۰-۳۵ ثانیه

توکسونومی دو (تشخیص و تفسیر)

وقتی ذهن دانشجو با انجام یک فرآیند به پاسخ می رسد

توکسونومی دو (تشخیص و تفسیر)

نوجوان ۱۶ ساله ای پس از برگشتن از مسابقه فوتبال شکایت از درد شدید در بیضه سمت چپ و شکم همراه با تهوع واستفراغ به مدت ۳ ساعت دارد در معاینه شکم نرم وسمع طبیعی داشت ومحتوی کیسه بیضه سفت ودردناک وکبود بود.بعلت درد معاینه حلقه انگوینال میسرنشد.گرافی ساده شکم وازمایش کامل ادرار طبیعی بود تشخیص شما چیست ؟

الف -اورکییت

ب -فتق انگوینال مختنق

ج -تورشن تستیکول

د -هماتوم ناشی از ضربه

۳۰-۲۰% از سوالات و زمان ۵۰-۴۵ ثانیه

تاکسونومی ۳ (تصمیم گیری)

وقتی ذهن با انجام دو مرحله فرآیند به پاسخ می رسد.

تاکسونومی ۳ (تصمیم گیری)

چه تصمیمی برای این جوان میگیرید ؟

الف -تجویز مسکن و پیگیری تا ۲۴ساعت

ب – دتورشن فوری بیضه

ج -اورکیکتومی الکتیو

د -هرنیورافی فوری

۳۰-۳۵% سوالات را تشکیل میدهند

زمان ۵۵-۶۰ ثانیه

یاداوری مهم

- بعضی از سوالات ظاهری شبیه توکسونومی ۳ دارد در حالیکه فقط محفوظات را می سنجند !!!!
- اگر Case / Problem فوق قبلا عینا درکلاس درس مطرح شده باشد و دانشجو بتواند با حافظه به آن پاسخ دهد تاکسنومی سطح یک می باشد

رعایت قواعد ساختاری سوال (چک لیست میلمن)

چک لیست میل من، همان پرسشنامه قواعد تهیه سوال های چند گزینه ای است. هر سوال باید طوری نوشته شود که تنها به وسیله کسانی جواب سوال را می دانند به درستی پاسخ داده شود و کسانی که تسلط کامل بر مطلب ندارند، نتوانند به آن پاسخ درست بدهند.

چک لیست میلمن جهت بررسی رعایت اصول ساختاری در سئوالات چند گزینه ای:

سوال	بلی	خیر
۱-آیا بخش اعظم اطلاعات در ساقه سوال گنجانده شده است؟		
۲ - آیا سوال یک هدف اختصاصی یادگیری را مورد ارزیابی قرار می-دهد؟		
۳ - آیا لغات استفاده شده در ساقه یا گزینه‌ها، شفاف و مستقیم بیان شده‌اند؟		
۴ - آیا از کاربرد گزینه منفی برای ساقه منفی خودداری شده است؟		
۵ – آیا از کاربرد گزینه‌های نظیر همه موارد هیچکدام و گزینه‌های ترکیبی خودداری شده است؟		
۶ – آیا ازکاربرد گزینه‌های متضاد یکدیگر خودداری شده است؟		

چک لیست میلمن جهت بررسی رعایت اصول ساختاری در سئوالات چند گزینه ای:

سوال	بلی	خیر
۷- - آیا لغات مثبت درساقه سوال استفاده شده است یا در صورت منفی بودن ساقه سوال لغایت منفی مشخص شده اند؟		
۸ – آیا هر سوال مستقل از سوالات دیگر می باشد؟		
۹ – آیا گزینه ها از نظر طول، ساختار لغوی و سبک نگارش هم سنگ هستند؟		
۱۰ – آیا تا حد امکان از کاربرد ساختار لغوی و سبک نگارش هم سنگ هستند؟		
۱۱ – آیا کلمات به کاررفته در ساقه و یا گزینه ها از نظر املایی صحیح هستند؟		
۱۲ – آیا گزینه ها بطور عمودی لیست شده اند.		

Item Analysis in Student Assessment

تحلیل آزمون ها

ضریب سهولت
Facility Index

- تعریف: نسبتی از افراد کل گروه، که به یک سوال پاسخ صحیح داده اند. و میزان آن بین صفر تا یک است
- اگر تمام افراد گروه یک آیتم خاص مثلا" سوال یک را درست زده باشند می گوییم **Facility Index** ما ۱ است.
- اگر **(FI) = 1** یعنی آن آیتم آسان است.
- اگر همه یک سوال را غلط زده باشند می‌گوییم **(FI) = 0** و معنی آن این است که این آیتم یک آیتم مشکلی است.
- لذا هر اندازه ضریب سهولت به ۱ نزدیکتر باشد آن سوال آسانتر است.

فرمول محاسبه
Facility Index

$$FI = \frac{\text{نسبت افراد قوی که به یک سوال پاسخ صحیح داده اند} + \text{نسبت افراد ضعیف که به یک سوال پاسخ صحیح داده اند}}{۲}$$

ضریب دشواری
Difficulty Index

- *تعریف:* درصد کل آزمون شوندگانی که به یک سؤال جواب درست می دهند.
- برای محاسبه ضریب دشواری اگر تعدادآزمون شوندگان ≤۲۰ نفرباشد، برگه‌های آزمون را به دو دسته بالا و پایین تقسیم می کنیم.
- اگر تعداد بیش از ۴۰ نفرباشد، بهترین رقم برای گروه بالا و پایین ۲۷٪است.
- بطور کلی می توان، از ۲۵ تا ۳۳ درصد را انتخاب نمود.

فرمول محاسبه
Difficulty Index

$$DI = \frac{\text{تعداد افرادی که از گروه قوی پاسخ صحیح داده اند} + \text{تعداد افرادی که از گروه ضعیف پاسخ صحیح داده اند}}{\text{کل افراد گروه قوی+ کل افراد گروه ضعیف}}$$

مثال

نتایج یک سؤال ازیک آزمون
(تعدادآزمون شوندگان=۸۹نفر)

گزینه	۲۷٪بالا	۲۷٪پایین	۴۶٪متوسط
الف	۲	۱۱	۲۰
ب	۴	۴	۵
ج	۱۸	۹	۱۵
د	۰	۰	۱
جمع	۲۴نفر	۲۴نفر	۴۱نفر

$$DI = 18 + 9 \div 48 = 0.5$$

- بنابراین هرچه ضریب دشواری **بالاتر** باشد ،سؤال **آسان** تر است.
- ضریب دشواری مناسب آن است که به **۰/۵** نزدیک باشد.
- بطورکلی ضریب های دشواری بین **۰/۳ تا ۰/۷** حداکثر اطلاع را در باره تفاوت بین آزمون شوندگان به دست می دهند.

Discrimination Index (DI)
ضریب افتراق

- حاصل تفاضل نسبتی از افراد گروه قوی یا High score group که به یک آیتم پاسخ درست داده اند از افراد گروه ضعیف یا Low score group که به همان آیتم پاسخ درست داده اند.

فرمول محاسبه ضریب افتراق
Discrimination index

تعداد افرادی که از گروه قوی پاسخ صحیح داده اند

–

تعداد افرادی که از گروه ضعیف پاسخ صحیح داده اند

تعدادافراد یک گروه (بالا یا پایین)

مثال

نتایج یک سؤال ازیک آزمون
(تعدادآزمون شوندگان=۸۹نفر)

گزینه	۲۷٪بالا	۲۷٪پایین	۴۶٪متوسط
الف	۲	۱۱	۲۰
ب	۴	۴	۵
ج	۱۸	۹	۱۵
د	۰	۰	۱
جمع	۲۴نفر	۲۴نفر	۴۱نفر

$$DI = 18 - 9 \div 24 = 0.375$$

- اگر یک آیتم را همه افراد **High score group** مان (۲۵٪بالا) درست زده باشند پس نسبت افراد می شود یک و همه افراد **Low score group** مان هم این را درست زده باشند این هم می شود یک بنابراین:

DI= 1-1= 0

- اگر همه افراد **High score group** مان این آیتم خاص را درست زده باشند و همه افراد **Low score group** مان این آیتم را غلط زده با شند:

DI= 1-0= 1

- یک بیشترین میزان و مثبت ترین مقدار **Discrimination Index** می باشند و این آیتم خیلی خوب می تواند بین افراد قوی و ضعیف تمایز بدهد و این آیتم بسیار مطلوب و ایده آل است.

- اگر ضریب افتراق منفی شود این آیتم نه تنها بی ارزش است بلکه مخدوش کننده ی آزمون ما می باشد.
- زیرا افراد قوی آن آیتم را اشتباه زده اند و افراد ضعیف آن را درست زده اند.
- ضریب افتراق بین ۱ تا ۱- است
- منفی یک کمترین مقدار Discrimination Index است و در جهت معکوس هدف ما کار می کند این آیتم موجب کسب نمره برای افراد ضعیف و از دست دادن نمره برای افراد قوی است و حتماً باید حذف شود.

تحلیل

Discrimination index

Discrimination Index > 0

آیتم مناسبی است

Discrimination Index = 0

بهتر است حذف شود ولی چنانچه حذف نشود مشکل خاصی ایجاد نمی کند (آیتم خنثی)

Discrimination Index < 0

حتما باید حذف شود


با سپاس
از توجه و همراهی شما
صفایی پور