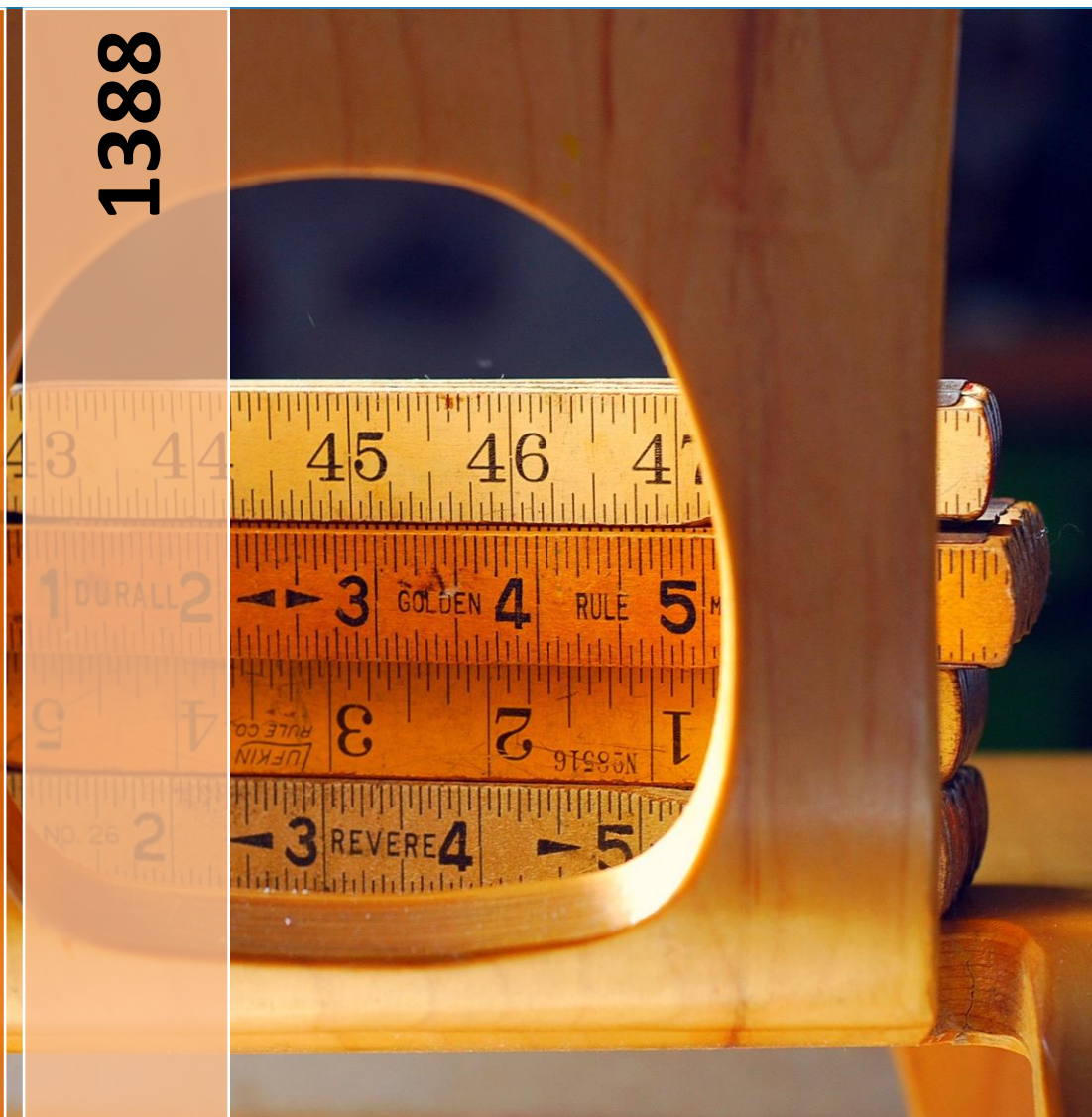


اندازه گیری، آزمون، سنجش و
ارزشیابی آموزشی

مرکز مطالعات و توسعه آموزش
پزشکی

1388



Education development center

جلسه اول

1388

فهرست مطالب

اهداف كلي	3
1 اصول و مباني اندازه گيري	3
1.1 تعريف اندازه گيري	4
1.2 انواع مقياس هاي اندازه گيري: مقياس اسمي	4
1.3 انواع مقياس هاي اندازه گيري: مقياس ترتيبی	5
1.4 انواع مقياس هاي اندازه گيري: مقياس فاصله اي	6
1.5 انواع مقياس هاي اندازه گيري: مقياس نسبي	6
2 اصول و مباني آزمون ها، آزمون و سنجش	7
2.1 تعريف آزمون	8
2.2 تعريف آزمون	8
2.3 تعريف سنجش	9
2.4 تفاوت سنجش و اندازه گيري و آزمون	10
3 اصول و مفاهيم ارزشيابي	11
3.1 تعريف ارزشيابي و ارزشيابي پيشرفت تحصيلی	12
3.2 مقايسه اندازه گيري، آزمون، سنجش و ارزشيابي با يکديگر	13
4 دسته بندي آزمون هاي مورد استفاده	15
4.1 دسته بندي آزمون ها با توجه به ويژگي هاي مورد اندازه گيري: آزمون هاي توانايي	15
4.2 دسته بندي آزمون ها با توجه به ويژگي هاي مورد اندازه گيري: آزمون هاي عاطفي	17
4.3 دسته بندي آزمون ها با توجه به چگونگي تهيه آنها: آزمون هاي ميزان شده	19
4.4 دسته بندي آزمون ها با توجه به چگونگي تهيه آنها: آزمون هاي معلم ساخته	19
20 خلاصه	
آزمون كلي	Error! Bookmark not defined.
منابع	



شناسه جستار

عنوان جستار: اندازه

گیری، آزمون، سنجش

و ارزشیابی آموزشی

نویسنده: دکتر میترا

امینی

دانشیار دانشگاه علوم پزشکی

شیراز

طراح آموزشی: دکتر مانوش

مهرابی

کارشناس دانشگاه علوم پزشکی

شیراز

ویرایش: مریم فخرزاد

آخرین تاریخ به روز

رسانی: 88/8/2

با همکاری:

مرکز مطالعات و توسعه آموزش

پزشکی

مرکز آموزش مداوم (قطب علمی

آموزش الکترونیکی)

دانشگاه علوم پزشکی شیراز

اهداف کلی

در تدوین این درس اهداف کلی زیر مورد نظر بوده است: آشنایی با

- 1- تعاریف اندازه گیری، آزمون، سنجش و ارزشیابی آموزشی
- 2- تفاوت های اندازه گیری، آزمون، سنجش و ارزشیابی آموزشی
- 3- انواع متغیرها
- 4- انواع آزمون ها

1 اصول و مبانی اندازه گیری¹

➤ مفاهیم کلیدی

- 1- اندازه گیری فرآیندی است که تعیین می کند یک شخص یا یک شیء چقدر از یک ویژگی برخوردار است.
- 2- بنا به تعریف مقیاس² مجموعه ای عدد یا نمادهایی دیگر است که ویژگی های آنها بر ویژگی های تجربی اشیایی که اعداد به آنها نسبت داده می شود منطبق است
- 3- مقیاس های اندازه گیری شامل مقیاس های اسمی، رتبه ای، فاصله ای و نسبی هستند.
- 4- طبقه بندی مقیاس ها از ساده به پیچیده و به صورت سلسله مراتبی تنظیم شده است. یعنی اینکه هر طبقه دارای تمامی ویژگی های طبقه قبل به اضافه یکسری ویژگی های اضافه است.

➤ اهداف یادگیری

شما بعد از مطالعه این نوشتار می توانید:

- 1- اندازه گیری را تعریف کنید.
- 2- مقیاس های اندازه گیری را لیست کنید.
- 3- مقیاس اسمی را توضیح دهید.
- 4- مقیاس رتبه ای را توضیح دهید.

¹ Measurement

² Scale

5- مقیاس فاصله ای را شرح دهید.

6- مقیاس نسبی را توضیح دهید.

1.1 تعریف اندازه گیری

در اندازه گیری ویژگی‌ها یا صفات اشیاء و افراد تعیین می‌گردند و مقدار آن ویژگی‌ها یا صفات به صورت عدد یا رقم گزارش می‌شود بنابراین می‌توان گفت اندازه گیری عبارت است از فرآیندی که تعیین می‌کند یک شخص یا یک شیء چه مقدار از یک ویژگی برخوردار است.

اندازه گیری می‌تواند شکل‌های مختلفی به خود بگیرد. یکی از ویژگی‌های متغیر قابلیت اندازه گیری آن است. چنانچه از وزن به عنوان یک متغیر نام ببریم بهترین راه اندازه گیری آن بر اساس کیلوگرم یا گرم می‌باشد در مورد قدم سانتی متر یا متر از عهده آن بر می‌آید.

اندازه گیری فرآیند یا شیوه ای دقیق و نظام‌دار است. نانالی در تعریفی که از اندازه گیری بدست داده برکار برد قواعد تأکید کرده است: اندازه گیری در برگیرنده قواعد است برای نسبت دادن اعداد به اشیاء یا افراد به گونه ای که صفت‌هایی از آنها را بصورت مشخص نشان دهد.

منظور از کاربرد قاعده در اندازه گیری این است که روش‌های نسبت دادن اعداد باید بطور روشن مشخص شوند مثلاً در نمره گذاری نتایج یک آزمون ریاضی وقتی که جواب‌های درست دانش‌آموزان شمارش می‌شوند و به تعداد جواب‌های درست به آنها نمره داده می‌شود شاهد نمونه ای از کاربرد قاعده در اندازه گیری هستیم.

1.2 انواع مقیاس‌های اندازه گیری: مقیاس اسمی

مقیاس اسمی پایین‌ترین سطح یا ساده‌ترین نوع اندازه گیری را معرفی می‌کند. این مقیاس به دو صورت مورد استفاده قرار می‌گیرد یکی برای اسم گذاری و دیگری برای طبقه بندی

• اسم گذاری

در این مقیاس از ارقام و اعداد برای اسم گذاری یا نام گذاری و تشخیص اشیاء و افراد استفاده می‌شود بدون اینکه این اعداد هیچ گونه مفهوم و معنای ریاضی داشته باشند به عنوان مثال شماره‌هایی که بر پشت پیراهن بازیکنان فوتبال نوشته می‌شوند و شماره خیابان‌های شهر نمونه‌هایی از این نوع هستند.

• طبقه بندی

در مقیاس اسمی، علاوه بر نام گذاری، از اعداد و ارقام برای طبقه بندی اشیاء و افراد نیز استفاده می‌شود. باز هم در اینجا اعداد هیچ گونه مفهوم و معنای ریاضی ندارند و تنها برای مشخص کردن طبقه‌های مختلف طبقه بندی به کار می‌روند. فرض کنید یک زمین‌شناس انواع سنگ‌ها را در کیسه‌های مختلف با شماره‌های مختلفی دسته بندی می‌کند. در این دسته بندی، یک عدد معرف کیسه حاوی سنگ‌های آهکی، عدد دیگر نشان دهنده کیسه محتوی سنگ‌های آتش‌زنه، و غیره خواهد بود. در اینجا، جز اینکه اعداد معرف انواع مختلف سنگ‌ها هستند، هیچ

معنی دیگری ندارند. مثالی دیگر برای این مقیاس طبقه بندی افراد به زن و مرد یا بهنجار و نابهنجار است که در آنها مثلاً عدد 1 زن یا فرد بهنجار و عدد 2 مرد یا فرد نابهنجار را معرفی می کند. از آنجا که در تعریف اندازه گیری استفاده از قواعد ضروری فرض شده است، پاره ای از دانشمندان به استفاده از مقیاس اسمی به عنوان یک مقیاس ایراد گرفته و گفته اند که این مقیاس فاقد قواعد اندازه گیری است.

1.3 انواع مقیاس های اندازه گیری: مقیاس ترتیبی

برای متغیرهایی که حالت درجه بندی رتبه ای داشته در مقیاس ترتیبی بیان می شود معمولاً بصورت معینی مثلاً از بهترین تا بدترین یا از بزرگترین تا کوچکترین به آنها رتبه داده می شود. وقتی پژوهشگر با داده هایی سروکار دارد که دارای مقیاسی با درجه بندی رتبه ای است می تواند وضع نسبی مشاهده های مختلف را با هم مقایسه کند. مثلاً بر حسب این مقیاس می توان گفت که درجه سختی الماس بیشتر از درجه سختی فولاد است. از این نقطه نظر، داده های مقیاس ترتیبی با داده های مقیاس طبقه ای مرتب شده شباهت دارد. برای هر دو نوع از داده ها، تصور رتبه یا مرتبه وجود دارد. تفاوت در این است که برای داده های مقیاس طبقه ای مرتب شده، تعداد طبقه ها محدود بوده و برای تفکیک و تشخیص مشاهده های که در داخل یک طبقه قرار دارد تلاشی صورت نمی گیرد. اما در مورد داده های مقیاس ترتیبی، سعی بر این است که همه مشاهده ها از یکدیگر تمیز داده شوند، به نحوی که به دو مشاهده تنها موقعی یک رتبه داده می شود که نتوان معلوم کرد کدام یک وضع نسبی بالاتری دارد. فنون آماری که در مورد داده های مقیاس ترتیبی بکار می رود از فنونی که در مورد داده های مقیاس طبقه ای مرتب شده بکار می رود کاملاً متمایز است. بنابراین تشخیص این دو نوع متغیر از لحاظ تجزیه و تحلیل آماری داده ها حائز اهمیت است.

اعداد در مقیاس ترتیبی، فقط سلسله مراتب اشیاء، افراد یا رویدادها را از لحاظ خصیصه ای معین در داخل یک گروه نشان می دهد و بی آنکه درباره فاصله بین افراد مطلبی بیان کند وضع نسبی آنها را از لحاظ یک خصیصه، مشخص می سازد. مثلاً بر حسب مقیاس ترتیبی می توان گفت که هوشنگ در بین همکلاسان خود نفر دوم، ایرج نفر سوم و مهشید نفر ششم است. اما نمی توان گفت که ایرج از لحاظ معلومات تا چه میزان بهتر از مهشید است و همچنین نمی توان نتیجه گرفت که ایرج دو برابر مهشید معلومات دارد. شرط اصلی داشتن یک مقیاس ترتیبی آن است که پژوهشگر بتواند درباره هر یک از اشیاء یا افراد مورد اندازه گیری معلوم کند که آن شیئی یا فرد از لحاظ خصیصه مورد نظر بیشتر، کمتر یا برابر شیئی یا فرد دیگر سهم داشته است. بنابراین پژوهشگر باید بتواند وضع نسبی موارد موجود را مشخص کند و قادر باشد از لحاظ خصیصه مورد نظر، هر فرد را در نقطه مناسبی از مقیاس قرار دهد.

در مقیاس ترتیبی چون اعداد فقط اطلاعاتی در مورد سلسله مراتب اشیاء و افراد بدست می دهد به جای آنها می توان سلسله اعداد طبیعی 1، 2، 3، 4 و ... و یا سلسله اعداد دیگری را که ترتیب وضع افراد

در آن ثابت بماند بکار برد. اما عملیاتی که با این اعداد می توان انجام داد فقط به بیشتر، کمتر یا برابر محدود می شود و سری عملیات ریاضی که در مورد آنها اعمال می شود نیز محدود و منحصر به محاسبه میانه، نقاط درصدی و ضریب همبستگی رتبه ای است^۱. اندازه گیری متغیرهای روانی- تربیتی به ندرت از سطح مقیاس ترتیبی تجاوز می کند مگر آنکه مفروضه ها و تعاریف خاصی را از پیش بپذیریم.

1.4 انواع مقیاس های اندازه گیری: مقیاس فاصله ای^۲

ویژگی عمده متغیر فاصله ای آن است که به ارزش های آن نمره های عددی تعلق می گیرد. نمره ها را می توان از هم کم کرد و تفاوت بین آنها را بدست آورد. درجات دماسنج سانتی گراد و فارنهایت نمونه بسیار خوبی برای مقیاس فاصله ای است. دماهای 40° و 30° را در نظر بگیرید. می دانیم این دو درجه از یکدیگر تفاوت دارند (داوری در سطح مقیاس طبقه ای). می دانیم دمای 40° از دمای 30° زیادتر (گرمتر) است (داوری در سطح مقیاس ترتیبی). اما در اینجا نیز می دانیم که گرمای 40° دقیقاً 10° از گرمای 30° بیشتر است. بنابراین مقیاس فاصله ای ضمن آنکه اطلاعات مربوط به مقیاس های طبقه ای و ترتیبی را در اختیار می گذارد، اطلاعات اضافی دیگری را نیز در مورد تفاوت ها بدست می دهد.

ویژگی دیگر مقیاس فاصله ای آن است که نقاط مختلف آن نه تنها به ترتیب کمی و زیادی قرار گرفته بلکه واحد اندازه گیری آن در سراسر طول مقیاس یکسان و ثابت است. واحد اندازه گیری در مقیاس سانتی گراد، درجه است. چون درجه در طول مقیاس خود مقدار ثابتی از تفاوت را نشان می دهد، نتیجه می گیریم که تفاوت بین 10° و 20° دقیقاً برابر با تفاوت بین 20° و 30° است. یک ویژگی دیگر این مقیاس این است که صفر مطلق ندارد یعنی صفر درجه سانتی گراد به معنی نبودن دما نیست.

1.5 انواع مقیاس های اندازه گیری: مقیاس نسبتی^۳

این نوع متغیر دارای همه ویژگی های متغیر فاصله ای است و علاوه بر آن صفر مطلق نیز دارد. پیشرفت در ریاضی، در بهترین صورت خود، یک متغیر فاصله ای است. دانش آموزی که در قسمت ریاضی نمره صفر می گیرد نباید فرض شود که اصولاً فاقد دانش ریاضی است. اما از دگر سو، مقدار پول موجود در بانک، یک متغیر نسبتی است. اگر مقدار آن صفر باشد، واقعاً هیچ پولی در آن بانک موجود نیست. اعداد روی مقیاس نسبتی نه تنها رتبه اشیاء و تفاوت نسبی آنها را نشان می دهد بلکه نسبت افراد را به یکدیگر از لحاظ خصیصه مورد اندازه گیری نیز مشخص می کند. بنابراین عملیات تجربی مورد لزوم برای داشتن چنین مقیاسی نه تنها عملیات تعیین برابری و نابرابری رتبه ها و فاصله های (بسا واحدهای مساوی) را در بر می گیرد، بلکه مستلزم برابر بودن نسبت ها نیز هست. اندازه گیری در چنین سطحی بر این مفروضه متکی است که

¹ Rank order correlation

² Interval scale variables

³ Ration scale variables

مقیاس، نقطه صفر مطلق دارد و از نقطه مزبور به بعد نیز واحدهای آن برابر است. چون نسبت بدون صفر مطلق مفهوم و معنایی ندارد لذا تنها با استفاده از مقیاس نسبی است که می توان حکم کرد: ایرج دو برابر هوشنگ پول دارد، یا ایرج نصف مهرداد وزن دارد. در واقع همه روابط بین اعداد در نظام قراردادی ریاضیات درباره نقاط مختلف این مقیاس صادق است. وقتی يك صفت از شیئی در سطح این مقیاس اندازه گیری می شود می توان این صفت را در شیئی مزبور به صورت عدد مطلق (بدون اطلاع از نمره اشیاء دیگر) نشان داد.

➤ آزمون

- 1- کدام يك از موارد زیر صحیح نیست؟
الف) اندازه گیری فرآیندی است که تعیین می کند يك شخص یا يك شيء چه مقدار از يك ویژگی برخوردار است.
ب) یکی از ویژگی های متغیر قابلیت اندازه گیری آن است.
ج) مقیاس های اندازه گیری به 4 نوع یا 4 سطح تقسیم می شوند.
د) در اندازه گیری از قضاوت ارزشی استفاده می شود.
- 2- در خصوص مقیاس های اندازه گیری کدامیک از موارد زیر صحیح است؟
الف) وقتی از ارقام برای اسم گذاری استفاده می شود از مفهوم ریاضی اعداد استفاده می کنیم.
ب) در مقیاس نسبی صفر مطلق وجود ندارد.
ج) در مقیاس ترتیبی سلسله مراتب اشیاء یا افراد یا رویدادها مشخص می شود.
د) مقیاس فاصله ای دارای همه ویژگی های مقیاس نسبی است.

2 اصول و مبانی آزمون ها، آزمودن و سنجش

➤ مفاهیم کلیدی

- 1- آزمون روشی نظام دار برای اندازه گیری نمونه ای از رفتار است.
- 2- سنجش کلاسی تمامی فرآیندهای مورد نیاز برای تصمیم گیری درباره یادگیری دانش آموزان است.
- 3- اندازه گیری و آزمودن با کمیت سروکار دارند اما سنجش الزاماً به کمیت منحصر نمی شود.

➤ اهداف یادگیری

شما پس از مطالعه این نوشتار باید بتوانید:

- 1- آزمون را تعریف کنید.
- 2- سنجش را تعریف کنید.
- 3- تفاوت های سنجش و اندازه گیری و آزمودن را شرح دهید.

2.1 تعریف آزمون^۱

بر خلاف ویژگی های فیزیکی (جسمی) و رفتارهای آشکار که به طور مستقیم اندازه گیری می شوند، ویژگی ها یا صفات روانی مانند هوش، خلاقیت، انگیزش، نگرش، یادگیری و مانند اینها را باید به طور غیر مستقیم اندازه گیری کرد، زیرا هیچ کس به طور مستقیم به این ویژگی ها دسترسی ندارد. معمول ترین وسیله اندازه گیری ویژگی ها یا صفات روانی **آزمون** است.

از آنجا که متخصص اندازه گیری به ویژگی روانی مورد نظرش دسترسی مستقیم ندارد تا مانند یک ویژگی جسمی یا یک رفتار آشکار آن را مستقیماً اندازه بگیرد، ناچار است به رفتارهایی از فرد که به گمان او معرف ویژگی روانی مورد نظر هستند مراجعه کند و به اندازه گیری آنها بپردازد. مثلاً رفتارهایی از افراد را که معرف هوشمندی یا کم هوشی هستند اندازه می گیرد و بر اساس آنها تعیین می کند که نمره هوش هر کسی چقدر است، یا با پرسیدن سؤال هایی از دانش آموزان کلاس تعیین می کند که هر یک از آنها چه مقدار از دانش و مهارت های درسی را که قرار بوده است یاد بگیرد یاد گرفته است.

رفتارهای معرف ویژگی های روانی بسیار متنوع اند، و بنابراین نمی توان همه آنها را اندازه گیری کرد. مثلاً اگر بخواهیم تمام رفتارهایی را که معرف هوشمندی یا یادگیری یک موضوع درسی هستند فهرست کنیم در هیچ آزمونی نمی گنجند. بنابراین، متخصص اندازه گیری نمونه ای از رفتارهای مورد نظرش را انتخاب می کند و آزمون خود را بر پایه آن نمونه بنا می نهد. پس، بنا بر تعریف، «آزمون وسیله یا روشی نظامدار (یا نظام مند) برای اندازه گیری نمونه ای از رفتار است» (گرانلاند،^۲ لین،^۳ 1990). هر چه نمونه رفتارهای انتخاب شده و گنجانیده شده در آزمون بیشتر معرف ویژگی روانی مورد نظر باشد، به همان میزان نتایج اندازه گیری دقیق تر و قابل اعتمادترند. علاوه بر آزمون، روش ها و وسایل دیگری برای اندازه گیری ویژگی های روانی وجود دارند. با این حال، معروف ترین وسیله اندازه گیری در آموزش و پرورش و روانشناسی، آزمون است که به صورت مجموعه ای سؤال و غالباً به طور کتبی برای پاسخگویی در اختیار آزمون شونده قرار داده می شود.

2.2 تعریف آزمودن

نیتکو (2001) در تعریف آزمون گفته است، «آزمون به عنوان یک ابزار یا یک شیوه نظامدار مشاهده و توصیف ویژگی های دانش آموزان که یا یک مقیاس عددی یا یک طرح طبقه بندی را به کار می بندد تعریف می شود» وی اضافه کرده است که در آموزشگاه منظور از آزمون یک ابزار مداد و کاغذی^۴ (کتبی) شامل تعدادی سؤال است که دانش آموزان به آن پاسخ می دهند. این نوع آزمون ها معمولاً از راه جمع کردن نمره

1 Test

2 Gronlund

3 Linn

4 Paper-and-pencil

هایی که دانش آموزان در هر يك از سؤال ها به دست می آورند نمره گذاری می شوند، یعنی اینها از مقیاس عددي استفاده می کنند. بنا به تعریف، وقتی که برای اندازه گیری یکی از ویژگی های روانی یا تربیتی يك فرد یا گروهی از افراد از آزمون استفاده می شود به این فعالیت آزمودن^۱ می گویند. در تعریف آزمون گفتیم که آزمون نمونه ای از رفتارهای مورد نظر را اندازه می گیرد. از این رو، «اندازه گیری متغیرهای روانشناختی به وسیله تدابیر یا روش هایی که به منظور کسب نمونه هایی از رفتار طراحی شده اند آزمودن روانی یا روان آزمایی نام دارد» (کوهن و سوردلیک، 2002). از آنجا که آزمون يك وسیله اندازه گیری است، هر وقت آزمودن صورت می پذیرد یعنی آزمونی به کار می رود نوعی اندازه گیری انجام می شود، اما هر نوع اندازه گیری الزاماً آزمودن نیست. اندازه گیری از آزمودن مفهوم گسترده تري دارد، زیرا آزمودن نوع خاصی از اندازه گیری است. ممکن است اندازه گیری صورت پذیرد اما آزمونی در کار نباشد؛ یعنی اندازه گیری می تواند با استفاده از وسیله دیگری بجز آزمون انجام شود، مانند زمانی که يك معلم، از راه مشاهده و با شمردن تعداد دفعاتی که يك دانش آموز هم کلاسی کنار دستی اش را آزار می دهد، رفتار پرخاشگرانه او را اندازه می گیرد.

2.3 تعریف سنجش

اصطلاح دیگر نزدیک به اندازه گیری و آزمودن اصطلاح سنجش^۲ است. سنجش از آزمودن و اندازه گیری مفهوم گسترده تري دارد. بنا به گفته نیتکو (2001) «سنجش يك اصطلاح كلي است و به صورت فرایندی تعریف می شود که برای گردآوری اطلاعات مورد نیاز تصمیم گیری درباره دانش آموزان، برنامه های درسی، و سیاست های آموزشی مورد استفاده قرار می گیرد.» اگن^۳ و کاوچاک^۴ (2001) **سنجش کلاسی^۵** را به عنوان «تمامی فرآیندهای مورد نیاز برای تصمیم گیری درباره یادگیری دانش آموزان» تعریف کرده اند.

چنان که دیدیم، وقتی که اندازه گیری صرفاً با استفاده از آزمون انجام می شود آزمودن صورت می پذیرد، اما در سنجش وسایل و فنون مختلفی برای جمع آوری اطلاعات به کار می روند. از جمله آنها می توان به آزمون، پرسشنامه، مقیاس درجه بندی، فهرست و ارسبی، کارآزمایشگاهی، پروژه تحقیقی، امتحان شفاهی، تکلیف درسی، مصاحبه، و مشاهده عملکرد و رفتار دانش آموزان در موقعیت های مختلف اشاره کرد. به عنوان نمونه، سانتروک^۶ (2001) گفته است «مشاهدات غیررسمی می تواند درباره میزان انگیزش دانش آموزان نسبت به یادگیری موضوع های درسی اطلاعات مفیدی به شما بدهد. همچنین اگر رابطه خوبی با دانش آموزانتان داشته باشید پرسش های مستقیم از آنان در يك گفتگوی

¹ Testing

² Assessment

³ Eggen

⁴ Kauchak

⁵ Classroom assessment

⁶ Santroc

خصوصی شما را از چگونگی نگرش آنان نسبت به درس و مدرسه آگاه خواهد ساخت»

ویگینز^۱ (1993) در توضیح اصطلاح سنجش گفته است، «سنجش به یک تحلیل جامع و چند وجهی از عملکرد گفته می‌شود» او همچنین در مقایسه سنجش با آزمودن از قول کرانباخ^۲ نقل کرده است که «سنجش شامل استفاده از فنون مختلف است و تأکید زیادی بر مشاهده عملکرد دارد سنجش نوعی تحلیل بالینی و پیش بینی عملکرد است، در حالی که آزمون یک ابزار است، یک وسیله اندازه گیری است»

اصطلاح سنجش بیشتر در حوزه روانشناسی بالینی و آموزش و پرورش به کار می‌رود، زیرا در روانشناسی بالینی و آموزش و پرورش، علاوه بر آزمون و پرسشنامه، از وسایل و روش‌های مختلف دیگری برای اندازه گیری رفتارها و ویژگی‌های روانی و تربیتی افراد استفاده می‌شود، مانند مصاحبه، ایفای نقش، استفاده از دستگاه‌های روانی-فیزیولوژی، و مانند اینها. آرنزد^۳ (1994) در توصیف سنجش در آموزش و پرورش مطالب زیر را بیان داشته است:

اصطلاح سنجش معمولاً به مجموعه کاملی از اطلاعات که توسط معلمان درباره دانش آموزان و کلاسهایشان جمع آوری و دسته بندی می‌شوند اشاره می‌کند. اطلاعات مربوط به دانش آموزان را می‌توان به طور غیر رسمی، مثلاً از راه مشاهده یا مبادله کلامی به دست آورد. همچنین اینگونه اطلاعات را می‌توان به صورت رسمی، مثلاً از طریق دادن تکالیف، آزمون‌های مختلف، و گزارش‌های کتبی کسب کرد. افزودن بر اینها، اطلاعات مربوط به کلاس‌های درس و آموزش معلمان نیز می‌تواند بخشی از سنجش به حساب آید.

سالویا و یسل دایک (1991) می‌گویند سنجش تنها به جمع آوری داده‌ها یا اطلاعات خلاصه نمی‌شود، بلکه به جمع آوری داده‌ها همراه با قصد یا منظور گفته می‌شود. به عقیده این متخصصان، سنجش در آموزش و پرورش به دو منظور انجام می‌گیرد: (1) تعیین و تشخیص مشکلات دانش آموزان، (2) تصمیم‌گیری درباره آنان. تشخیص مشکل در سه زمینه تحصیلی، رفتاری، و جسمی انجام می‌شود. تصمیم‌گیری با هدف‌های ارجاع دادن دانش آموزان مشکل‌آفرین یا ناسازگار به متخصصان، جدا کردن دانش آموزان نیازمند کمک‌های ویژه، گروه بندی دانش آموزان برای مقاصد آموزشی، طراحی و آموزشی، و بالاخره تعیین میزان پیشرفت تحصیلی دانش آموزان صورت می‌پذیرد.

2.4 تفاوت سنجش و اندازه گیری و آزمودن

تفاوت دیگر میان سنجش و اندازه گیری و آزمودن این است که هم اندازه گیری و هم آزمودن عموماً با کمیت سروکار دارند، اما سنجش الزاماً به کمیت منحصر نمی‌شود. به سخن دیگر، زمانی که اندازه گیری یا آزمودن صورت می‌پذیرد، نتیجه به صورت اعداد یا ارقام (کمیت) نشان داده می‌شود، اما نتیجه سنجش می‌تواند به صورت غیر کمی

¹ Wiggins

² Cronbach

³ Arends

(توصیفی) گزارش شود. به عنوان مثال، وقتی که، به کمک یک آزمون ریاضی، یادگیری یک دانش آموز مشخص می شود، نتیجه آن در قالب یک نمره که جنبه کمی دارد تعیین می گردد، اما می توان رفتار اجتماعی یک دانش آموز را در کلاس درس با عنوان های پرخاشگر، سازگار، و مانند اینها توصیف کرد که جنبه کیفی دارند.

➤ آزمون

- 1- کدام یک از موارد زیر در خصوص سنجش صحیح نیست؟
الف) سنجش الزاماً به کمیت منحصر نمی شود و نتیجه آن می تواند بصورت غیر کمی بیان شود.
ب) سنجش کلاسی به تمام فرآیندهای مورد نیاز برای تصمیم گیری در خصوص یادگیری دانش آموزان اطلاق می شود.
ج) در سنجش وسایل و فنون مختلفی برای جمع آوری اطلاعات بکار می رود.
د) هیچکدام

- 2- کدام یک از جملات زیر صحیح نیست؟
الف) معروف ترین وسیله اندازه گیری در آموزش و پرورش و روانشناسی آزمون است.
ب) آزمون روشی نظامدار جهت اندازه گیری نمونه ای از رفتار است.
ج) آزمون از اندازه گیری مفهوم گسترده تری دارد.
د) اندازه گیری متغیرهای روانشناختی، آزمون روانی یا روان آزمایی نام دارد.

3 اصول و مفاهیم ارزشیابی

➤ مفاهیم کلیدی

- 1- ارزشیابی به یک فرآیند نظام دار برای جمع آوری، تحلیل و تفسیر اطلاعات گفته می شود به این منظور که تعیین شود آیا هدف های مورد نظر تحقق یافته اند یا به چه میزانی؟
- 2- ارزشیابی پیشرفت تحصیلی مفهومی بسیار نزدیک به مفهوم سنجش یادگیری است اما ارزشیابی آموزشی در مفهوم کلی آن معنایی بسیار وسیع دارد.

➤ اهداف یادگیری

- شما پس از مطالعه این نوشتار باید بتوانید:
- 1- ارزشیابی را تعریف کنید.
 - 2- ارزشیابی پیشرفت تحصیلی را شرح دهید.
 - 3- تفاوت های اندازه گیری، آزمون، سنجش و ارزشیابی را فهرست کنید.

3.1 تعریف ارزشیابی و ارزشیابی پیشرفت تحصیلی

اصطلاح ارزشیابی^۱ (یا ارزیابی)، به طور ساده، به تعیین ارزش^۲ برای هر چیزی یا دایره ارزشی^۳ کردن گفته می‌شود. با این حال، تعریف جامع تری از ارزشیابی به شرح زیر می‌توان به دست داد: «ارزشیابی به یک فرایند نظام‌دار برای جمع‌آوری، تحلیل، و تفسیر اطلاعات گفته می‌شود به این منظور که تعیین شود آیا هدف‌های مورد نظر تحقق یافته‌اند یا در حال تحقق یافتن هستند و به چه میزان». یکی از ویژگی‌های مهم ارزشیابی تعیین **کیفیت**^۴ است. در فرآیند ارزشیابی دایره ارزشی با توجه به کیفیت به عمل می‌آید. نیتکو (2001) کیفیت را در ارزشیابی پیشرفت تحصیلی به عنوان دانش، مهارت‌ها و توانایی‌هایی که از دانش‌آموزان پس از آموزش انتظار می‌رود تعریف کرده است. وی به عنوان مثالی از ارزشیابی پیشرفت تحصیلی گفته است فرض کنید شما دانش‌آموزی را در توانایی نوشتن، نسبت به پایه تحصیلی‌اش، بسیار خوب ارزشیابی کرده‌اید. به منظور انجام این ارزشیابی شما توانایی نوشتن او را سنجش کرده‌اید. برای این کار، مجموعه نوشته‌های دانش‌آموز را مرور کرده‌اید، و نوشته‌های او را با نوشته‌های سایر دانش‌آموزان هم‌کلاسی‌اش و با معیارهای دیگری که برای نوشتن در سطح پایه این دانش‌آموز می‌شناسید مقایسه کرده‌اید (تعیین کیفیت). سرانجام دایره شما از توانایی نوشتن این دانش‌آموز به این نتیجه‌گیری انجامیده است که نوشته‌های او از کیفیت خوبی برخوردارند. پس توصیه می‌کنید که این دانش‌آموز می‌تواند در یک مسابقه ملی مربوط به مقاله نویسی شرکت کند. بنابراین، ارزشیابی مبنای تصمیم‌گیری برای فعالیت‌های عملی و اجرایی را تشکیل می‌دهد. مثال بالا در رابطه با ارزشیابی توانایی نوشتن نوعی ارزشیابی پیشرفت تحصیلی است که نمونه‌های دیگر آن می‌تواند ارزشیابی از پیشرفت دانش‌آموزان در درس‌های ریاضی، علوم، تاریخ و مانند اینها باشد. علاوه بر پیشرفت تحصیلی دانش‌آموزان، ارزشیابی آموزشی با موضوع‌های دیگری نیز سروکار دارد. از جمله آنها می‌توان عملکرد معلمان و مدیران، روش‌های آموزشی، برنامه‌های درسی، دوره‌های آموزشی، مواد آموزشی، پروژه‌های آموزشی، گروه‌های آموزشی و سازمان‌های آموزشی را نام برد. اینگونه ارزشیابی‌های آموزشی از ارزشیابی پیشرفت تحصیلی که در آن صرفاً موفقیت‌ها و دستاوردهای یادگیری دانش‌آموزان و دانشجویان مورد نظر است معنی گسترده‌تری دارند. **ارزشیابی پیشرفت تحصیلی**^۵ مفهومی بسیار نزدیک به مفهوم سنجش یادگیری است، اما **ارزشیابی آموزشی**^۶ در مفهوم کلی آن معنایی بسیار وسیع دارد.

1 Evaluation

2 Value

3 Value judgement

4 Quality

5 Achievement evaluation

6 Educational evaluation

ورتن^۱ و سندرز^۲ (1987) دو تن از صاحب نظران در حوزه ارزشیابی آموزشی گفته اند: «در آموزش و پرورش، ارزشیابی به یک فعالیت رسمی گفته می شود که برای تعیین کیفیت، اثربخشی، یا ارزش یک برنامه، فرآورده، پروژه، فرآیند، هدف یا برنامه درسی به اجرا در می آید.» این متخصصان ارزشیابی آموزشی برای روشن ساختن تعریف خود مثال زیر را به دست داده اند: یک ارزشیاب برنامه درسی^۳ ابتدا هدف های کلی یا غایت های^۴ برنامه درسی را تعیین می کند و بعد، با استفاده از درون داد به دست آمده از گروه های مرجع مربوط، تعیین می کند که این هدف ها برای دانش آموزان، پدران و مادران و جامعه خوب هستند یا نه. پس از آن، ارزشیاب اطلاعات لازم را برای ارزشیابی از میزان تحقق یافتن آن هدف ها و آثار جانبی ناشی از برنامه درسی جمع آوری می کند. سپس به تحلیل و تفسیر اطلاعات به دست آمده می پردازد. و وقتی که این کار نیز انجام گرفت، آنگاه ارزشیاب درباره ارزش برنامه درسی قضاوت یا داوری می کند و معمولاً به فرد یا افراد مسئول در تصمیم گیری های نهایی پیشنهادهایی ارائه می دهد.

3.2 مقایسه اندازه گیری، آزمودن، سنجش و ارزشیابی با یکدیگر

چنان که دیدیم، ارزشیابی شامل داوری ارزشی درباره کیفیت موضوع مورد ارزشیابی است که مبتنی بر تحلیل و تفسیر اطلاعات جمع آوری شده است. اطلاعات مورد نیاز ارزشیابی آموزشی به کمک اندازه گیری (عموماً آزمودن) و سنجش گردآوری می شوند، و هر چه اطلاعات گردآوری شده جامع تر و کامل تر باشد ارزشیابی درست تری صورت می پذیرد. به مثال زیر توجه کنید: فرض کنید معلمی می خواهد میزان اثربخشی روش آموزشی خود را در درس علوم با توجه به مقدار یادگیری دانش آموزانش ارزشیابی کند. به این منظور، تعدادی سؤال از درس علوم تهیه می کند و در اختیار دانش آموزان قرار می دهد تا به آنها جواب دهند. بعد جواب های دانش آموزان را تصحیح می کند و با شمارش تعداد سؤال هایی که دانش آموزان درست جواب داده اند به آنان نمره می دهد. سپس بر پایه نمرات به دست آمده دانش آموزان و انتظاراتی که از آنان می رفته است درباره مقدار یادگیری آنان و میزان توفیق خودش به داوری می پردازد.

در مثال بالا، مجموعه سؤال هایی که به دانش آموزان داده می شود آزمون به حساب می آید. اجرای آزمون و نمره گذاری آن آزمودن یا همان اندازه گیری به کمک آزمون نام دارد که در آن تبدیل نتایج آزمون به اعداد طبق یک قاعده مشخص (شمارش پاسخ های درست دانش آموزان) صورت پذیرفته است. و سرانجام داوری درباره نتایج آزمون یا نتایج اندازه گیری ارزشیابی به شمار می آید. ناگفته پیداست که اگر نتایج آزمون مورد استفاده در مثال بالا مخدوش باشند. ارزشیابی انجام شده نیز قابل اعتماد نخواهد بود. بنابراین ارزشیابی های درست بر پایه اطلاعات دقیق استوارند.

1 Worthen

2 Sanders

3 Curriculum evaluator

4 Goals

البته معلم می‌توانست، برای انجام ارزشیابی دقیق‌تری از میزان توفیق دانش‌آموزان در یادگیری و موفقیت خودش در آموزش، علاوه بر آزمون، از راه‌های دیگری نیز به جمع‌آوری اطلاعات بپردازد: مثلاً از دانش‌آموزان بخواهد تا به صورت فردی یا در گروه‌های کوچک یک پروژه تحقیقی انجام دهند، در آزمایشگاه به آزمایش‌های علمی بپردازند و نتایج کار خود را به صورت کتبی به معلم گزارش نمایند، یا به هنگام اجرای آزمایش در آزمایشگاه عملکرد دانش‌آموزان را مشاهده کند و به کمک یک فهرست و ارسلی (چک لیست) یا مقیاس درجه بندی یادگیری آنان را تعیین کند. مجموعه این اقدامات به اضافه اجرای آزمون علوم سنجش یادگیری به حساب می‌آید که مبنای محکم‌تری را برای ارزشیابی فراهم می‌آورد. بنابراین، ارزشیابی آموزشی می‌تواند بر اساس شواهد و داده‌های گوناگون صورت پذیرد. بسته به شرایط و موقعیت‌هایی که با آن سروکار دارید، این شواهد و داده‌ها می‌توانند تنها کمی باشند، کمی و کیفی با هم باشند، یا حتی تنها کیفی باشند. نیتکو (2001) در این باره گفته است:

شما می‌توانید دانش‌آموزان را بر اساس سنجش حاصل از مشاهدات نظام‌دار و توصیف‌های کیفی ارزشیابی کنید. بدون آنکه به اندازه‌گیری اقدام نمایید. حتی زمانی که اطلاعات عینی (نتایج اندازه‌گیری) در دسترس هستند و مورد استفاده قرار می‌گیرند، متخصصان ارزشیابی باید آنها را با تجارب و دیدگاه‌های خود درهم آمیزند و آنگاه اتخاذ تصمیم کنند. بنابراین، در همه ارزشیابی‌ها، میزانی از ذهنیت، ناهمخوانی، و سوگیری وجود دارد. آزمودن و اندازه‌گیری، از آنجا که از سایر روش‌های سنجش عینی‌تر هستند، پاره‌ای از ناهمخوانی‌ها و ذهنیت‌ها را کاهش می‌دهند. اما گفتن اینکه استفاده از آزمودن و اندازه‌گیری (یا به طور کلی فنون کمی اندازه‌گیری) ارزشیابی را خیلی زیاد بهبود می‌بخشد خود نوعی سوگیری نسبت به تکنولوژی است.

➤ آزمون

1- کدام یک از جملات زیر صحیح نیست؟
 الف) ارزشیابی به تعیین ارزش برای هر چیز گفته می‌شود.
 ب) یکی از ویژگی‌های مهم ارزشیابی تعیین کیفیت است.
 ج) ارزشیابی پیشرفت تحصیلی مفهوم وسیع‌تری به نسبت ارزشیابی آموزشی دارد.
 د) یک ارزشیابی برنامه درسی ابتدا هدف‌های کلی برنامه خود را مشخص می‌کند.

2- کدام یک از جملات زیر در مقایسه اندازه‌گیری، آزمودن، سنجش و ارزشیابی صحیح است؟
 الف) اطلاعات مورد نیاز ارزشیابی آموزشی همیشه از طریق اندازه‌گیری گردآوری می‌شوند.
 ب) شواهد و داده‌های ارزشیابی فقط کمی هستند.
 ج) در همه ارزشیابی‌ها میزانی از سوگیری وجود دارد.
 د) داده‌های ارزشیابی در بیشتر مواقع کیفی هستند.

4 دسته بندی آزمون های مورد استفاده

➤ مفاهیم کلیدی

- 1- آزمون ها با توجه به ویژگی های مورد اندازه گیری به دو دسته آزمون های توانایی و آزمون های عاطفی تقسیم می شوند.
- 2- آزمون های توانایی به دو دسته آزمون های استعداد و آزمون های پیشرفت تحصیلی تقسیم می شوند.
- 3- از مهمترین اقسام آزمون های عاطفی، آزمون های شخصیت، آزمون های نگرش و آزمون های علاقه هستند.
- 4- آزمون ها از نظر چگونگی تهیه آنها به دو دسته آزمون های استاندارد شده و آزمون های معلم ساخته تقسیم می شوند.

➤ اهداف یادگیری

شما پس از مطالعه این نوشتار باید بتوانید:

- 1- تقسیم بندی آزمون ها را بر اساس ویژگی های مورد اندازه گیری نام برید.
- 2- آزمون های توانایی را توضیح دهید.
- 3- آزمون های عاطفی را شرح دهید.
- 4- آزمون های استاندارد و معلم ساخته را توضیح دهید.

4.1 دسته بندی آزمون ها با توجه به ویژگی های مورد اندازه

گیری: آزمون های توانایی

آزمون های توانایی، آموخته ها، مهارت ها و استعداد های افراد را می سنجند. دسته ای از این آزمون ها، توانایی های شناختی¹ افراد را می سنجند و تعدادی دیگر توانایی های روانی- حرکتی² را اندازه می گیرند. هر یک از این دو دسته آزمون کاربرد خاص خود را دارد. آزمون های توانایی های شناختی بیشتر در زمینه های تحصیلی و فعالیت های فکری و ذهنی مورد استفاده قرار می گیرند، اما آزمون های روانی- حرکتی غالباً در موقعیت های صنعتی و نظامی و تربیت بدنی به کار می روند. صرف نظر از کاربردهای اختصاصی، این دو دسته آزمون، به طور عمده، برای تعیین جنبه های مختلف توانایی های افراد تهیه و اجرا می شوند. ویژگی مهم آزمون های توانایی (شناختی و روانی- حرکتی) این است که هنگام اجرای آنها از آزمون شونده خواسته می شود تا حداکثر سعی خود را به کار بندد تا بهترین نمره ممکن را به دست آورد.

آزمون های توانایی به دو دسته آزمون های **استعداد³** و آزمون های **پیشرفت تحصیلی⁴** تقسیم می شوند. آزمون های استعداد مقدار توانایی یا آمادگی فرد را برای انجام دادن کارهایی که در پیش دارد و همچنین ظرفیت او را برای یادگیری های مختلف می سنجند؛ اما آزمون های پیشرفت تحصیلی دانش و مهارت هایی را که فرد تا لحظه اجرای آزمون کسب کرده است اندازه می گیرند. به سخن دیگر، آزمون های استعداد ناظر به آینده اند و برای پیش بینی موفقیت های آتی فرد به

¹ Cognitive

² Psychomotor

³ Aptitude

⁴ Academic achievement

کار می روند، در حالی که آزمون های پیشرفت تحصیلی ناظر به گذشته اند و برای تعیین آموخته های قبلی فرد مورد استفاده قرار می گیرند. آزمون های هوش و استعداد های مختلف معرف آزمون های استعداد هستند و آزمون ها و امتحان هایی که برای موضوع های گوناگون درسی تهیه می شوند آزمون های پیشرفت تحصیلی نام دارند.

➤ تفاوت آزمون های استعداد و پیشرفت تحصیلی

آزمون های استعداد و پیشرفت تحصیلی را به روشنی نمی توان از یکدیگر متمایز کرد، زیرا آنچه را که فرد در گذشته آموخته است غالباً می توان برای پیش بینی موفقیت یادگیری او در آینده مورد استفاده قرار داد. در واقع یکی از عوامل مهم پیش بینی یادگیری فرد در آینده، یادگیری های قبلی او هستند. ترندایک^۱ و هیگن^۲ (1979) در کتاب معروف خود با عنوان اندازه گیری و ارزشیابی در روانشناسی و آموزش و پرورش، در رابطه با آزمون های استعداد و پیشرفت تحصیلی، به گونه زیر نظر داده اند:

باید به خاطر داشت که همه آزمون های توانایی، صرف نظر از نامی که دارند، توانایی های موجود فرد را اندازه می گیرند. ما نمی توانیم مستقیماً توانایی «ذاتی» یا «ارثی» کسی را اندازه گیری کنیم. آنچه را که ما می توانیم در فرد اندازه بگیریم توانایی او برای انجام دادن اعمال در زمان حال است... بنابراین، تمایز بین آزمون های استعداد و پیشرفت تحصیلی به قصد و منظوری مربوط می شود که نتایج آزمون برای آن قصد و منظور مورد استفاده قرار می گیرند، نه به ماهیت یا محتوای آن آزمون.

هر چند که این دو نوع آزمون از لحاظ ساخت و محتوا شباهت های زیادی با یکدیگر دارند، اما از آنجا که آنها غالباً برای مقاصد مختلف به کار می روند، دسته بندی آزمون های توانایی به استعداد و پیشرفت تحصیلی یک دسته بندی مفید است. چنان که قبلاً گفته شد، آزمون های استعداد را معمولاً برای پیش بینی موفقیت های آینده افراد به کار می بندند، در حالی که آزمون های پیشرفت تحصیلی را برای تعیین مقدار آموخته ها یا توانایی های موجود آزمون شوندگان مورد استفاده قرار می دهند.

آزمون های پیشرفت تحصیلی با موضوع یا محتوایی که اندازه می گیرند مشخص می گردند و در نتیجه با عنوان آن موضوع ها نامگذاری می شوند، مانند آزمون پیشرفت ریاضی، آزمون پیشرفت فیزیک، آزمون پیشرفت مهارت های فنی و جز اینها. اصطلاح آزمون های استعداد نیز به تعدادی آزمون با مقاصد مختلف اطلاق می شود. معروف ترین آزمون های استعداد آزمون های هوش^۳ یا استعداد کلی هستند. هدف اکثر آزمون های هوش کلی اندازه گیری آن دسته از توانایی های ذهنی است که تقریباً در همه انواع فعالیت های فکری و شناختی ضروری اند.

علاوه بر آزمون های هوش یا استعداد کلی که توانایی های عمومی فرد را اندازه می گیرند، تعداد دیگری آزمون با نام آزمون های

¹ Thorndike

² Hagen

³ Intelligence

استعدادهاي ویژه وجود دارند که استعدادهاي اختصاصي افراد را در زمينه هاي مختلف اندازه گيري مي کنند. به تعبیر دیگر، هدف از کاربرد آزمون هاي هوش کلي پيش بيني موفقیت هاي همه جانبه افراد است، در حالی که آزمون هاي استعدادهاي ویژه براي پيش بيني موفقیت یا شایستگی افراد در زمینه هاي فکري و عملي مشخص به کار مي روند. آزمون هاي استعدادهاي ویژه با توجه به زمینه اي که استعداد افراد در آن سنجیده مي شوند نام گذاري مي شوند، مانند آزمون استعداد رياضي، آزمون استعداد فني، آزمون استعداد موسيقي و غيره.

4.2 دسته بندي آزمون ها با توجه به ویژگی هاي مورد اندازه گيري: آزمون هاي عاطفي

آزمون هاي شناختي و رواني - حرکتی توانايي هاي فکري و عملي افراد را اندازه مي گیرند. اما آزمون هاي عاطفي ویژگی هاي عاطفي و شخصيتي افراد را مي سنجند. آزمون هاي توانايي حداکثر عملکرد آزمون شونده را اندازه مي گیرند؛ بدین معنی که نشان مي دهند آزمون شونده در بهترین شرایط قادر به انجام چه کاري است، در حالی که آزمون هاي عاطفي وضع موجود آزمون شونده را در موقعیت هاي طبیعی زندگی مي سنجند. یکی دیگر از وجوه اختلاف این دو دسته آزمون آن است که آزمون هاي توانايي به طور عمده براي مقاصد پيش بيني موفقیت هاي تحصيلي و شغلي مورد استفاده قرار مي گیرند، اما آزمون هاي عاطفي بیشتر به امور مشاوره در حل مسائل رواني، تربيتي و حرفه اي کمک مي کنند. یکی از معایب آزمون هاي عاطفي این است که این آزمون ها به اندازه آزمون هاي شناختي عيني و دقیق نیستند. به این دلیل بسیاری از روانشناسان کاربرد اصطلاح آزمون را براي این ابزارهاي اندازه گيري درست نمی دانند. با وجود این، به قول آیکن¹ (1985) «بسیاری از ابزارهاي اندازه گيري عاطفي داراي ویژگی هاي آزمون هاي خوب، از جمله روایی و پایایی، هستند.» بنابراین استفاده از اصطلاح آزمون براي ابزارهاي اندازه گيري شخصیت و سایر ویژگی هاي عاطفي کار اشتباهي نیست. آزمون هاي عاطفي داراي انواع زیادی هستند. ما در اینجا تنها به مهم ترین آنها یعنی آزمون هاي **شخصیت**، آزمون هاي **نگرش**، و آزمون هاي **علاقه** اشاره مي کنیم.

➤ آزمون هاي شخصیت

از شخصیت² تعریف هاي زیادی به عمل آمده است. در واقع مي توان شخصیت را به صورت مجموعه رفتارهاي فرد در موقعیت هاي اجتماعي تعریف کرد (ایبل³ 1979) بعضي روانشناسان، علاوه بر رفتارهاي آشکار، ویژگی هاي غیر رفتاري از جمله هوش، پیشرفت، قد و قامت، وضع سلامتی، کیفیت صدا و وضع ظاهر را به این تعریف شخصیت اضافه مي کنند. از این توصیف کلي شخصیت چنین بر مي آید که روش هاي سنجش شخصیت باید شامل متغیرهاي مختلف شناختي، رواني - حرکتی و عاطفي باشند. از جمله این متغیرها مي توان پیشرفت، هوش، استعدادهاي ویژه، علاقه،

¹ Aiken

² Personality

³ Ebel

نگرش، ارزش ها و غیره را نام برد. اما از آغاز نهضت روانسنجی تاکنون رسم بر این بوده است که کار اندازه گیری متغیرهایی چون هوش، استعداد، پیشرفت تحصیلی و بعضی ویژگی های دیگر به عهده آزمون های ویژه آنها گذاشته شود و آزمون های شخصیت به طور عمده به اندازه گیری متغیرهای انگیزش^۱، مزاج^۲، سازگاری^۳، منش^۴ و از این قبیل اختصاص یابد.

➤ آزمون های علاقه و نگرش

دو بخش دیگر شخصیت علاقه^۵ و نگرش^۶ هستند که به خاطر اهمیت زیادی که در امور مشاوره شغلی و تربیتی دارند معمولاً به طور جداگانه سنجش می شوند. هم علاقه و هم نگرش به دوست داشتن و دوست نداشتن های افراد اشاره می کنند و هر دو به رجحانهای فرد درباره فعالیت ها، نهادهای اجتماعی، یا گروه ها ارتباط دارند. هر دو اینها شامل احساس های شخصی نسبت به امور هستند. اما آنچه علاقه و نگرش را از هم مجزا می کند موضوع این علاقه ها یا بی علاقه گی هاست. «علاقه به احساس فرد نسبت به یک فعالیت اشاره می کند، در حالی که نگرش حاکی از احساس فرد نسبت به یک شیء، یک نهاد اجتماعی، یا یک گروه است.» (مهرنزی^۷ و لهمان^۸، 1984، ص 465) ساکس^۹ (1197) در مقایسه علاقه و نگرش با یکدیگر گفته است علاقه به رجحان یک فعالیت بر سایر فعالیت ها اشاره می کند؛ در مقابل، نگرش به معنی ترجیح دادن گروه ها، نهادها یا اشیا است.

برای سنجش علاقه ها و نگرش های افراد روش های مختلفی وجود دارند. معروف ترین وسیله اندازه گیری در این دو زمینه پرسشنامه های علاقه و نگرش است. پرسشنامه های علاقه جمله هایی درباره مشاغل و فعالیت های گوناگون را شامل می شوند. آزمودنی به هر یک از این سؤال ها بر حسب علاقه ای که به آن حرفه یا فعالیت دارد پاسخ می دهد. پرسشنامه های نگرش سنج نیز حاوی تعدادی پرسش درباره نهادهای اجتماعی، گروه های نژادی و قومی، یا مفاهیم مختلف هستند که در اختیار آزمون شونده گذاشته می شوند و او باید نظر خود را نسبت به آنها مشخص کند.

-
- 1 Motivation
 - 2 Temperament
 - 3 Adjustment
 - 4 Character
 - 5 Interest
 - 6 Attitude
 - 7 Mehrens
 - 8 Lehman
 - 9 Sax

4.3 دسته بندی آزمون ها با توجه به چگونگی تهیه آنها: آزمون های میزان شده

آزمون های میزان شده یا معیار شده^۱ آزمون هایی هستند که با استفاده از نمونه های بزرگ آزمون شوندگان و غالباً به وسیله مؤسسه ها یا بنگاه های آزمون سازی دولتی یا خصوصی و یا مراکز آموزشی و علمی و برای اندازه گیری پیشرفت تحصیلی یا هوش و استعداد گروه های بزرگ دانش آموزان و با مقاصد ارزشیابی از دوره های تحصیلی تهیه و توزیع می شوند. در این نوع آزمون ها، ملاک نسبی اندازه گیری مورد استفاده قرار می گیرد.

4.4 دسته بندی آزمون ها با توجه به چگونگی تهیه آنها: آزمون های معلم ساخته

آزمون های معلم ساخته آزمون هایی هستند که معلمان برای سنجش پیشرفت تحصیلی دانش آموزان خود تهیه می کنند و در طول دوره آموزش و یا در پایان دوره به اجرا در می آورند. این نوع آزمون ها غالباً برای تعیین میزان توفیق یادگیرندگان در رسیدن به هدف های دقیق آموزشی طرح ریزی می شوند و از این جهت معمولاً در آنها ملاک مطلق اندازه گیری به کار گرفته می شود. هر چند که ظاهر این دو نوع آزمون خیلی به هم شبیه است، بین آنها تفاوت هایی اساسی وجود دارد.

➤ آزمون

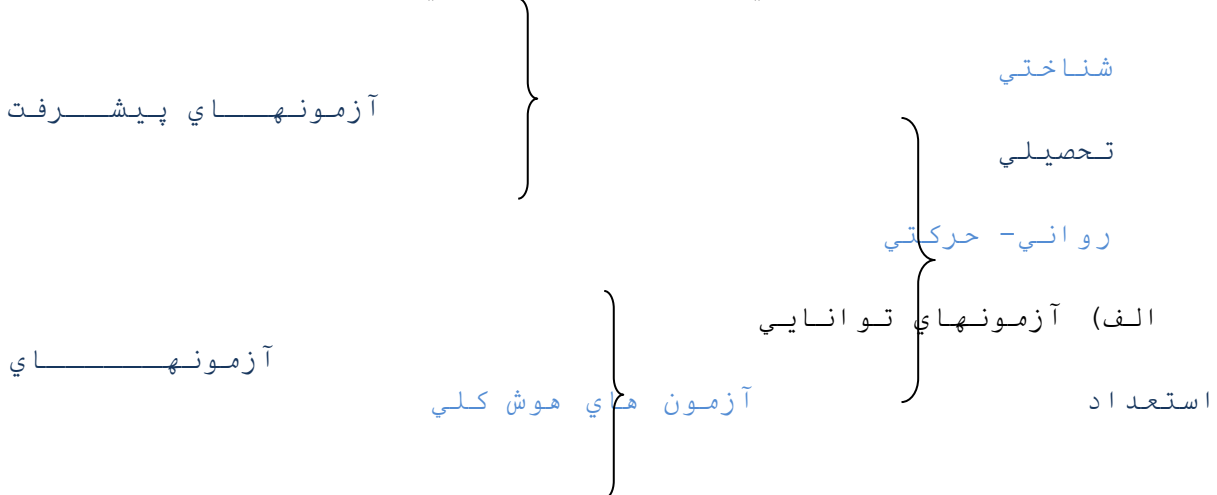
- 1- آزمون های معیار شده کدام نوع آزمون ها را شامل می شوند؟
 - الف) آزمون هایی که با نمونه های بزرگ آزمون شوندگان پس از استاندارد شدن تهیه شده اند.
 - ب) آزمون هایی که معلمان برای سنجش پیشرفت تحصیلی دانش آموزان خود تهیه می کنند.
 - ج) این آزمون ها نگرش و علاقه را می سنجند.
 - د) این آزمون ها همان آزمون های شخصیت هستند.
- 2- کدام يك از موارد زیر مهمترین انواع آزمون های عاطفی هستند؟
 - الف) آزمون های شخصیت
 - ب) آزمون های نگرش
 - ج) آزمون های علاقه
 - د) هر سه مورد

¹ Standardized

خلاصه

ما در این بحث اصطلاحات مهم حوزه اندازه گیری و ارزشیابی آموزشی شامل اندازه گیری، آزمون، سنجش، ارزشیابی را تعریف کردیم. سپس اصطلاحات و شباهتها و تفاوتهاي آنها را مورد بحث قرار دادیم. در مقایسه میان آزمودن و سنجش و اندازه گیری گفتیم که آزمودن نوعی اندازه گیری است که صرفاً به کمک آزمون صورت می پذیرد و فقط جنبه کمی دارد. اما در سنجش هم از آزمون و سایر وسایل تبدیل اطلاعات به کمیت استفاده می شود و هم از روشهای کیفی. مقیاسهای مهم اندازه گیری را نیز مورد بحث قرار دادیم. این مقیاسها عبارتند از مقیاس اسمی، فاصله ای و نسبی. سپس آزمونها را به ترتیب زیر تقسیم بندی کردیم:

1- با توجه به ویژگیهای مورد اندازه گیری



آزمون های استعدادهاي ویژه



پایایی

درس روشهای ارزشیابی
دانشجو جلسه

دوم
Validity

1388



فهرست مطالب

24	اهداف كلي	24
24	دسته بندی ارزشیابیهای آموزشی براساس موضوع ارزشیابی	1
24	1.1 ارزشیابی از دانشجویان	24
25	1.2 ارزشیابی از برنامه درسی و مواد آموزشی	25
25	1.3 ارزشیابی از مؤسسه آموزشی	25
25	1.4 ارزشیابی از کارکنان	25
26	2 دسته بندی ارزشیابی بر اساس ملاک مورد استفاده	26
27	2.1 تفسیر هنجار - مرجع	27
27	2.2 تفسیر ملاک - مرجع	27
28	2.3 وجوه تمایز آزمون های هنجار - مرجع - و ملاک - مرجع	28
29	3 دسته بندی ارزشیابی بر اساس زمان و هدف استفاده و ارزشیابان	29
30	3.1 ارزشیابی آغازین	30
30	3.2 ارزشیابی تکوینی	30
31	3.3 ارزشیابی تشخیصی :	31
31	3.4 ارزشیابی تراکمی :	31
32	3.5 افتراق ارزشیابی تکوینی از تراکمی	32
32	3.6 دسته بندی ارزشیابی با توجه به ارزشیابان	32
33	4 مراحل اساسی در ارزشیابی برنامه	33
34	4.1 طراحی	34
34	4.2 جمع آوری اطلاعات	34
34	4.3 خلاصه کردن اطلاعات	34
34	4.4 مقایسه با استاندارد (معیارهای اثر بخشی برنامه)	34
35	4.5 قضاوت و ارزش گذاری	35
	35 خلاصه	
36	منابع	36



شناسه جستار
 عنوان جستار: دسته بندی
 ارزشیابی ها:
 هنجار-مدار و معیار-مدار
 نویسنده: لیلیا بذرافکن
 عضو هیئت علمی دانشگاه علوم
 پزشکی شیراز
طراح آموزشی: دکتر مانوش
 مهربانی
 کارشناس آموزشی دانشگاه
 علوم پزشکی شیراز
ویرایش: مریم فخرزاد
 کارشناس مرکز مطالعات و
 توسعه آموزش علوم پزشکی
 تاریخ آخرین به روز
 رسانی: 1388/8/17
با همکاری:
 مرکز مطالعات و توسعه آموزش
 پزشکی
 قطب علمی آموزش الکترونیکی
 دانشگاه علوم پزشکی شیراز

اهداف کلی

- در تدوین این درس اهداف کلی زیر مد نظر بوده اند:
- 1- تقسیم بندی ارزشیابی بر اساس موضوع ارزشیابی
 - 2- تقسیم بندی ارزشیابی با توجه به ملاک مورد استفاده
 - 3- تقسیم بندی ارزشیابی با توجه به زمان و هدف ارزشیابی
 - 4- تقسیم بندی ارزشیابی با توجه به نظر ارزشیابان
 - 5- مراحل اساسی در ارزشیابی برنامه

دسته بندی ارزشیابیهای آموزشی براساس موضوع ارزشیابی

➤ مفاهیم کلیدی

- 1- برنامه های درسی و مواد آموزشی برنامه ها و پروژه های آموزشی ویژه نیز مورد ارزشیابی قرار می گیرند.
- 2- هدف از ارزشیابی تعیین میزان دستیابی به اهداف آموزشی دانشگاه و یا مؤسسه آموزشی و تعیین نقاط ضعف و قوت آن است.
- 3- ارزشیابی از کارکنان یعنی: ارزشیابی از همه کسانی که بطور مستقیم یا غیر مستقیم مسئول بازده آموزشی هستند.
- 4- یکی از روشهای تقسیم بندی انواع ارزشیابی ها با توجه به هدف یا موضوع ارزشیابی است. از این لحاظ ارزشیابی های مختلف را می توان در گروههای مختلف انجام داد.

➤ اهداف یادگیری

پس از مطالعه این قسمت شما می توانید:

- 1- انواع مختلف ارزشیابی بر اساس موضوع ارزشیابی را نام ببرید.
- 2- ملاک های مهم در مورد ارزشیابی از برنامه درسی را توضیح دهید.
- 3- تفاوت برنامه، پروژه و برنامه درسی را با ذکر مثال بیان نمایید.

ارزشیابی از دانشجویان

عمدتاً ارزشیابی از پیشرفت تحصیلی یا ارزشیابی از عملکرد و میزان یادگیری آنان است.

ارزشیابی از برنامه درسی و مواد آموزشی

به منظور تعیین اثر بخشی برنامه درسی یا مواد آموزشی انجام می شود

شامل ارزشیابی از عوامل مختلفی همانند روشهای آموزشی؛ کتابهای درسی؛ وسائل سمعی بصری؛ سازماندهی برنامه درسی و عوامل فیزیکی در رابطه با ارزشیابی از برنامه درسی ملاک مهم میزان یاد گیری دانشجویان است اما از ملاکهای دیگرمانند نگرش اساتید و دانشجویان و یا موضوع هزینه اثر بخشی برنامه نیز مهم هستند.

علاوه بر برنامه های درسی و مواد آموزشی برنامه ها و پروژه های آموزشی ویژه نیز مورد ارزشیابی قرار می گیرند. منظور از پروژه ها و برنامه های آموزشی ویژه کوشش هایی هستند موقعیتی که به عنوان بخشی از برنامه های معمول موسسه آموزشی به حساب نمی آید. بنابراین مفاهیم پروژه، برنامه و برنامه درسی از هم متمایز هستند. پروژه یک کوشش مجزا و احتمالاً یک باره است که با استفاده از یک نوآوری می کوشد تا تغییری به وجود آورد. این نوآوری می تواند به یک روش آموزش مهارت های بالینی یا یک رویکرد بهبود نگرش دانشجویان در مورد طب سرپایی مربوط باشد. بنابراین در ارزشیابی از یک پروژه آموزشی ارزشیابی باید به دنبال پاسخ این پرسش باشد که آیا پروژه مورد نظر از جهانی اثربخش بوده است یا خیر.

اگر جواب مثبت باشد آنگاه پروژه به صورت بخشی از برنامه در خواهد آمد. برنامه از پروژه گسترده تر و پیچیده تر است. یک پروژه خاص ممکن است یکی از چند نوآوری هایی باشد، که به منظور کمک به رفع مشکل دانشجویان در دوره آموزش بالینی کمک کند. در حالی که برنامه یک رویکرد چند وجهی برای حل مشکلات کلی نظام آموزشی است. برنامه درسی از پروژه و برنامه جامع تر است. برنامه درسی را به عنوان مجموعه ای از برنامه ها تعریف کرده اند.

ارزشیابی از مؤسسه آموزشی

هدف از آن تعیین میزان دستیابی به اهداف آموزشی دانشگاه و یا مؤسسه آموزشی و تعیین نقاط ضعف و قوت آن است. از ابزارهای گوناگونی مثل مصاحبه؛ آزمون؛ پرسشنامه و مشاهده می توان استفاده نمود.

ارزشیابی از کارکنان

ارزشیابی از همه کسانی که بطور مستقیم یا غیر مستقیم مسئول بازده آموزشی هستند.

شامل اساتید؛ مدیران؛ کارکنان اداری می شوند.

ارزیابی از اساتید می تواند به اشکال مختلف زیر انجام گیرد.

مشاهدات کلاسی از شیوه آموزشی استاد توسط مدیران و مسئولین.

نظر خواهی از دانشجویان از طریق پرسشنامه.

ارزشیابی اساتید از کار خود.

ارزشیابی از اساتید بوسیله همکاران.

بهترین و سالمترین روش ارزشیابی از اساتید استفاده از همه

روشهاست.

➤ آزمون

- 1- این نوع ارزشیابی مبتنی بر موضوع یا هدف است :
- الف- ارزشیابی از دانشجویان
ارزشیابی از مواد آموزشی
ج- ارزشیابی از اساتید
موارد
- ب- پروژ
- د- همه
- 2- تماس زودرس با بیمار يك :
- الف- برنامه درسي است.
است
- ج- برنامه آموزشی است
موارد
- ب- پروژ
- د- همه

دسته بندی ارزشیابی بر اساس ملاک مورد استفاده

➤ مفاهیم کلیدی

- 1- تفسیر هنجار - مرجع برای تفاوت گذاری بین افراد یا تمیز قایل شدن میان افراد يك گروه معین (تعریف شده) در زمینه هر چیزی که مورد اندازه گیری است مورد استفاده قرار می گیرد.
- 2- در تفسیر هنجار - مرجع، موقعیت فرد در گروه هنجاری مورد توجه است
- 3- ملاک یا معیار به معنای سطح معینی از عملکرد است که مستقل از اینکه میزان عملکرد دیگران چگونه می تواند باشد تعیین گردیده است.
- 4- هنگامی که ما اندازه گیری هنجار- مرجع¹ را با اندازه گیری ملاک- مرجع² مقایسه می کنیم اساساً به دو نوع تفسیر متفاوت از اطلاعات توجه کرده ایم. هر يك از دو نوع اندازه گیری فوق دارای ویژگی های معینی بوده، و نتایج آزمون های هنجار - مرجع نمی تواند با شیوه های ملاک - مرجع و یا به عکس مورد تفسیر قرار گیرد.
- 5- آزمون های هنجار- مرجع بیشتر کلی و جامع بوده و حیطة وسیعی از محتوا و تکالیف یادگیری را در بر دارند.
- 6- نمرات آزمون هنجار - مرجع به موقعیت هایی در گروه هنجاری تبدیل می شوند.
- 7- سؤالات آزمون هنجار - مرجع از نظر دشواری به سوی متوسط گرایش دارد.

➤ اهداف یادگیری

- پس از مطالعه مطالب این بخش شما می توانید:
- 1- انواع ارزشیابی از نوع ملاک مورد استفاده را شرح دهید.
- 2- ارزشیابی هنجار - مرجع را با ذکر مثال توضیح دهید.
- 3- ارزشیابی ملاک- مرجع را با ذکر مثال توضیح دهید.
- 4- ارزشیابی ملاک - مرجع را از هنجار - مرجع افتراق دهید.

¹ Norm- Referenced

² Criterion - Referenced

- 5- ارزشیابی تکوینی را از ارزشیابی تکمیلی افتراق دهید.
- 6- تفاوت آزمون هنجار- مرجع و معیار- مرجع را بیان نماید.
- 7- تفاوت سطح دشواری در آزمون های هنجار - مرجع و معیار- مرجع را با ذکر مثال توضیح دهد.

تفسیر هنجار - مرجع

تفسیر هنجار - مرجع برای تفاوت گذاری بین افراد یا تمیز قایل شدن میان افراد یک گروه معین (تعریف شده) در زمینه هر چیزی که مورد اندازه گیری است مورد استفاده قرار می گیرد. در اندازه گیری هنجار - مرجع، نمره فرد از طریق مقایسه آن با نمرات یک گروه معین که غالباً گروه هنجاری^۱ نامیده می شود مورد تفسیر قرار می گیرد. هنجارها یا نورمها نمایانگر نمراتی هستند که به وسیله یک یا چند گروه از دانشجویانی که در آزمون شرکت کرده اند تحصیل شده است.

تفسیر هنجار- مرجع یک تفسیر نسبی مبتنی بر موقعیت یک فرد در مقایسه با برخی گروه هاست که غالباً گروه هنجاری نامیده می شوند. هنجارها مرکب از نمراتی هستند که معمولاً به صورت آمار توصیفی از گروه هنجاری به دست آمده اند.

به این ترتیب این نوع از موقعیت یابی عملکرد فرد را به صورت مطلق مشخص نمی کند. هنجار مورد استفاده اساس مقایسه بوده و هر نمره ای از طریق موقعیتش در گروه هنجاری مشخص می شود.

تفسیر ملاک - مرجع

مفاهیم مربوط به آزمودن های ملاک - مرجع با معنای دوگانه ای از ملاک- مرجع گسترش یافته است. از طرفی، معنای آن ارتباط دادن عملکرد یک فرد به برخی ملاک هایی است که عبارت از یک سطح عملکردی تعریف شده است. نمره فرد در این حالت به صورت مطلق تفسیر می شود. در این موقعیت ملاک یا معیار به معنای سطح معینی از عملکرد است که مستقل از اینکه میزان عملکرد دیگران چگونه می تواند باشد تعیین گردیده است.

معنای دیگر ملاک - مرجع عبارت از اندیشه یک حیطه رفتاری معین می باشد، یعنی یک مجموعه تعریف شده از رفتارهای فراگیر. عملکرد فراگیر در آزمون به گروهی از رفتارها که مشخصاً تعریف شده اند ارتباط می یابد. معیار در این موقعیت عبارت از رفتارهای مطلوب است.

تفسیر ملاک- مرجع بیش از آن که تفسیری نسبی باشد تفسیری مطلق است که به مجموعه تعریف شده ای از رفتارهای فراگیر و یا همان طور که غالباً انجام می شود، به سطوحی معین از عملکرد فراگیر ارتباط

دارد.

آزمون های ملاک- مرجع مستلزم این است که رفتارهای فراگیر قبل از تهیه آزمون مشخص شود. رفتارها باید از هدف های آموزشی قابل

¹ Normative Group

شناسایی باشد؛ آزمون های ملاک - مرجع بیشتر بر رفتارهای خاص فراگیر متمرکز بوده، و معمولاً در هر آزمون تعداد محدودی از این رفتارها تحت پوشش قرار می گیرد.

وجوه تمایز آزمون های هنجار - مرجع- و ملاک - مرجع

اگر چه عامل تفسیر و نه ویژگی ها، فراهم آورنده تمایز بین آزمون های هنجار - مرجع و ملاک- مرجع می باشد، ولی این دو نوع آزمون از چند نظر با یکدیگر متفاوتند. آزمون های هنجار- مرجع بیشتر کلی و جامع بوده و حیطه وسیعی از محتوا و تکالیف یادگیری را در بر دارند. این آزمون ها برای آزمودن زمینه یابی¹ مورد استفاده اند. اگر چه این مورد استفاده کاربرد انحصاری آنها نیست. آزمون های هنجار- مرجع بر گروه خاصی از رفتارهای فراگیر متمرکزند.

نمرات آزمون هنجار - مرجع به موقعیت هایی در گروه هنجاری تبدیل می شوند. نمرات آزمون ملاک - مرجع معمولاً به صورت درصد پاسخ های درست و یا به صورت دیگری که بیانگر تسلط و یا عدم تسلط فرد مورد نظر باشد ارائه می گردد. آزمون های ملاک- مرجع بیشتر از آزمون های هنجار - مرجع به آموزش انفرادی مربوط می گردند. در آموزش انفرادی به جای مقایسه عملکرد دانش آموز با یک گروه، عملکرد او به طور اختصاصی تر از طریق مقایسه آن با رفتارهای مطلوب همان دانش آموز معین مورد تفسیر قرار می گیرد.

سؤالات آزمون هنجار - مرجع از نظر دشواری به سوی متوسط گرایش دارد، در حالی که آزمون های ملاک- مرجع متناسب با تکالیف یادگیری، سؤالات دشوار نیز خواهد داشت. چنین تفاوتی از نظر دشواری سؤال ضروری است. زیرا آزمون های هنجار- مرجع بر تمایز میان افراد، و آزمون های ملاک - مرجع بر توصیف عملکرد تأکید دارند. مثلاً سؤالات ساده نمی تواند میان افراد تمیز قایل شود ولی برای توصیف عملکرد ممکن است ضروری باشد.

تذکر نهایی در استفاده از آزمون های استاندارد هنجار - مرجع این است که: تفسیر نتایج هنگام اندازه گیری گرایش ها، علایق و استعداد تحصیلی بدون مقایسه آن با یک گروه مرجع غیر ممکن است. گروه های مرجع در چنین مواردی معمولاً دانش آموزان برجسته یا دانش آموزانی با علایق سطح بالا در زمینه های معین هستند. معلمان برای پیش بینی چنین نمراتی هیچ مبنای مقایسه قبلی ندارند، بنابراین برای استناد معنی به چنین نمره ای، یک گروه مرجع باید مورد استفاده قرار گیرد. به عنوان مثال، یک نمره 80 در یک پرسشنامه علایق به خودی خود هیچ معنایی ندارد. از طرف دیگر، اگر نمره 80 عبارت از پاسخی شاخص به وسیله یک گروه علاقه مند در زمینه های مکانیکی باشد، آن گاه نمره معنی دار می شود.

➤ آزمون

1- در این ارزشیابی بایستی ضریب تمیز سؤالات بالا باشد؟

¹ Survey Testing

- الف- هنجار- مرجع
ب- معیار- مرجع
ج- هر دو
د- هیچکدام
- 2- در این ارزشیابی عملکرد فرد به صورت اختصاصی تر مورد نظر است؟
الف- هنجار- مرجع
ب- معیار- مرجع
ج- تراکم
د- پایانی
- 3- آزمون کنکور دانشگاه ها یک آزمون
الف- ملاک مرجع است
ب- هنجار مرجع است
ج- هر دو
د- هیچکدام
- 4- تفسیر ملاک مرجع
الف- مطلق است
ب- نسبی است
ج- نسبی است
د- تمام موارد

دسته بندی ارزشیابی بر اساس زمان و هدف استفاده و ارزشیابان

➤ مفاهیم کلیدی

- 1- منظور کمک به اصلاح مورد ارزشیابی است و در طول دوره آموزشی یعنی زمانی که فعالیت آموزشی هنوز در جریان است.
- 2- برای بهره گیری کامل از نتایج ارزشیابی تکوینی بیان اهداف آموزشی به صورت دقیق و رفتاری ضروری است.
- 3- آزمون مورد استفاده در این روش، ملاکی یا وابسته ملاک مطلق است.
- 4- در ارزشیابی تراکمی تعداد سوالات زیاد و در یک زمان تعداد زیادی اهداف آموزشی را می سنجد.
- 5- ارزشیابی و آموزش به معنای سنجیدن ارزش حالات یا رویدادهای گوناگون، از کوچک تا بزرگ، از دقیق تا بسیار کلی است. نظام برنامه درسی همچنین شامل طرحی برای ارزشیابی برنامه درسی در ارتباط با هدفهای کلی آن است.
- 6- ارزشیابی آموزشی با توجه به زمان و هدف استفاده از آن شامل ارزشیابی آغازین، ارزشیابی تکوینی، ارزشیابی تشخیصی و ارزشیابی تراکمی می باشد.

➤ اهداف یادگیری

- پس از مطالعه این قسمت شما می توانید:
- 1- هدف از ارزشیابی تکوینی را شرح دهد.
 - 2- مراحل اجرای یک ارزشیابی تکوینی را توضیح دهد.
 - 3- اهداف ارزشیابی تراکمی را توضیح دهد.
 - 4- مخاطبان دو نوع ارزشیابی تکوینی و تکمیلی را با هم مقایسه نماید.

ارزشیابی آغازین

منظور از ارزشیابی آغازین آن دسته از ارزشیابی های آموزشی است که پیش از انجام فعالیت آموزشی جهت پاسخگویی به دو پرسش عمده زیر صورت می پذیرد:

1. آیا یادگیرندگان بر دانش و مهارتهای پیش نیاز درس تازه مسلط هستند؟
2. یادگیرندگان چه مقدار از اهداف و محتوایی درس تازه را قبلاً یاد گرفته اند؟

ارزشیابی تکوینی

منظور کمک به اصلاح مورد ارزشیابی است و در طول دوره آموزشی یعنی زمانی که فعالیت آموزشی هنوز در جریان است و یادگیری دانشجویان در حال تکوین یا شکل گیری است انجام می شود و منظور آن رفع مشکلات است. برای بهره گیری کامل از نتایج ارزشیابی تکوینی بیان اهداف آموزشی به صورت دقیق و رفتاری ضروری است. چون این نتایج به طور عمده برای پیشرفت یادگیرندگان و تعیین نواقص روش آموزش معلم بکار می رود لذا بهتر است از آنها به منظور نمره گذاری استفاده نشود. آزمون مورد استفاده در این روش، ملاکی یا وابسته ملاک مطلق است زیرا هدف تعیین میزان توفیق یادگیرنده در رسیدن به تک تک اهداف درسی است و هیچگونه مقایسه میان افراد مختلف مورد نظر نمی باشد.

سودمندیهای این نوع ارزشیابی عبارتند از پرداختن و جستجوی شواهدی دال بر ارزشمند بودن یک برنامه و دوم استفاده کردن از شواهد به دست آمده در جهت تصمیم گیری درمورد چگونگی ویرایش برنامه در خلال تولید آن است. زمانی که فرایند آموزش اعتقاد به استفاده از شیوه ارزشیابی تکوینی داشته باشد، می توان این سه مرحله را برای اجرای آن استفاده نمود این مراحل توسط دیک و کاری 1985 پیشنهاد شده است.

آزمون کردن يك به يك	آزمون کردن گروهی	آزمون کردن يك به يك
انواع اطلاعاتی که ضمن کسب می شود	انواع اطلاعاتی که ضمن این آزمون کسب می شود	انواع اطلاعاتی که ضمن این آزمون کسب می شود
اشتباه کردن در تخمین قابلیت های ورودی فراگیران	از طریق مقایسه نمرات پیش آزمون و پس آزمون به سوالات مربوط به روی دادن یادگیری و مقدار آن پاسخ داده می شود	انجام آزمون دقیق از آموزش
عدم وضوح ارائه آموزش	عملی بودن کاربرد و کارآیی	عملی بودن کاربرد و کارآیی
سوالها مبهم	بهبتر کردن ویرایش و درسهها	بهبتر کردن ویرایش و درسهها

		انتظارات نامناسب از بازده هاي يادگيري و نگرش هاي معلم و فراگير
--	--	---

ارزشيابي تشخيصي :

- نوعي از ارزشيابي تكويني است و هدف آن تشخيص مشكلات يادگيري دانشجويان در يك موضوع درسي است.
- مشخصات ارزشيابي تشخيصي :
- معمولاً در جريان آموزش انجام مي شود.
 - روشهاي مناسب جهت رفع مشكلات را نيز نشان مي دهد.
 - نه تنها معلم را از ضعف يك دانشجو آگاه مي كند بلكه زمينه هاي ضعف او را نيز نشان مي دهد.
 - روابط علي را نشان نمي دهد.
 - آزمون مورد استفاده در اين نوع نيز ملاكي يا وابسته به ملاك مطلق است.

ارزشيابي تراكمي :

تمامي آموخته هاي دانشجويان در طول يك دوره آموزشي تعيين مي شوند و هدف آن نمره دادن به دانشجويان و قضاوت درباره اثر بخشي كار معلم و برنامه درسي يا مقايسه برنامه هاي مختلف درسي با يكديگر است پس ارزشيابي در پايان دوره آموزشي انجام مي شود (ارزشيابي پاياني)

اهداف :

- نمره دادن و اتخاذ تصميم (ارتقاء دانشجو) دادن امتياز تحصيلي ، اعزام به خارج و بورس
- تعيين اثر بخشي معلم و ميزان توفيق او

آزمون مورد استفاده : سوالات مورد استفاده در آزمونهاي تكويني،

تشخيصي و تراكمي كمابيش يكسان است و تركيب اين سوالات و مقصد و منظور از کاربرد اين سوالات سبب تمايز آنها مي شود.

در آزمونهاي تكويني و تشخيصي تعداد سوالات كم و تمرکز بر اهداف خاص است. در ارزشيابي تراكمي تعداد سوالات زياد و در يك زمان تعداد زيادي اهداف آموزشي را مي سنجد.

قصد آزمون تكويني و تشخيصي، تشخيص مشكلات يادگيري و آموزشي در ضمن جريان يادگيري است.

ولي در ارزشيابي تراكمي تعيين آموخته هاي كلي فراگيران در پايان دوره قضاوت در باره كل برنامه طرح آموزشي مورد نظر است مخاطب و استفاده كننده در آزمون تكويني طراحان و توليد كنندگان برنامه آموزشي است ولي در آزمون تراكمي كساني كه برنامه را مورد

استفاده قرار می دهند یا فعلاً از آن استفاده می کنند مورد نظر هستند.

آزمون تکوینی به تصمیمات مربوط به توسعه و تولید برنامه کمک می کند و تصمیمات مربوط به ادامه، قطع یا گسترش برنامه را یاری می دهد.

افتراق ارزشیابی تکوینی از تراکمی

بهترین راه افتراق ارزشیابی تکوینی از تراکمی نظریه Bob stake است:

- آشپز سوپ را مزه می کند Formative
 - مهمان سوپ را مزه می کند Summative
- در یک دیدگاه کلی مقایسه بین ارزشیابی تکوینی و تراکمی در جدول زیر آمده است.

ارزشیابی تراکمی	ارزشیابی تکوینی
<p>هدفش ارزشیابی تمام برنامه پس از اجرای آن است امتیاز آن این است که بعنوان مبنایی جهت تجدید نظر در طرحهای برنامه درسی شناخته شده است</p> <p>این نوع ارزشیابی غالباً براساس آزمونهای همه جانبه است که از جمله عکس العمل فراگیرنسبت به آموزش، نظر معلم در رابطه با کارایی آموزش و....</p>	<p>هدف کمک به برنامه ریزان و تولید کنندگان (برنامه درسی، کتابها و ...) از طریق استفاده از تحقیق تجربی است.</p> <p>وسایله ای جدید برای تصمیمات منطقی و روا برای طراحان است</p> <p>یکی از ارزشیابی های این ارزشیابی تاکید بر اهمیت استفاده از فرایندهای ارزشیابی بعنوان بخشی از خود طراحی است.</p>

نهایتاً می توان چنین استنباط کرد که تفاوت این دو ارزشیابی بیشتر در مقاصد که به خاطرش ارزشیابی انجام شده و زمان استفاده از ارزشیابی است تا مواردی نظیر متدولوژی و یا تکنیکهای تحلیل آن.

دسته بندی ارزشیابی با توجه به ارزشیابان

بر دو گونه است: ارزشیابی درونی (internal) و ارزشیابی بیرونی (external)

- ارزشیابی درونی این است که ارزشیابی بوسیله دست اندرکاران تهیه و مجریان برنامه یعنی افراد درون برنامه انجام می شود.

- ارزشیابی بیرونی این است که ارزشیابی بوسیله اشخاص یا سازمانهایی خارج از برنامه انجام می شود.
- ارزشیابی درونی بیشتر مناسب ارزشیابی تکوینی است و ارزشیابی بیرونی مناسب ارزشیابی تراکمی است.
- ارزشیابی درونی خیلی بیشتر از بیرونی درباره برنامه اطلاع دارد اما به دلیل نزدیکی به برنامه نمی تواند بی طرف بماند.
- لذا ارزشیابی بیرونی از ارزشیابی درونی عینی تر و بی طرفانه تر است اما مشکل است که به اندازه ارزشیابان درونی از برنامه ها اطلاع داشته باشند.

➤ آزمون

- 1- در ارزشیابی تراکمی مهم نمی باشد.
 - الف- در پایان دوره انجام می شود. ب-
 - نمره دادن به دانشجویان هدف است. ج-
 - اثربخشی کار معلم هدف است. د- تشخیص مشکلات یادگیری هدف است.
- 2- اگر ارزشیابی از مراحل CIPP به قصد پاسخگویی انجام شود به آن ارزشیابی
 - الف- تراکمی گفته می شود. ب- تکوینی گفته می شود.
 - ج- تشخیصی گفته می شود. د-
 - توجهی گفته می شود.
- 3- ارزشیابی تکوینی :
 - الف- برای ارائه نمره دانشجویان است. ب-
 - برای ارائه بازخورد می باشد. د- ب و ج
 - ج- برای اصلاح یادگیری است.
- 4- ارزشیابی تکوینی یک نوع ارزشیابی
 - الف- ملاک مطلق است ب- ملاک نسبی است.
 - ج- هر دو د- هیچکدام

مراحل اساسی در ارزشیابی برنامه

➤ مفاهیم کلیدی

- 1- استانداردها باید از قبل تدوین شده باشند به نحوی که همه افراد درگیر در برنامه بدانند که به طرف چه شاخصهایی باید حرکت کنند و برنامه با چه شاخصهایی مورد ارزشیابی قرار می گیرد.
- 2- ارزشیابی اساساً مستلزم قضاوت است و قضاوت یعنی مقایسه اطلاعات جمع آوری شده در مقابل استانداردهای از قبل تعیین شده
- 3- در ارزشیابی برنامه باید مراحل پنجگانه زیر مورد توجه قرار گیرد: طراحی، جمع آوری اطلاعات، خلاصه کردن اطلاعات، مقایسه با استاندارد و قضاوت و ارزش گذاری.

➤ اهداف یادگیری

- 1- پس از مطالعه مطالب این بخش شما می توانید:
 - 1- مراحل ارزشیابی برنامه را شرح دهد.

- 2- حداقل 4 سوال مهم که در طراحی برنامه مطرح می شود را بیان نماید.
- 3- منابع جمع آوری اطلاعات را تفکیک نماید.
- 4- ویژگی های استاندارد در ارزشیابی برنامه را ذکر نماید.

طراحی

در طراحی ارزشیابی برنامه باید به سوالات زیر پاسخ گفت:

- 1- آیا برنامه قابل ارزشیابی است؟
- 2- آیا برنامه دارای اهداف روشن و فعالیت های تعریف شده است؟
- 3- آیا منابع کافی (مالی زمانی نیروی انسانی متخصص) برای ارزشیابی در دسترس است؟
- 4- چه نوع اطلاعات و داده هایی مورد نیاز است؟
- 5- چگونه به اطلاعات و داده ها دسترسی پیدا کنیم؟
- 6- چگونه از اطلاعات و داده ها استفاده کنیم؟

جمع آوری اطلاعات

اطلاعات را می توان از منابع اولیه و اصلی (مستندات اصلی و گزارشات دست اول و خام) یا از منابع ثانویه (مثل گزارشاتی که از منابع اولیه جمع بندی و جمع آوری می شوند) تهیه نمود. همچنین منابع اطلاعاتی می توانند متعدد و متنوع باشند و شامل مکاتبات افراد یا اسناد باشند.

در انتخاب روش مناسب جمع آوری اطلاعات باید به نکاتی مانند نوع اطلاعات مورد نیاز، زمان در دسترس هزینه و میزان تخصص و تجربه مورد نیاز توجه نمود.

خلاصه کردن اطلاعات

اطلاعات جمع آوری شده باید به شکل نمودارها یا جداول خلاصه شوند. در این مرحله صحت و تفسیر بدون سوگیری یافته ها کلید ارزشیابی موفق است.

مقایسه با استاندارد (معیارهای اثر بخشی برنامه)

اغلب افراد به جمع آوری مستندات و اطلاعات پرداخته و قضاوت درباره یک برنامه را بدون در دست داشتن استانداردهای دقیق و شفاف انجام می دهند درحالی که استانداردها باید از قبل تدوین شده باشند به نحوی که همه افراد درگیر در برنامه بدانند که به طرف چه شاخصهایی باید حرکت کنند و برنامه با چه شاخصهایی مورد ارزشیابی قرار می گیرد.

استانداردها نباید در ذهن ارزشیاب مخفی بماند باید قادر باشیم که بگوییم آیا به نقطه مطلوب رسیده و یا در راه رسیدن به نقطه مطلوب هستیم. استانداردها ارزشیابی را عینی و معنی دار می کند. به عنوان مثال استاندارد ما در درس ارزشیابی برنامه این است که صد در صد دانشجویان قادر باشند که ارزشیابی یک برنامه آموزشی را طراحی و اجرا کنند.

استاندارد باید دارای دو ویژگی باشد اول داشتن (نشانگری که قابل مشاهده بوده و نشاندهنده تغییر در یک چیز باشد) ثانیاً استاندارد باید میزان تغییر (مثلاً درصد تغییر) را نشان دهد.

قضاوت و ارزش گذاری

ارزشیابی اساساً مستلزم قضاوت است و قضاوت یعنی مقایسه اطلاعات جمع آوری شده در مقابل استانداردهای از قبل تعیین شده حاصل قضاوت در فرمهای ارزشیابی، ارائه و گزارش می شوند و این یافته ها با همه افراد ذینفع مثل ارزشیابی شوندگان، مدیران در سطوح مختلف به شکل کتبی، شفاهی، مقاله، روزنامه، نوار ویدئویی، گفتگوی رادیو و تلویزیونی و غیره ارائه می شود. به هر حال بسیاری معتقدند ارزشیابی دارای دو عنصر است هنر و علم.

هنر در ارزشیابی در گرو کار و تلاش برای دستیابی به یک توافق عمومی در مورد اهداف، طراحی نحوه جمع آوری اطلاعات در موقعیتهای خاص و ایجاد زمینه مناسب است. علم در ارزشیابی در گرو تعیین استانداردها، انتخاب روش مناسب جمع آوری اطلاعات به نحوی منظم و تفسیر اطلاعات برای ارزش گذاری و قضاوت در برنامه است.

➤ آزمون

1- این سوال که آیا برنامه قابل ارزشیابی است مربوط به کدام مرحله است؟

الف- جمع آوری اطلاعات

ب-

قضاوت

ج- طراحی

د- مقایسه با استاندارد

2- دو ویژگی اساسی استاندارد عبارت است از:

الف- نشان دهنده تغییر و میزان تغییر باشد. ب- از قبل تدوین شده باشد.

ج- الف و ب

د-

هیچکدام

خلاصه

به طور کلی ارزشیابی و آموزش به معنای سنجیدن ارزش حالات یا رویدادهای گوناگون، از کوچک تا بزرگ، از دقیق تا بسیار کلی است. نظام برنامه درسی همچنین شامل طرحی برای ارزشیابی برنامه درسی در ارتباط با هدفهای کلی آن است در این راستا ارزشیابی آموزشی با

توجه به زمان و هدف استفاده از آن شامل ارزشیابی آغازین Reassessment ارزشیابی تکوینی Formative ارزشیابی تشخیصی Diagnostic و ارزشیابی تراکمی Summative می باشد . منظور از ارزشیابی آغازین آن دسته است از ارزشیابی های آموزشی که پیش از انجام فعالیت آموزشی جهت پاسخگویی به دو پرسش عمده زیر صورت می پذیرد .

1. آیا یادگیرندگان بر دانش و مهارتهای پیش نیاز درس تازه مسلط هستند .

2. یادگیرندگان چه مقدار از اهداف و محتوای درس تازه را قبلاً یاد گرفته اند .

ارزشیابی تکوینی به منظور کشف مشکلات یادگیری و اصلاح آن است در حالیکه ارزشیابی تکمیلی یا پایانی به منظور نمره دادن و قضاوت در مورد اثربخشی کار معلم و برنامه درسی یا مقایسه برنامه های مختلف درسی بکار می رود . اگر ارزشیابی با توجه به ملاک از قبل تعیین شده و یا سطح معینی از عملکرد مستقل از عملکرد دیگران صورت گیرد معیار مدار و اگر ارزشیابی برای تمایز قایل شدن میان افراد یک گروه معین باشد، هنجار مدار می گویند .

منابع

1. سیف دکتر علی اکبر ، اندازه گیری ، سنجش و ارزشیابی آموزشی ویرایش سوم 1383 فصل هفدهم صفحه 447 تا 480
2. آلن مری جی، وندي ام ين. مقدمه اي بر نظريه هايي اندازه گيري ترجمه دکتر علي دلاور انتشارات سمت، چاپ اول پاییز 1374 .
3. رضایی دکتر حمید رضا، شوقی شفق آریا فرنگیس، دولتی ، دکتر یحیی، محقق دکتر محمد علی، ملک زاده ، دکتر رضا ، تکنیک های ارزیابی در پزشکی چاپ اول انتشارات آمه 1377 .
4. مهجور دکتر سیامک رضا- ارزشیابی آموزشی نظریه ها، مفاهیم ، اصول، الگوها. نشر ساسان چاپ 1376 .

5. Worthen , Blaine R. and James R. Sanders. Educational evaluation alternative approaches and practical guidelines .Newyork longman.1987.

پایایی


Validity

مرکز مطالعات و توسعه آموزش علوم پزشکی



فهرست مطالب

39	اهداف كلي	39
39	1 اصول و مباني روايي	39
	1.1 تعريف	40
	1.2 اهداف سنجش روايي	40
40	2 روايي محتوايي (اعتبار محتوا)	40
	2.1 تعريف	41
	2.2 جدول آزمون	41
	2.3 روشهاي تعيين روايي محتوايي	42
	2.4 روايي صوري	42
43	3 روايي ملاكي	43
	3.1 روايي ملاكي	43
	3.2 روايي پيش بيني	43
	3.3 مقايسه روايي پيش بين با همزمان	44
	3.4 ويژگيهاي مطلوب ملاك	44
45	4 روايي سازه	45
	4.1 تعريف	45
	4.2 روايي چند روشي - چند صفتي	46
	4.3 اعتبار عملي	48
	خلاصه 48	
49	منابع	49



شناسه جستار
عنوان جستار: روايي
نويسنده: دكتور مهراي
اميني
دانشيار دانشگاه علوم پزشکی
شیراز
طراح آموزشی: دكتور مانوش
مهرابي
كارشناس آموزشی دانشگاه
علوم پزشکی شیراز
ويرايش: مريم فخرزاد
آخرين تاريخ به روز

اهداف كلي

- در تدوين اين درس اهداف كلي زير مد نظر است:
- آشنايي با:
- تعريف روايي (اعتبار) آزمونها
 - مهمترين انواع روايي (روايي صوري ، محتوايي ، ملاكي ، سازه)
 - روشهاي تعيين روايي
 - عوامل دروني و بيروني مؤثر بر روايي
 - ترسيم جدول مشخصات آزمون

رساني: 5/ 1388/8

با همكاري :

مرکز مطالعات و توسعه آموزش پزشکی
قطب علمی آموزش الکترونیکی
دانشگاه علوم پزشکی شیراز

اصول و مباني روايي

➤ مفاهيم كليدي

- 1- آزمونهاي مورد استفاده بايد معتبر باشند.
- 2- يك آزمون در صورتي معتبر است كه چيزي را كه قرار است اندازه بگيرد، اندازه گيري كند.
- 3- روايي محتوايي ، ملاكي و سازه انواع مختلف روايي هستند .

➤ اهداف يادگيري

شما پس از مطالعه اين مطلب بايد بتوانيد:

- 1- روايي را تعريف كرده و مثال بزنيد.
- 2- اهداف سنجش اعتبار را فهرست نماييد.

تعريف

روايي اصطلاحي است كه به هدفی كه آزمون برای تحقق بخشیدن به آن درست شده است اشاره می کند. لین و گرانلاند (2000) روايي را به صورت يك ارزشيابي از كفايت و مناسبت تفسيرها و استفاده هاي نتايج سنجش تعريف كرده اند. كاپلان¹ و سوکوز² گفته اند روايي را می توان به صورت توافق بين نمره آزمون و كيفيتي كه قرار است اندازه بگيرد تعريف كرد.

بنابراين آنچه گفته شد آزموني دارايي روايي است (معتبر است) كه آنچه را در نظر دارد اندازه بگيرد. براي مثال آزموني كه براي انتخاب داوطلبان يك شغل به كار مي رود در صورتي داراي اعتبار است كه نمره آن با عملکرد آتي آزمودنيها در آن شغل رابطه قوي داشته باشد. يك آزمون معتبر هوش افراد مختلف را از لحاظ هوش از يكدیگر تفكيك مي كند. نمره هايي كه از يك آزمون معتبر شخصيت به دست مي آيد نشان دهنده تفاوتهاي معني داري در شخصيت است. سازندگان و

¹ - Kaplan

² - Saccuzzo

استفاده کنندگان آزمون هر دو مسئول تضمین اعتبار آزمونهای هستند که تهیه میشوند و به کار می روند.

اهداف سنجش روایی

آزمونهای مورد استفاده جهت سنجش دارای سه هدف عمده هستند و به سه پرسش زیر پاسخ می دهند.

1. آیا سوالهای آزمون از محتوای درس یا موضوع مورد نظر یک نمونه خوب ارائه می دهند؟

2. آیا نمرات آزمون عملکرد فعلی یا آتی دانشجو یا دانش آموز را پیش بینی می کنند؟

3. آیا نمرات آزمون به مفاهیم نظری یا سازه هایی که آزمون برای سنجش آنها درست شده است مربوطند؟

سه مورد بالا به ترتیب سه نوع روایی محتوایی، ملاکی و سازه را نشان می دهند تعیین اعتبار های وابسته به ملاک و سازه مستلزم محاسبه و بررسی همبستگیها و شاخصهای آماری دیگری است ولی در اعتبار محتوا از هیچ یک از روشهای آماری استفاده نمی شود.

➤ آزمون

1- کدامیک از موارد زیر **غلط** است ؟

الف : روایی اصلاحی است که به هدفی که آزمون برای تحقق بخشیدن به آن درست شده است اشاره می کند.

ب : آزمونی روا است که آنچه را در نظر دارد اندازه بگیرد اندازه گیری کند

ج : آزمونی رواست که قابل تکرار باشد

د : یک آزمون معتبر هوش افراد مختلف را از لحاظ هوش از یکدیگر تفکیک می کند.

2- کدامیک از موارد زیر انواع مهم روایی هستند ؟

الف : محتوایی

ج : سازه

ب : ملاکی

د : هر سه

روایی محتوایی¹ (اعتبار محتوا)

➤ مفاهیم کلیدی

1- روایی محتوایی یکی از انواع مهم روایی است.

2- پوشش دادن محتوا و مرتبط بودن سئوالات با محتوا در امتحانات ضروری است .

3- روایی محتوایی با نظر متخصصین تعیین می شود.

➤ اهداف یادگیری

شما پس از مطالعه این مطلب باید بتوانید:

1- روایی محتوایی را تعریف نمایید.

¹ - Content Validity

- 2- اهمیت روایی محتوایی در آزمونها را شرح دهید.
- 3- روش تعیین روایی محتوایی را بیان نمایید .
- 4- روایی صوری را توضیح دهید.

تعریف

روایی محتوایی به این مطلب اشاره می کند که نمونه سئوالهای مورد استفاده در يك آزمون تا چه حد معرف كل جامعه سئوالهاي ممكن است که می توان از محتوا یا موضوع مورد نظر تهیه کرد . هرچه آزمون از این لحاظ بهتر باشد دارای روایی بیشتری است اعتبار محتوا به تحلیل منطقی محتوایی يك آزمون بستگی دارد. اگر يك مدرس بخواهد برای درس ریاضی خود يك آزمون پیشرفت تحصیلی از نظر محتوا روا بسازد باید آزمون او نمونه درست و دقیقی از موارد ریاضی باشد که به دانشجویان درس داده شده است. از طرف دیگر این آزمون نباید چیزی خارج از محتوا و هدفهای درس ریاضی این مدرس را اندازه گیری کند. دو مفهوم در این زمینه اهمیت بسیاری دارد . پوشش دادن ، محتوا¹مرتبط بودن سوالات با محتوا² : در بحث پوشش دادن محتوا بیان می شود که از هر مبحث اساسی حتماً بایستی سئوال طرح شود و در بحث مرتبط بودن سوالات با محتوا این موضوع را طرح می کند که هر سوال امتحان باید با مبحثی که درس داده شده ارتباط داشته باشد و چیزی خارج از محتوای درس را نسنجد.

جدول آزمون

بهترین روش برای اینکه مشخص کنیم سئوالات آزمون نمونه کاملی از هدفها و محتوای درس است. تهیه جدول مشخصات آزمون است این جدول يك جدول دو بعدی است با تعدادی ردیف و ستون متناسب با محتوا و هدفهای آموزشی درس در ردیف بالای جدول اجزای مطالب درسی نوشته می شوند و در ستون کنار آن فهرست هدفهای آموزشی قرار می گیرد.

• چگونگی توزیع سئوالها به بخشهای مختلف در جدول

بعد از تکمیل محتوا و هدف جدول مشخصات آزمون ، باید مشخص شود که سوالات آزمون چگونه در میان هدفها و عنوانهای درس توزیع گردد، یعنی تعداد سئوالهایی را که به هر يك از موضوعات درسی و هدفهای آموزشی اختصاص خواهند داشت تعیین می گردد و این تعداد را یا به صورت تعداد واقعی سوالها یا بصورت درصد در خانه های جدول مشخصات وارد کند برای تعیین درصد سوالات باید دید که کدام بخش از محتوا و کدام دسته از هدفها دارای حجم بیشتری است و وقت بیشتری صرف آموزش آنها شده یا اهمیت بیشتری دارند. سپس براساس این موارد تعداد سئوالهای آزمون را بین بخشهای مختلف تقسیم کرد. نمونه ای از جدول مشخصات آزمون درس قلب و عروق در زیر ارائه شده است.

سطوح محتوای آموزشی	تجزیه و تحلیل	کاربرد	درک	دانش
--------------------------	------------------	--------	-----	------

¹ - Content Coverage

² -Content Relevance

0	0	1	2	نارسايي قلب
سوال	سوال	سوال	سوال	
1	1	1	2	شوك
سوال	سوال	سوال	سوال	
0	1	1	1	مسموميت با ديگوكسين
سوال	سوال	سوال	سوال	

اگر بخشهاي مختلف از نظر محتوا از اهميت يكساني برخوردار باشند و در توزيع سئوالها در ميان بخشهاي مختلف تنها بخواهيم مقدار ساعتهاي تدريس را براي هر بخش منظور كنيم از فرمول زير استفاده مي كنيم.

تعداد ساعتهايي كه صرف تدريس بخش شده است = نسبت ساعتهاي تدريس هر بخش

تعداد كل ساعتهاي تدريس

روشهاي تعيين روايي محتوايي

روش آماري خاصي جهت تعيين اين نوع روايي وجود ندارد و از قضاوت متخصصين براي اين امر استفاده مي شود. البته امكان خطا بين متخصصين وجود دارد.

روايي صوري

روايي صوري يكي از مشتقات روايي محتوايي است از نظر تعريف منظور از روايي صوري اين است كه سئوالات آزمون تا چه حد در ظاهر شبیه به موضوعي هستند كه براي اندازه گيري تهيه شده اند. بعضي از صاحب نظران روايي صوري را روايي نمي دانند البته عنوان ميشود كه هرگاه صورت ظاهر يك آزمون از لحاظ قابليت پذيرش و معقول بودن براي آزمون شوندگان داراي اهميت باشد و تا آنجايي كه ظاهر آزمون بر انگيزش آزمون شونده اثر مي گذارد روايي صوري مي تواند مهم باشد به عنوان مثال اگر قرار است آزموني داوطلبان يك دوره حسابداري را انتخاب كند اگر سئوالات درخصوص تراز نامه و دفتر حساب باشد بيشتر موجب برانگيختگي آزمون شوندگان ميشود تا اينكه در خصوص مصرف سوخت باشد. در بعضي آزمونها مثل آزمونهاي روانپزشكي طبق نظر ديك و هگرتي بهتر است سئوالات روايي صوري كمی داشته باشند تا بتوانند بيماري رواني را تشخيص دهند در خصوص روش تعيين اين نوع روايي نيز نظر متخصصين شرط اساسي است.

➤ آزمون

1- كداميك از موارد زير نادرست است؟

- الف : بهترين روش براي اينكه مشخص كنيم سئوالات آزمون نمونه كاملي از هدفها و محتواي درس است تهيه جدول مشخصات آزمون است
 ب : براي تعيين روايي محتوايي از ضريب كاپا استفاده مي شود.
 ج : روايي صوري يكي از مشتقات روايي محتوايي است
 د : توزيع سئوالات آزمون بايستى متناسب با مقدار زمان صرف شده جهت هر موضوع باشد.

- 2- مهمترین مفهوم پوشش دادن محتوا چیست؟
 الف : از هر مبحث اساسی حتماً باید سوال طراحی شود
 ب : ظاهر سئوالات باید شبیه موضوع آزمون باشد
 ج : باید آزمون ملاک خوبی از عملکرد فرد باشد
 د : هیچکدام

روایی ملاکی¹

➤ مفاهیم کلیدی

- 1- روایی ملاکی شامل دو زیر گروه روایی پیش بین و همزمان است.
- 2- جهت تعیین روایی ملاکی از ضریب همبستگی بین آزمون اولیه و آزمون ملاک محاسبه می شود.

➤ اهداف یادگیری

- شما بعد از مطالعه این متن باید بتوانید :
- 1- روایی ملاکی و انواع آن را توضیح دهید.
 - 2- محاسبه روایی ملاکی را به طور ساده شرح دهید .
 - 3- خصوصیات یک آزمون مناسب ملاکی را نام ببرید.

روایی ملاکی

زمانی بکار برده می شود که نمرات یک آزمون را بتوان با ملاکی مرتبط دانست.

ملاک رفتاری است که نمره های آزمون برای پیش بینی آن بکار برده می شوند. مثلاً برای داشتن اعتبار وابسته به ملاک نمره های آزمونی که برای انتخاب داوطلبان یک شغل طرح ریزی شده است باید با ملاک کارایی در آن شغل رابطه داشته باشد. هر چقدر رابطه بین نمرات آزمونی که می خواهیم روایی ملاکی آن را بسنجیم با آزمون ملاک بیشتر باشد ضریب روایی ملاکی آن آزمون به همان نسبت بیشتر خواهد بود. اگر آزمون ملاک پس از گذشت یک فاصله زمانی از اول انجام شود به آن روایی پیش بین² می گویند و زمانی که این دو آزمون بدون فاصله و در حال حاضر صورت گیرد به آن روایی همزمان³ گویند. در گذشته این دو نوع روایی را بطور مستقل مورد بحث قرار می دادند ولی از سال 1985 به بعد هر دو به عنوان زیر گروه روایی ملاکی شناخته شدند، چرا که روش تعیین هر دو نوع روایی مشابه است.

روایی پیش بینی

روایی پیش بین ویژه آزمونهایی است که برای مقاصد پیش بینی بکار می روند بطور مثال آیا نمرات دوران اینترنتی دانشجویان پزشکی پیش بین خوبی برای عملکرد آینده آنان به عنوان یک پزشک است و یا مثلاً آیا نمرات کنکور پیش بین خوبی برای عملکرد آتی فرد در دانشگاه هست یا نه .

¹ - criterion related validity

² - Predictive Validity

³ - Concurrent Validity

• روش تعیین روایی پیش بین

ضریب همبستگی بین آزمون پیش بین و آزمون ملاک روش تعیین روایی پیش بین است. این همبستگی با PXY نشان داده می شود در این نماد X نمره آزمون Y نمره ملاک است. در تعیین روایی پیش بین گاهی می توان از چند ملاک استفاده کرد مثلاً اگر قرار باشد آزمون ورودی انتخاب معلمین خوب با عملکرد آنان در آینده به عنوان معلم مرتبط باشد می توان ملاک را میزان موفقیت معلمان از نظر مدیران، دانش آموزان و همکاران حساب کرد. بعضی از مواقع می توان از چند آزمون پیش بین استفاده کرد. مثلاً، افزودن چند عامل پیش بینی دیگر نظیر نمرات دبیرستان فرد به نمرات کنکور می تواند قدرت پیش بینی آن را اضافه کند.

مقایسه روایی پیش بین با همزمان

گفتیم که هر دو نوع روایی زیر مجموعه روایی ملاکی هستند ولی یک تفاوت بین آنها وجود دارد در روایی پیش بین داده های مربوط به ملاک در آینده بدست می آیند بنابراین واقعاً نمرات برای پیش بینی بکار می روند ولی در روایی همزمان هدف جایگزین ساختن آزمون پیش بین با آزمون ملاک است مثلاً آیا می توان نمره یک آزمون روانپزشکی را بجای معاینه بالینی در تشخیص بیماری بکار برود یا نه .

ویژگیهای مطلوب ملاک

شرانیدیک و هیگن در سال 1979 برای داده های ملاکی چهار ویژگی مطلوب ذکر کرده اند این ویژگیها به ترتیب اهمیت عبارتند از : مرتبط بودن ، بی طرفی ، پایایی و دردسترس بودن.

✓ مرتبط بودن

ربط داشتن داده های یک آزمون ملاکی از نظر تعریف قابلیت انطباق آنها با موقعیت واقعی می باشد. در این راه بهترین روش قضاوت مصححان است. برای مثال بعضی معتقدند که موفقیت در دانشگاه یعنی میزان دانش کسب شده در طول سالیان تحصیل در دانشکده برخی دیگر می گویند میزان دانش کسب شده ملاک درستی از موفقیت دانشگاهی است. لیکن نمرات دانشگاهی شاخص خوبی برای اندازه گیری آن نیستند و عوامل دیگر مثل روابط اجتماعی خوب، شغل خوب مهم هستند.

✓ بی طرفی

باید آزمون ملاک به گونه ای باشد که همه افراد در کسب نمره خوب فرصت مساوی داشته باشند مثلاً شناخت قبلی معلم از دانش آموزان سبب نشود که به طور ناخودآگاه یا آگاهانه نمرات بالاتر یا کمتری به او بدهد.

✓ پایایی

در مورد پایایی قسمت مربوطه به طور مفصل توضیح داده خواهد شد. پایایی بطور مختصر یعنی با ثبات بودن و تکرارپذیری که شرط مهم یک آزمون ملاکی است.

✓ دردسترس بودن

ملاک عملی باید آسان و ساده باشد و امکان دسترسی به آن وجود داشته باشد. درباره این ویژگی با توجه به زمان، هزینه و امکانات لازم قضاوت می شود.

➤ آزمون

- 1- ویژگی های مهم داده های ملاکی شامل موارد زیر است **بجز**:
الف: _____ مرتبط با _____ بودن
ب: بی طرفی
ج: دردسرس _____ ترس با _____ بودن
د: شناخت قبلی معلم از دانشجویان
- 2- اگر بخواهیم نمرات کنکور و عملکرد فرد در دانشگاه مقایسه کنیم از کدامیک از انواع روایی زیر استفاده می کنیم:
الف: _____ ص _____ وری
ب: محتوایی
ج: پیش _____ ین
د: سازه

روایی سازه

➤ مفاهیم کلیدی

- 1- روایی سازه جدیدترین شکل روایی است.
- 2- روایی چند روشی، چند صفتی جنبه ای از روایی سازه به حساب می رود و اعتبار عاملی شکل دیگری از اعتبار سازه است.
- 3- دو نوع مهم روایی چند روشی - چند صفتی شامل اعتبار همگرا و اعتبار تمیز است.

➤ اهداف یادگیری

- پس از مطالعه این مطلب شما باید بتوانید:
- 1- روایی سازه را تعریف نمایید.
 - 2- روش روایی چند روشی - چند صفتی را توضیح دهید.
 - 3- روایی همگرا و تمیز را شرح دهید.
 - 4- با استفاده از جدول روایی چند روشی، چند صفتی روایی تست های مختلف جهت سنجش موارد مختلف را مورد تفسیر قرار دهید.

تعریف

روایی سازه جدیدترین شکل روایی است از نظر تعریف روایی سازه یک آزمون عبارتست از میزان صحت آن آزمون در اندازه گیری ساخت نظری یا ویژگی مورد نظر روایی سازه یک فرایند مداوم است براساس نظریه کنونی سازنده آزمون درباره صفت مورد اندازه گیری و چگونگی کاربرد نمره های آزمون در موقعیت های مختلف پیش بینی های می کند و سپس آنها را آزمایش می کند چنانچه نتیجه داده ها پیش بینی های انجام شده را تأیید کند اعتبار سازه افزایش می یابد در غیر اینصورت یا آزمون اشتباه است یا فرضیه غلط است و یا اینکه آزمون اصلاً نتوانسته ویژگی مورد نظر را اندازه گیری کند.

کرلینجر در سال 1965 سازه را اینگونه تعریف کرده است: سازه یک مفهوم است اما سازه یک معنی اضافی بر مفهوم نیز دارد و آن این است که سازه برای مقاصد ویژه عملی بطور عمده و از روی آگاهی ابداع می شود.

مورفي و ديويد شوفر نيروي گرانش را به عنوان يك سازه علمي مثال آورده اند وقتي كه ميوه اي از درخت مي افتد نيروي جاذبه يا نيروي گرانش براي پيش بيني بكار مي رود. نيروي گرانش قابل ديدن نيست تنها چيزي كه ديده ميشود افتادن ميوه از درخت است با اين وجود كاملاً منطقي است كه نيروي گرانش اندازه گيري و درباره آن نظريه پردازي شود.

هرگونه پيش بيني آزمائش پذير از قبيل پيش بيني هاي محتوا و اعتبار وابسته به ملك براي تائيد اعتبار سازه انجام پذير است. پيش بيني هاي ديگر عبارتند از تفاوتهاي گروهی، تغييرات، همبستگی ها، فرآیندها.

• تفاوت هاي گروهی

بستگی به اين دارد يك آزمون دنبال تفاوت باشد يا نباشد به عنوان مثال در يك آزمون رشد اجتماعي پژوهشگر مي تواند بين كودكان و بزرگسالان را پيش بيني كند يا برعكس در آزمون عدم تفاوت بين گروههاي فرهنگي اندازه گيري شود.

• تغييرات

بعضي از موارد معتبر در نمرات آزمون با گذشت زمان داريم بعنوان مثال اگر قرار باشد يك آزمون مهارت هاي ارتباطات كلامي را اندازه گيري كند با بزرگ شدن كودك نمره او بالاتر مي رود.

• همبستگی ها

ممکن است در بعضي موارد همبستگی مثبت، منفي يا صفر وجود داشته باشد. مثلاً يك آزمون حافظه کوتاه مدت ممکن است با سن همبستگی مثبت داشته باشد ولي با جنس همبسته نباشد.

• فرآیندها

فرض كنيد يك آزمون استدلال رياضي از لغات دشواري تشكيل شده باشد. در اينصورت آن آزمون در عين حالي كه اعمال رياضي را مي سنجد ممكن است توانايي درك واژگان را هم بسنجد.

روايي چند روشي - چند صفتي¹

اين روايي جنبه اي از روايي سازه به شمار مي رود. زماني به كار مي رود كه دو يا چند صفت به وسيله دو يا چند روش اندازه گيري مي شود. فرض كنيد دو صفت درون گرایی و روان رنجوري با دو روش اندازه گيري مي شوند يكي با روش سئوالهاي صحيح، غلط و ديگري با استفاده از پرسشهاي چند گزينه اي چهار نوع آزمون به نمونه اي از افراد داده شود و نتايج به صورت ماتريس در جدول نشان داده مي شود همبستگی هايي كه در محل تقاطع هر ردیف و ستون نشان داده شده اند همبستگی بين نمرات سطر و ستوني است. معمولاً ضريب همبستگی بين قطرها بايد مساوي يك باشد البته به علت ضريب پايي هاي برآورد شده معمولاً كمتر از 1 گزارش مي شود.

¹ Multitrait- multi method validity

✓ خصوصیات ماتریس به شرح زیر است:

- میزان پایایی هایی که در قطر اصلی ماتریس وجود دارد بایستی بالا باشد
 - همبستگی‌های بین دو اندازه‌گیری مختلف از یک صفت باید بالا و همبستگی بین اندازه‌گیری صفت‌های نامربوط باید پایین باشد
 - دو نوع اعتبار مهم چند روش، چند صفتی عمده وجود دارد یکی اعتبار همگرا¹ دیگری اعتبار تمیز².
- اعتبار همگرا از طریق همبستگی بالا بین نمره‌های آزمون‌های نشان داده می‌شود که به روش‌های مختلف ویژگی‌های یکسانی را اندازه‌گیری می‌کند. این همبستگی بالا نشان می‌دهد که آزمون‌ها در ویژگی‌های مورد اندازه‌گیری همگرا هستند.
- اعتبار تمیز از طریق همبستگی‌های پایین بین نمره‌های آزمون‌هایی که ویژگی‌های مختلفی دارند نشان داده می‌شود همبستگی‌های محاسبه شده پایین نشان می‌دهد که آزمون‌ها بین ویژگی‌های مختلف تمیز و تشخیص داده شد.
- همبستگی‌های جدول زیر روایی چند روشی، چند صفتی همگرا و تمیز را نشان می‌دهد.

جدول ماتریس اعتبار چند روشی - چند صفتی

روان رنجوری چند گزینه‌ای	درون گرایی چند گزینه‌ای	روان رنجوری صحیح غلط	درون گرایی صحیح غلط	
%17	%75	%20	%80	درون گرایی صحیح غلط
%71	%15	%85	%20	روان رنجوری صحیح غلط
%21	%83	%15	%75	درون گرایی چند گزینه‌ای
91	%21	%71	%17	روان رنجوری چند گزینه‌ای

زمانی که بیش از دو صفت یا روش در کار باشد این ماتریس کمی پیچیده تر و بزرگتر است.

¹ - Convergent validity

² - Discriminative validity

بطور كلي الگوي همبستگی در این روش باید برای همبستگی‌هاي درون صفتي داراي ارزش بالا و برای همبستگی‌هاي بين صفات داراي ارزش پائين باشد از مزایاي این روش امکان محاسبه روایي همگرا و تمیز بطور مستقل و همچنین استاندار و مستقلي برای روایي سازه است. از معایب این روش سختي کار است و همچنین نیاز به تجزیه و تحلیل دارد.

اعتبار عاملي

اعتبار عاملي شکلي از اعتبار سازه است که از طریق تحلیل عاملي انجام میشود.

تحلیل عاملي يك روش آماری است که از طریق آن تعداد و ماهیت متغیرهاي را که يك آزمون اندازه می‌گیرد مشخص می‌کنند. در تحلیل عاملي این کار از طریق ادغام تعداد زيادي از متغیرها با همديگر و ایجاد تعداد محدودی متغیر قابل بررسی به نام عامل ایجاد می‌شوند. این روش روابط دروني میان داده هاي مورد تحلیل را تعیین می‌کند و برای ایجاد سهولت متغیرها را به خوشه‌ها یا عوامل کاهش می‌دهد و سپس این خوشه ها را نام گذاری می‌کند.

➤ آزمون

- 1- کدامیک از خصوصیات ماتریس چند روش چند صفتي است؟
 - الف : میزان پایايي در قطر اصلي ماتریس باید بالا باشد
 - ب : همبستگی هاي بين دو اندازه گيري مختلف از يك صفت باید پایين باشد
 - ج : همبستگی بين بين اندازه گيري صفت هاي نامربوط باید بالا باشد
 - د : ضریب همبستگی قطرها مساوي يك است
- 2- کدامیک از موارد زیر صحیح است :
 - الف : اعتبار عاملي شکلي از اعتبار سازه است که از طریق تحلیل عاملي انجام میشود.
 - ب : تحلیل عاملي يك روش آماری جهت تعیین اعتبار سازه است
 - ج : هر دو
 - د : هیچکدام

خلاصه

روایي اصطلاحی است که به هدفی که آزمون برای تحقق بخشیدن به آن درست شده است اشاره می‌کند. روایي داراي انواع روایي محتوایي، روایي ملاکی و روایي سازه می‌باشد.

روایي محتوایي به این مطلب اشاره می‌کند که آیا آزمون از مجموعه سئوالهاي درست شده است که معرف يك نمونه واقعي از محتوای درس و هدفهاي آموزشی آن باشد یا نه. موردی از روایي محتوایي روایي صوري است روایي صوري نشان می‌دهد که ظاهر سئوالهاي يك آزمون به آنچه آزمون می‌خواهد بسنجد شباهت دارد یا نه؟

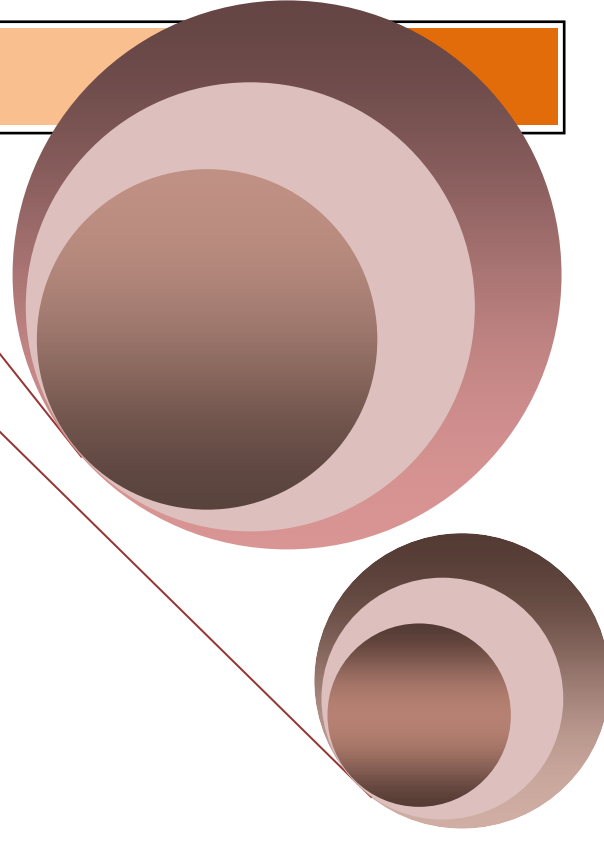
روایي ملاکی دو بخش مهم دارد که عبارتند از روایي پیش بین و روایي همزمان. روایي پیش بین نشان می‌دهد که آیا براساس عملکرد

آزمون شوندگان در يك آزمون مي توان عملكرد آنها را در آزمون ديگري پيش بيني كرد يا نه ؟ به آزمون اول كه براي پيش بيني بكارمي رود آزمون پيش بين و به آزمون دوم كه نمرات آزمودني در آن از روي آزمون اول پيش بيني ميشود آزمون ملاك مي گویند اگر آزمون پيش بين و همزمان همراه انجام شود. در اين حالت آزمون داراي روايي همزمان است.

نوع ديگري از روايي، روايي سازه است. روايي سازه ويژه آزمونهائي است كه درستي نظريه هاي علمي را مي سنجند يك آزمون در صورتي داراي روايي سازه است كه نمرات حاصل از اجراي آن به مفاهيم يا سازه هاي نظريه مورد نظر مربوط باشند.

منابع

1. سيف دكتر علي اكبر، اندازه گيري ، سنجش و ارزشيابي آموزشي ويرايش سوم چاپ 1383.
2. آلن مري جي، وندي ام ين. مقدمه اي بر نظريه هايي اندازه گيري ترجمه دكتر علي دلاور انتشارات سمت، چاپ اول پاييز 1374.
3. رضايي دكتر حميد رضا، شوقي شفق آريا فرنگيس، دولتي ، دكتر يحيي، محققي دكتر محمد علي، ملك زاده ، دكتر رضا ، تكنيك هاي ارزيابي در پزشكي چاپ اول انتشارات آمه 1377 .
4. Renger R, Meadow LM. Testing for predictive validity in health care education research : a critical review. Academic medicine .1994;69(8): 685-7
5. Assessing construct validity in organizational research. Administrative science . 1991;36(1): 1-10



پا پی

Advanced Topics

فهرست مطالب

52	اهداف كلي
52	1 اصول و مباني پايي
52	1.1 مقدمه
53	1.2 تعريف و مفهوم تئوري پايي
54	2 روشهاي تعيين پايي در آزمونهاي نوشتاري
54	2.1 روش بازآزمائي
55	2.2 روش فرمهاي موازي يا هم ارز
55	2.3 روشهاي همساني دروني
57	3 روش تعيين پايي در آزمونهاي شفاهي و آزمونهاي عملکردي
57	3.1 روش تعيين پايي در آزمونهاي شفاهي
	3.2 پايي در امتحانات عملکردي نظير آسكي و استفاده از بيمار
58	استاندارد
59	3.3 چه مقدار پايي كافي است؟
59	3.4 راههاي بهتر پايي در ارزشيابيها
	60 خلاصه
60	منابع



شناسه جستار

عنوان جستار: پایایی

نویسنده: دکتر مهرابی

امینی

دانشیار دانشگاه علوم پزشکی

شیراز

طراح آموزشی: دکتر مانوش

مهرابی

کارشناس آموزشی دانشگاه

علوم پزشکی شیراز

ویرایش: خانم صفاری

آخرین تاریخ به روز

رسانی: 6/ 8/ 1388

با همکاری:

مرکز مطالعات و توسعه آموزش

پزشکی

قطب علمی آموزش الکترونیکی

دانشگاه علوم پزشکی شیراز

اهداف کلی

- در تدوین این درس اهداف کلی زیر مورد نظر است:
- انتظار می رود شما بعد از خواندن مطالب این نوشتار به اهداف زیر نائل آئید:
- آشنایی با
- تعریف پایایی و انواع آن
 - محاسبه ضریب پایایی با روشهای مختلف
 - روش آزمون مجدد و روش دو نیم کردن آزمون
 - مفاهیم K₂₀ و K₂₁ (کودریچاردسون 20 و کودریچاردسون 21)
 - رابطه پایایی با روایی

اصول و مبانی پایایی¹

➤ مفاهیم کلیدی

- 1- پایایی درحقیقت تکرار پذیری یک آزمون است .
- 2- پایایی شرط لازم روایی است.
- 3- همیشه در عالم واقعی در اندازه گیریها خطا وجود دارد.

➤ اهداف یادگیری

- پس از مطالعه مطالب این بخش شما قادر خواهید بود:
- 1- پایایی را تعریف کنید.
 - 2- انواع پایایی را ذکر کنید.
 - 3- رابطه روایی با پایایی را بیان کنید.
 - 4- براساس تئوری سنجش کلاسیک فرمول ضریب پایایی را تفسیر کنید.

مقدمه

پایایی یک وسیله اندازه گیری عمدتاً به دقت نتایج حاصل از آن اشاره می کند. پایایی در حقیقت به تکرار پذیر بودن نتایج یک آزمون است نه ابزار آن. کاپلان و ساکوزا در سال 2001 گفته اند پایایی به دقت ، اعتماد پذیری ، ثبات یا تکرار پذیری نتایج آزمون اشاره می کند. برنان² چنین بیان می کند که کمی سازی نتایج ثبات یا عدم ثبات در یک آزمون اساس مفهوم پایایی را تشکیل می دهد. اصولاً یکی از خصوصیات اصلی هر فعالیت علمی از قبیل پژوهش برآن است که بایستی اصول کاربه گونه ای باشد که تکرار پذیر باشد یعنی اگر محقق دیگری

¹-Reliability

² - Brennan

خواست پژوهش را تکرار نماید نتایج مشابهی را بدست آورد در غیر این صورت یعنی اگر پژوهش تکرار پذیر نباشد میتوان چنین بیان کرد که نتایج حاصل از آن مشکوک است. این موضوع در مورد داده‌های ارزشیابی نیز صدق می‌کند بدین صورت که نتایج یک آزمون نیز بایستی تکرار پذیر باشد به عبارت دیگر در صورتی که آزمونی انجام شود و نتایج آن تکرار پذیر نباشد به چه درد می‌خورد؟

بنابراین پایایی شرط لازم برای روایی است و در تمام ارزشیابی‌ها پایایی یکی از اصول مهم روایی به حساب می‌آید. داده‌های ارزشیابی اگر دارای پایایی نباشند یک منبع جهت خطا به حساب می‌آیند.

تعریف و مفهوم تئوری پایایی

در خصوص انواع مختلفی از پایایی در مقالات آموزشی بحث شده است تمام محاسبه‌های پایایی به این موضوع اشاره می‌کنند که در هر آزمون یک مقدار خطا به همراه نتیجه آزمون وجود دارد.

بر اساس تئوری سنجش کلاسیک¹ پایایی در حقیقت واریانس نمره واقعی تقسیم بر واریانس کل می‌باشد. معمولاً ضریب پایایی نظیر ضریب همبستگی حساب می‌شود.

$$X = T + E$$

که در این فرمول X نمره مشاهده شده²

T نمره واقعی³

E خطا⁴ می‌باشد.

در دنیای ایده آل فرض بر این است که هیچ خطایی وجود ندارد و نمره مشاهده شده درحقیقت نمره واقعی می‌باشد ولی در عالم واقعیت همیشه این خطا وجود دارد و ضرایب پایایی به منظور تعیین این نوع خطا بکار می‌رود.

در این مبحث پایایی سه روش مهم ارزشیابی مورد بحث قرار می‌گیرد. اول پایایی تست‌های نوشتاری در قسمت دوم پایایی مصاحبه‌های شفاهی و در قسمت سوم پایایی آزمون‌های عملکردی ساختار یافته مثل امتحان آسکی و امتحانات با استفاده از بیمار استاندارد مورد بحث قرار می‌گیرد.

➤ آزمون

1- کدامیک از موارد زیر در مورد پایایی صحیح است؟

الف: پایایی یعنی یک وسیله اندازه گیری چیزی را باید اندازه بگیرد، اندازه گیری کند.

ب: پایایی یعنی تکرارپذیری نتایج آزمون

ج: پایایی یعنی سئوالات آزمون شبیه مفهومی باشند که اندازه گیری می‌شود.

د: پایایی یعنی سازه های یک آزمون بهم مرتبط باشند.

2- از نظر تئوری سنجش کلاسیک کدامیک صحیح است

1 -Classic Measurement Theory (CMT)

2 - Observed Score

3 - Total Score

4 - Error

- الف : خطا + نمره مشاهده شده = نمره واقعي
 ب : خطا + نمره واقعي = نمره مشاهده شده
 ج : خطا * نمره مشاهده شده = نمره واقعي
 د : خطا * نمره واقعي = نمره مشاهده شده

روشهاي تعيين پايي در آزمونهاي نوشتاري

➤ مفاهيم كليدي

- 1- روشهاي تعيين پايي متفاوت هستند. و معمولاً اين روشها از پيشنهادهاي اپيل، ویرسما، جورس، مرفي و ديوييد شوfer گرفته شده اند.
- 2- روشهاي تعيين پايي آزمونهاي نوشتاري شامل روش دونيمه کردن، روش کودر ریچاردسون و روش ضريب آلفاي کرونباخ هستند.

➤ اهداف يادگيري

- شما بايد پس از مطالعه متن بتوانيد:
- روش بازآزمائي را شرح دهيد .
 - روش موازي را توضيح دهيد .
 - روش KR20 و KR21 را شرح داده و تفاوت آنها را بيان كنيد.
 - روش آلفاي کرونباخ را شرح داده و از داده ها آن را محاسبه كنيد.

روش بازآزمائي¹

ساده‌ترین روش تعیین پایایی يك آزمون روش بازآزمائي است در اين روش آزمون را در دو نوبت به يك گروه از شرکت کنندگان مي دهند و نمرات حاصل را با هم مقايسه مي کنند. ضريب همبستگي بين نمرات حاصل از دو بار اجراي آزمون ضريب پايي آزمون است. به اين نوع پايي، پايي بازآزمائي مي گویند. اين روش تنها زماني قابل استفاده است که ويژگي مورد نظر در طول زمان تغيير نکند يا مثلاً دانشجويان در طول زمان مطالب جديدي را در آن حيطه ياد نگرفته باشند به عنوان مثال هميشه فرض ما بر اين است که هوش در طول زمان تغيير نمي کند لذا مي توان آزمون هوش را بازآزمائي کرد.

در اين روش در صورتي تست پايي است که نمرات مشابهي را دانشجويان در دو آزمون کسب نمايند. گر چه روش باز آزمائي يکي از مفاهيم اساسي در پايي را مورد بحث قرار مي دهد. ليکن در محيط واقعي بندرت استفاده مي شود. چرا که تکرار آزمون مشکل است. از طرفي متخصصين ارزشيابي به اين روش انتقاد کرده اند از جمله اينکه گفته اند تجربه نوبت اول آزمون سبب مي شود که آزمون شونندگان با سئوالات آزمون آشنا شوند و اين آشنائي باعث خواهد شد که به پاسخ دادن به همان سواها درنوبت دوم اجراي آزمون بهتر عمل کنند.

¹ - Test retest

روش فرمهاي موازي يا هم ارز¹

اين روش جهت رفع مشكلات روش بازآزمائي بكار مي رود در اين روش دو آزمون معادل يا موازي براي يك مطلب يا موضوع تهيه مي كنند و آنها را در فاصله زماني كوتاهي به يك گروه واحد از آزمون شونندگان مي دهند. دو فرم يك آزمون در صورتي هم ارز يا معادل هستند كه ميانگين و واريانس آنها برابر باشد ضريب همبستگي بين نمرات حاصل از اجراي اين فرم آزمون ضريب پايائي آن آزمونها به حساب مي آيند. اين روش نيز مشكلات خاص خود را دارد از جمله اينكه تهيه آزمونهاي موازي مشكل است و بعضي مواقع آزمون شونندگان در جواب دادن به سوالهاي فرم اول آزمون با اصولي آشنا مي شوند كه عملکرد آنان را در فرم دوم تحت تاثير قرار مي دهد.

روشهاي همساني دروني²

براي رفع مشكلات روشهاي قبلي اين روش استفاده مي شود كه در اين روش برخلاف روشهاي قبلي كه بر ثبات نتايج آزمون (روش بازآزمائي) يا بر برتري نتايج آزمون (روش فرمهاي هم ارز) تاكيد دارند در روش همساني دروني بر همساني يا يكنواختي ماده ها يا اجزاي تشكيل دهنده يك آزمون تاكيد مي شود . در كل روشهاي همساني دروني آزموني را كه مي خواهند ضريب پايائي اش را تعيين نمايند يك بار با گروه واحدي از آزمون شونندگان اجرا مي كنند و سپس با يكي از روشهاي زير به تعيين ضريب پايائي آن مي پردازند.

➤ روش دو نيمه كردن آزمون³

آزمون در اين روش يك بار اجرا مي شود بعد از اجرا به دو نيمه تقسيم مي شود كه بهترين راه دو نيمه كردن اين است كه همه سئوالهاي فرد را جزء يك آزمون به حساب آوريم و همه سئوالهاي زوج را نيز آزمون ديگري بدانيم ضريب همبستگي حاصل از نمرات دو نيمه آزمون ضريب پايائي هر يك از دو نيمه خواهد بود .
سپس ضريب همبستگي بين نيمه ها را در فرمول اسپيرمن براون⁴ مي گذاريم .
نتيجه حاصل ضريب پايائي كل آزمون است:

$$r_{tt} = \frac{2r_{11}}{1 + r_{\frac{11}{22}}}$$

ضريب پايائي كل آزمون = r_{tt}

ضريب همبستگي بين نيمه هاي آزمون $r_{\frac{1}{2}} * \frac{1}{2}$

-
- 1 - Parallel form reliability
 - 2 - Internal Consistency Reliability
 - 3 - Split halves
 - 4 - Spearman Brown Formula

➤ روش کودر - ریچاردسون¹

در روش کودر- ریچاردسون نیز آزمون تنها یکبار اجرا می شود ولی در این روش همه ماده های آزمون تحلیل می شوند در این روش از دو فرمول KR₂₀ و KR₂₁ استفاده می شود.

فرمول KR₂₀ به شرح زیر است:

$$rtt = \left[\frac{n}{n-1} \right] \left[1 - \frac{\sum pq}{\sigma^2} \right]$$

ضریب پایایی آزمون = rtt

n = تعداد سئوالهای آزمون

واریانس کل سئوالهای آزمون = σ^2

p = نسبت دانش آموزانی که به سئوال پاسخ درست داده اند

q = 1 - p = نسبت دانش آموزانی که به سوال پاسخ غلط داده اند

مجموع pq ها برای تمام سئوالها $\sum pq$

در فرمول KR₂₀ واریانس هر سئوال را برحسب pq به حساب می آورد اگر بتوان تصور کرد که همه سئوالهای آزمون از لحاظ درجه دشواری شبیه هم هستند بجای $\sum pq$ ، npq قرار داده می شود که n معرف تعداد سئوالات آزمون است. P ضریب دشواری سئوال است و ضریب KR₂₁ به شرح زیر است.

$$rtt \left(\frac{n}{n-1} \right) \left(1 - \frac{npq}{\sigma^2} \right)$$

استفاده از این فرمول ساده تر از ضریب KR₂₀ است.

➤ روش ضریب آلفای کرونباخ²

سومین روش تعیین پایایی آزمون با تاکید بر همسانی درونی روش آلفای کرونباخ نام دارد. این روش جهت زمانی بکار می رود که آزمون از قسمتهای مختلفی تشکیل شده باشد. اگر به جای سئوالها یا ماده ها آزمون از بخشها یا قسمتهای تشکیل شده باشد مانند زمانی که آزمون از تعدادی خرده آزمون تشکیل شده است از فرمول زیر استفاده می شود:

$$r_{\alpha} = \frac{J}{J-1} \left(1 - \frac{\sum \sigma_j^2}{\sigma^2} \right)$$

J = تعداد بخشهای آزمون

σ_j^2 = واریانس نمرات هر بخش آزمون

σ^2 = واریانس نمرات کل آزمون

اگر روش آلفا و روش دو نیمه کردن آزمون را با هم مقایسه کنیم خواهیم دید که روش دو نیمه کردن آزمون نوعی روش ضریب آلفا است. به عبارت دیگر ضریب آلفا معرف میانگین ضرایب پایایی حاصل از همه دو نیمه های ممکن یک آزمون است

¹ - Kuder Richardson

² - Coefficient alpha Cronbach

تنها تفاوت این دو روش این است که در روش دو نیمه کردن آزمون یک نیمه آزمون یا نیمه دیگرش قایسه می شود، در حالی که در روش ضریب آلفا هر سوال آزمون با تک تک سئوالهای دیگر مقایسه می شود.

➤ آزمون

- 1- کدامیک از روشهای زیر جهت سنجش پایایی تست های نوشتاری بهتر است؟
 - 1- روش بازآزمایی
 - 2- روش فرمهای موازی یا هم ارز
 - 3- روشهای همسانی درونی
 - 4- هیچکدام
- 2- کدام یک از موارد زیر جزء روشهای همسانی درونی است؟
 - الف : روش دو نیمه کردن آزمون
 - ب : روش کودر ریچادرسون
 - ج : روش آلفای کرونباخ
 - د : همه موارد

روش تعیین پایایی در آزمونهای شفاهی و آزمونهای عملکردی

➤ مفاهیم کلیدی

- 1- در آزمون شفاهی توافق آزمون گیرندگان خیلی مهم است
- 2- در آزمونهای عملکردی علاوه بر تست های امتحان و توافق امتحان گیرندگان فاکتورهای محیطی و نحوه رفتار بیماران استاندارد نیز بر پایایی موثرند.

➤ اهداف یادگیری

- شما باید پس از مطالعه این متن بتوانید:
- 1- روش تعیین پایایی در آزمونهای شفاهی را ذکر نمایید.
 - 2- ضریب فی و کاپا را تعریف کنید.
 - 3- تئوری عمومی سازی را توضیح دهید.
 - 4- اهمیت پایایی در امتحانات عملکردی را شرح دهید .

روش تعیین پایایی در آزمونهای شفاهی

آزمون شفاهی در بخشهای بالینی پزشکی به عنوان یکی از روشهای مهم ارزشیابی در کلینیک و بیمارستان بکار می رود. در بعضی از مواقع این آزمون شفاهی بوسیله تعدادی از اساتید که نمره دهی را انجام می دهند برگزار می شود. در این حالت پایایی به هماهنگی بین نمرات افراد امتحان گیرنده بستگی دارد. مهمترین تهدید این روش تعیین پایایی تفاوت نظر امتحان گیرندگان است روشهای مختلفی جهت محاسبه این نوع پایایی وجود دارد که عبارتند از :

❖ روش پایایی مصححان¹

برای تعیین پایایی آزمونهای شفاهی راههای مختلفی وجود دارد یکی از ساده ترین راه ها اینست که از دو یا چند مصحح که مستقیماً پاسخهای آزمون شوندگان را تصحیح می کنند استفاده کرد. همبستگی بین نمرات مصححان مختلف شاخص پایایی مصححان به حساب می آید. هرچه میزان توافق مصححان یا نمره گذاران بیشتر باشد پایایی مصححان هم بیشتر است.

تعیین پایایی به این روش برای استفاده روزانه موثر است ولی روشهای آماری بهتری نیز نظیر محاسبه ضریب $kappa$ بخصوص در زمانی که دو نفر مصحح وجود دارد بدرجاء بخور است.

روش دیگر تعیین ضریب phi می باشد که شانس تصادفی بودن توافق را مد نظر قرار نمی دهد برای همین مقداری توافق را بیشتر از حد واقعی محاسبه می نمایند.

❖ تئوری عمومی سازی²

برای برقراری صحیح این تئوری بایستی تمام فاکتورهایی را که بر پایایی تاثیر می گذارند مد نظر است از قبیل افراد، مصححان، تست ها در این روش واریانس مربوط به هر عامل را حساب کرده و خطای اندازه گیری نیز محاسبه می گردد.

❖ روش آنالیز واریانس³

این روش نیز مثل تئوری عمومی سازی است در این روش ضریب همبستگی فاکتورها با تست آنالیز واریانس محاسبه می شود. نقطه قوت این روش امکان استفاده از آن در بیشتر نرم افزارهای آماری است.

پایایی در امتحانات عملکردی نظیر آسکی و استفاده از بیمار استاندارد

در سالهای اخیر استفاده از روشهای سنجش عملکرد در علم آموزش پزشکی نقش مهمی را ایفا نموده است. عملکرد دانشجویان در حیطه های مختلفی نظیر مهارتهای ارتباطی، نحوه اخذ تاریخچه بیماری، تشخیص، حل مسئله و... قابل سنجش است. با توجه به اینکه دو روش آسکی و استفاده از بیمار استاندارد، روشهای استانداردتری نسبت به یک آزمون شفاهی آخر بخش هستند از این روشها در سالهای اخیر استفاده بیشتری شده است.

پایایی در آزمون های عملکردی نقش مهمی دارد چرا که پایایی در این آزمونها به عوامل دیگری علاوه بر سئوالات امتحان و مصححان نیز بستگی دارد.

برای محاسبه پایایی نیاز به مدلهای پیشرفته ای است که بطور مثال از تئوری عمومی سازی استفاده میشود که در تئوری تمام موارد از قبیل سئوالات، مصححان رفتار بیماران استاندارد و... مورد بررسی قرار گرفته و خطای معیار نیز محاسبه می شود.

¹ - Inter rater or inter observer reliability

² - Generalisability theory

³ - Analysis of variances (anova)

چه مقدار پایایی کافی است؟

همانگونه که می دانید ضریب پایایی از صفر تا یک متغیر است و هرچه بیشتر باشد آزمون پایاتر است. پاسخ این سؤال بستگی به مقصود ارزشیابی، استفاده نهایی از آن و نتیجه ای که از استفاده از آن عاید می شود بستگی دارد. منابع مختلف اعداد متفاوتی را در این زمینه ذکر کرده اند ولی بیشتر صاحب نظران معتقدند برای امتحانات مهم مثل امتحان ورود، امتحاناتی که منجر به مدرک بخصوص مدارک سطح بالا میشود از آنجایی که دادن مدرک به یک پزشک متخصص، دندانپزشک و... تاثیر مستقیمی بر روی جامعه دارد باید این ضریب باید بالای 9/ باشد. برای امتحاناتی که یک درجه اهمیت کمتری دارند مثلاً امتحانات ارتقاء پایان سال بهتر است ضریب بالای 8/ باشد. برای امتحاناتی که توسط اعضای هیات علمی همان دانشگاه گرفته شده و یا امتحانات کلاسی پایان دوره ضریب بالای 7/ قابل قبول است. موارد دیگر نظیر کوئیز و امتحانات تراکمی می توانند ضرایب پایایی کمتری داشته باشند.

راههای بهتر پایایی در ارزشیابیها

راههای متفاوتی جهت بهبود پایایی آزمونها وجود دارد مهمترین آنها بالابردن تعداد سئوالات و تعداد مصححان است باید مطمئن شویم که سئوالات به درستی نوشته شده اند و توضیحات کافی در مورد سئوالات بخصوص در آزمونهای عملکردی داده شده است و همچنین افراد صاحب نظر سئوالات را بررسی نموده اند. از نظر درجه سختی و دشواری بهتر است جهت افزایش پایایی سئوالات نه خیلی سخت و نه خیلی آسان باشند. سئوالات متوسط بهتر هستند استفاده از بانک سئوالات استاندارد قبلی نیز در این زمینه مفید است.

➤ آزمون

1- کدامیک از موارد زیر جهت بهبود پایایی مؤثر نیست؟

- الف : بالابردن تعداد سئوالات
- ب : بالابردن تعداد مصححان
- ج : بالابردن درجه دشواری سؤال
- د : ارائه توضیحات کافی به دانشجویان

2- کدامیک از موارد زیر در خصوص مقدار پایایی صحیح است؟

- الف : در آزمونهای مهم این ضریب باید بالای 5/ باشد
- ب: در آزمونهای کلاسی ضریب می تواند مقادیر کمتری داشته باشد
- ج : در امتحانات ارتقاء باید ضریب بالای 4/ باشد
- د : مقدار ضریب مهم نیست و سطح معنی داری اهمیت دارد.

3- کدامیک از ضرایب زیر جهت توافق مصححان بکار نمی رود؟

- الف : کاپا
- ب : فی
- ج : آلفا

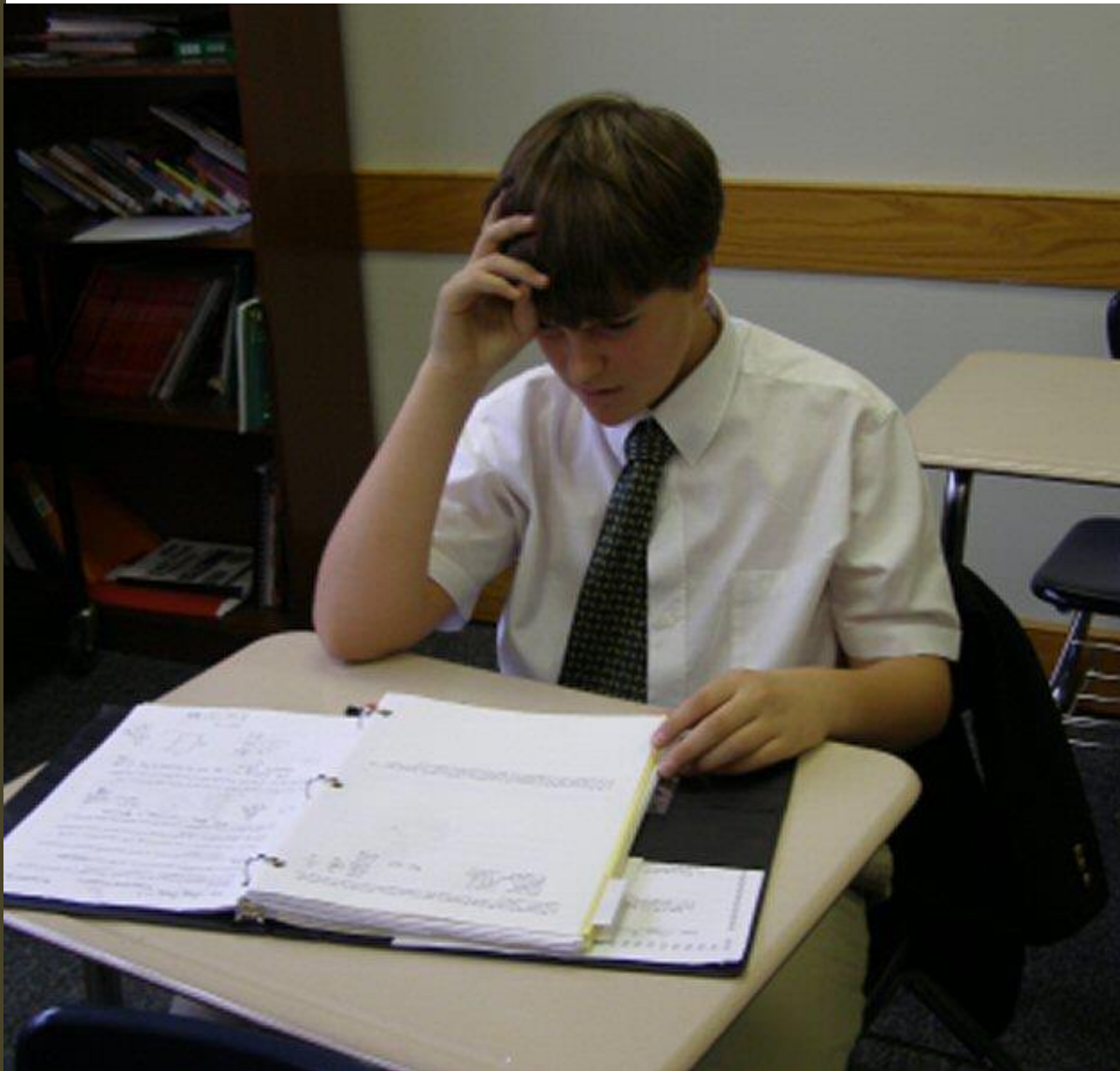
خلاصه

در این مبحث یکی از ویژگی‌های مهم آزمون یعنی پایایی را معرفی کردیم پایایی به میزان دقت یک آزمون اشاره می‌کند. آزمون پایا آزمونی است که اگر چندین بار به یک گروه از دانشجویان یا دانش آموزان داده شود نتایج یکسانی به دست می‌دهد برای تعیین پایایی تست های نوشتاری روشهای مختلفی از قبیل بازآزمایی ، فرمهای موازی یا هم ارز برای تعیین پایایی آزمونهای شفاهی از میزان توافق بین امتحان گیرندگان و ضرایب فی و کاپا استفاده می‌شود و برای تعیین پایایی آزمونهای عملکردی بایستی کلید فاکتور های مدنظر شامل شرایط محیطی ، عملکرد بیمار استاندارد و ... در نظر گرفته شود در این حالت از مدلهای پیشرفته ای نظیر مدل عمومی سازی استفاده می‌شود .

در آزمونهای مهم و سرنوشت ساز بهتر است ضریب پایایی بالای 90 % باشد در خصوص سایر آزمونها ضرایب ممکن است کمتر باشد ولی به هر حال این ضریب بهتر است برای کلیه آزمونهای بالای 70% باشد.

منابع

1. Steven M Downing . Reliability : on the reproducibility of assessment data. Medical Education 2004 ; 38 : 1006-102
2. سیف دکتر علی اکبر ، اندازه گیری ، سنجش و ارزشیابی آموزشی ویرایش سوم 1383 فصل هفدهم صفحه 447 تا 480



**EDUCATION &
DEVELOPMENT
CENTER**

آزمون کتبی

ESSAY TEST

فهرست مطالب

62	اهداف کلی	
62	1. اصول و مبانی آزمون های تشریحی	
63	1.1 تعریف آزمون های تشریحی	
63	1.2 انواع آزمون های تشریحی: آزمون گسترده پاسخ	
64	1.3 انواع آزمون های تشریحی: آزمون محدود پاسخ	
65	2. خصوصیات محاسن و معایب آزمون های تشریحی	
65	2.1 محاسن آزمون های تشریحی	

درس روشهای ارزشیابی دانشجو
جلسه هفتم

66	معایب آزمون های تشریحی	2.2
66	موارد کاربرد آزمون تشریحی	2.3
67	اصول طراحی سؤالات تشریحی	3
67	مقدمه	3.1
67	اصل اندازه گیری اعمال ذهنی	3.2
68	اصل وضوح	3.3
68	اصل اندازه گیری هدف های عالی آموزشی	3.4
68	اصل استفاده از جدول مشخصات	3.5
68	اصل ارزشیابی از سطوح بالای شناختی	3.6
69	سایر اصول	3.7
70	تصحیح سؤالات تشریحی	4
70	اصول تصحیح سؤالات تشریحی	4.1
71	شیوه های تصحیح	4.2
	خلاصه 73	
73	منابع	

اهداف کلی

در تدوین این درس اهداف کلی زیر مورد نظر است.

- انتظار می رود شما بعد از خواندن مطالب این نوشته به اهداف زیر نائل آید.
- آشنایی با:
1. تعریف و انواع آزمون های تشریحی
 2. خصوصیات محاسن و معایب آزمون های تشریحی
 3. شیوه طراحی این نوع سؤالات
 4. قواعد و روش های تصحیح سؤالات تشریحی

اصول و مبانی آزمون های تشریحی

➤ مفاهیم کلیدی

- 1- آزمون های تشریحی برای سنجش سطوح بالای اهداف یادگیری مناسب است.

شناسه جستار
عنوان جستار:
 کتبی
نویسنده: دکتر محسن صابر
 صابر
 کارشناس مسئول ارزشیابی مرکز مطالعات و توسعه آموزش پزشکی آخرین تاریخ به روز رسانی: 1388/6/26
طراح آموزشی: دکتر مانوش مهرابی
 کارشناس آموزشی دانشگاه علوم پزشکی
ویرایش: مریم فخرزاد
 کارشناس مرکز مطالعات و توسعه آموزش پزشکی
با همکاری:
 مرکز مطالعات و توسعه آموزش پزشکی
 قطب علمی آموزش الکترونیکی
 دانشگاه علوم پزشکی شیراز

- 2- سوالات تشریحی دو نوع گسترده پاسخ و محدود پاسخ دارد.
 3- برای سنجش خلاقیت و ارزشیابی باید سوالات گسترده پاسخ طراحی نماییم.

➤ اهداف یادگیری

- شما پس از مطالعه مطالب این بخش خواهید توانست به اهداف زیر نایل آید:
- 1- انواع آزمون های تشریحی و خصوصیات هر یک را تشریح کنید.
 - 2- برای هر یک از انواع آزمون های تشریحی یک سؤال طراحی نمایید.

تعریف آزمون های تشریحی

آزمون تشریحی¹ ترجمه واژه "Essay" با منشأ فرانسوی "essayer" به معنای تلاش کردن گرفته شده. این واژه پس از قرن شانزدهم با معنای انشاء یا داستان کوتاه مصطلح شد. واژه "Essay" به معنای آزمون هایی است که در آن فراگیر پاسخ سؤالات را با آزادی عمل بیان می کند. به نظر می رسد آزمون های تشریحی برای بررسی و سنجش سطوح بالای اهداف یادگیری مناسب تر است چون فراگیر با استفاده از واژه ها و توان ذهنی خود فرایند استدلالی را که به پاسخ مورد نظر استاد می انجامد را به نگارش در می آورد. در آزمون های عینی فراگیران پاسخ ها را انتخاب می کنند در حالیکه در آزمون های تشریحی پاسخ ها باید توسط فراگیر خلق و ارائه شود.

انواع آزمون های تشریحی: آزمون گسترده پاسخ

در آزمون های تشریحی گسترده پاسخ² هیچ گونه محدودیتی برای آزمون شونده منظور نمی شود و فراگیر هر طور که مایل است پاسخ خود را پرورانده و ارائه می کند. فراگیر در پاسخ دادن به سؤال های تشریحی از نظر مقدار پاسخ و زمان پاسخ دهی آزادی دارد. آزمون شونده باید اطلاعات و دانش خود را ارزیابی کند و اندیشه های خود را به نحوی منطقی، منسجم و بدیع بر روی کاغذ آورد بنابراین سؤال های آزمون های تشریحی گسترده پاسخ برای سنجش هدف ها در سطح تحلیل³، ترکیب⁴ و ارزشیابی⁵ متناظر با بالاترین سطوح طبقه بندی اهداف آموزشی در حوزه های شناختی مناسب ترین سؤال ها هستند.

این نوع سؤالات برای بررسی فرایند استدلال⁶ ذهنی فراگیر در کاربرد مطالب یاد گرفته شده در شرایط و موقعیت های خاص مناسب است. مثال:

امین و ایمان هر دو 15 ساله روزهای ابتدای بهار برای شنا به استخر در باغ رفتند آب هنوز بسیار سرد بود. امین گفت: چه کسی می

¹ Essay tests

² Extended Response

³ Analysis

⁴ Synthetic

⁵ Evaluation

⁶ Reasoning Process

تواند برای مدت طولانی تری زیر آب باقی بماند هر دو تصمیم گرفتند با هم مسابقه ای داشته باشند امین یک نفس عمیق کشید و زیر آب رفت و ایمن ده نفس عمیق کشید و به زیر آب سرد رفت. امین وقتی که دیگر نمی توانست نفس خود را نگه دارد از آب بیرون آمد و در کمال تعجب دید که ایمن در کف استخر است. امین ایمن را به کنار استخر آورد و شروع به احیاء دوستش نمود.

"پاتوفیزیولوژی اتفاقی که برای ایمن رخ داد را شرح دهید. چرا ایمن بیهوش شد و برای امین این امر رخ نداد؟"
روشن است که برای فراگیران شرایط توصیف شده امین و ایمن جدید بود و در طول بحث ارائه شده توسط استاد نیز این شرایط دقیقاً توصیف نشده بود.

سؤال قبل در حقیقت این سؤال بود.

اثر بهر را بر میزان اشباع اکسیژن در خون را شرح دهید.

چنانچه سؤال تنها برای بررسی دانش فرد طراحی شده باشد طبیعی است که شکل دوم سؤال راحت تر خواهد بود. اما در شکل اول سؤال توان و فرایند استدلال و تحلیل اطلاعات را بهتر می توان مورد ارزیابی قرار دهیم.

انواع آزمون های تشریحی: آزمون محدود پاسخ

در سؤالات تشریحی محدود پاسخ² آزمون شونده در دادن پاسخ به سؤالات آزادی کامل ندارد بلکه صورت سؤال او را ملزم می کند تا در چهارچوب خاصی پاسخ دهد. این محدودیت هم از نظر نحوه پاسخ گویی و هم از نظر زمان و میزان پاسخ است.
مثال:

برای کاهش زمان لیست انتظار در سیستم مراقبت های سلامت یک برنامه اصلاحات را بر اساس مدل تحلیلی SWOT طراحی کنید (زمان 45 دقیقه پاسخ در 300 واژه)

در این سؤال آزمون شونده محدود به جواب دادن درباره یک موضوع خاص با شیوه ی تعیین شده (بر اساس مدل تحلیلی SWOT) در زمان و حجم مشخص می شود. این محدودیت سبب می شود تصحیح پاسخ ها هماهنگ تر و آسان تر شود اگر چه بهر دلیل آزمون شونده در ارائه توانایی خود را برای ترکیب کردن اندیشه ها و بیان آنها به صورتی منسجم و منطقی محدودیت دارد. سؤال های آزمون محدود پاسخ برای اندازه گیری توانایی یادگیرندگان در سطوح فهمیدن³، بکار بستن⁴ و تحلیل⁵ مناسب است ولی برای سنجش توانایی ترکیب و ارزشیابی زیاد مناسب نیست.
نکته:

نوعی از سؤالات تشریحی محدود پاسخ، سؤالات کوتاه پاسخ است که در آن پاسخ به چند واژه محدود می شود.

مثال:

¹ Bohr Effect

² Restricted Response

³ Understanding

⁴ Apply

⁵ Analysis

محل اتصال کدام عضو بر وضعیت «آرنج تنیس بازان» مؤثر است؟

➤ آزمون

1- سوال زیر چه نوع سوالی است و چه سطحی از اهداف را می‌سنجد؟
"آموزش بصورت راند بالینی را نقد کنید."

الف) گسترده پاسخ تحلیل

ب) محدود پاسخ کاربرد

ج) گسترده پاسخ ارزشیابی

د) محدود پاسخ ترکیب

2- مدرس درس در ابتدای ترم به دانشجویان کارشناسی ارشد برای درس روش تحقیق تکلیفی را تعیین کرده بودند "نوشتن یک پورپوزال" این تکلیف چه نوع سوالات و در چه سطحی از دانشته‌های فراگیران را می‌سنجد؟

الف) گسترده پاسخ تحلیل

ب) محدود پاسخ بکار بستن

ج) گسترده پاسخ ترکیب

د) گسترده پاسخ ارزشیابی

خصوصیات محاسن و معایب آزمون‌های تشریحی

➤ مفاهیم کلیدی

- 1- آزمون تشریحی امکان‌سنجش تولید سوال را فراهم می‌کند نه انتخاب سوال را.
- 2- تصحیح آزمون تشریحی فرآیندی وقت‌گیر است.

➤ اهداف یادگیری

- شما می‌توانید پس از مطالعه مطالب این بخش به هدف زیر نایل آید:
- 1- با شناخت نقاط قوت و ضعف آزمون تشریحی شرایطی که بیشترین بهره را می‌توان از این نوع آزمون برد را تشریح کنید.

محاسن آزمون‌های تشریحی

- مدافعان این نوع آزمون محاسن آن را به شرح ذیل بر شمرده‌اند:
- تهیه آزمون‌های تشریحی از تهیه آزمون‌های عینی ساده‌تر است.
 - این نوع آزمون تنها وسیله موجود‌سنجش توانایی آزمون‌شونده در پروراندن جواب سؤال‌ها و بیان آنهاست.
 - سؤالات تشریحی توانایی پاسخ دادن به سؤال را می‌سنجد نه توانایی انتخاب پاسخ صحیح از میان سایر پاسخ‌ها را.
 - امکان روبرو نمودن آزمون‌شونده با موقعیت‌های واقعی‌تر به نسبت آزمون‌های عینی فراهم می‌شود.
 - آزمون تشریحی سبب می‌شود فراگیران عادات صحیح مطالعه و یادگیری عمیق‌تر را در خود پرورش دهند و توانایی درک و برقراری رابطه بین آموخته‌ها را به دست آورند.

معایب آزمون های تشریحی

- مخالفان کاربرد آزمون های تشریحی دلایل زیر را علیه استفاده از این آزمون ها ارائه می کنند:
- این آزمون ها نمونه کوچکی از محتوای درس و هدف های آموزشی را اندازه می گیرند.
- پایایی این نوع آزمون ها مورد تردید است.
- نمره فراگیر می تواند تحت تأثیر برداشت ذهنی استاد از دانشجو قرار گیرد.
- عوامل دیگری بجز دانسته ها و توان فراگیر در ارائه پاسخ برنمره نهایی او موثر است عواملی چون : مهارت های نگارشی ، خط و ساختار متن
- زمان زیادی برای تصحیح این نوع آزمون ها صرف می شود.

موارد کاربرد آزمون تشریحی

- صاحب نظران آموزشی معتقدند که در برخی شرایط استفاده از آزمون های تشریحی ارجح است.
- در شرایطی که تعداد دانشجویان محدود باشد و آموزش در گروه کوچک صورت پذیرد.
- زمانی که بخشی از اهداف کلیدی ما را اهداف نگرشی تشکیل داده است و بررسی این اهداف بیش از مباحث شناختی برای ما اولویت دارد.
- اگر مدرس بر توان خود برای تصحیح سوالات تشریحی بطور نقادانه بیش از توانایی طراحی سوالات عینی اعتماد داشته باشد.

➤ آزمون

- 1- آموزش دانشکده در بررسی نمرات و برگه های دانشجویان پس از اعتراض آن ها متوجه شدند که برگه های افراد خوش خط نمرات بالاتری را دریافت نموده اند آنها می خواهند دلیل لغو امتحان را اعلام کنند نقص در کدام یک از موارد ذیل دلیل لغو آزمون است.

الف) روایی

ب) انصاف و پایایی

ج) روایی و پایایی

د) انصاف و روایی

- 2- کدامیک از موارد ذیل را به عنوان دلیلی برای برگزاري آزمون پایان دوره به صورت تشریحی نمی توان بیان کرد.
- الف) آزمون تشریحی توان پروراندن پاسخ را می سنجند نه انتخاب پاسخ ها
- ب) استاد می تواند تجربه خود را از توانایی دانشجو که در طول دوره آموزش به دست آورده را برای تصحیح آزمون های آنها به کار گیرد.

- ج) سنجش اهداف كليدي را با استفاده از آزمونهاي تشریحي بهتر مي توان سنجيد.
- د) امکان روبرو نمودن فراگیر با شرایط واقعي در آزمون تشریحي بیشتر است.

اصول طراحی سؤالات تشریحي

➤ مفاهيم كليدي

- 1- اگر چه معروف است که طراحی سؤالات انشايي راحت تر است اما بايد به موارد زيادي در طراحی آن توجه نمود.
- 2- جدول اختصاص آزمون براي اين گونه آزمون ها ضروري است.
- 3- استفاده از اين نوع سؤالات را بايد به سنجش سطوح بالاي اهداف يادگيري محدود نمود.

➤ اهداف يادگيري

- شما می توانید پس از مطالعه مطالب این بخش به اهداف زیر نایل آييد:
- 1- مراحل تهیه سؤالات تشریحي را بيان کنید.
 - 2- ساختار سؤالاتي که براي سنجش سطوح بالاي اهداف است را تشریح کنید.
 - 3- مواردی که سبب افزايش اعتبار و پايایي آزمون مي شود را شرح دهید.

مقدمه

براي طراحی سؤالات انشايي بايد زمان مناسبی اختصاص داده شود. چون به علت محدودیت تعداد سؤالات، لازم است کلیه سؤالات روشن، مشخص و مناسب باشد و هیچ ابهامی در سؤالات دیده نشود. ارزش و ثمربخشي سؤالات انشايي نیز مانند انواع پرسش ها بسته به آن است که تا چه اندازه هدف ها و محتوای آموزشی را مورد ارزیابی قرار می دهد براي آنکه این نوع سؤالات از کیفیت مطلوبی برخوردار باشد و به عنوان ابزاري سودبخش براي اندازه گیری مفاهيم درسي به کار آيند بايد موارد ذیل مورد توجه قرار گیرد.

اصل اندازه گیری اعمال ذهني

طراح سؤال بايد قبل از نوشتن سؤال اعمال ذهني را که مي خواهد اندازه بگیرد را به دقت مشخص کند آنگاه سؤال را طوري بنویسد که پاسخ دادن به آن دقیقاً مستلزم انجام همان اعمال ذهني باشد. مثال:

اگر هدف آن است تا توانایی فراگیر در بکار بستن مفاهيم تحليل آزمون در طراحی سؤال آزمون را مورد سنجش قرار دهيم آنگاه سؤال بايد به گونه اي نوشته شود که نگارش پاسخ آن مستلزم همان عمل باشد.

براي انتخاب فراگیران برتر دانشکده در یکی از ورودي ها، مي خواهيم آزمونی را برگزار نماييم، تحليل آزمون و بطور اختصاصي ضريب دشواري و ضريب تمیز در این راه چه کمکی به ما مي کند.

اصل وضوح

مسأله ابهام در سؤالات در آزمون هاي تشریحي بیش از آزمون هاي عینی مطرح است سؤال هاي انشایی باید کاملاً واضح و دور از ابهام نوشته شود تا فراگیر مفهوم آن را به خوبی دریابند. در سؤالات انشایی محدود پاسخ مطالب و مسایلی که باید به آنها پاسخ داده شود باید در متن سؤال ذکر شود. طول پاسخ، زمان لازم برای جواب دادن به سؤال و نیز ملاک هاي ارزشیابی نیز باید در همه ي انواع سؤالات انشایی مشخص شود.

مثال:

سؤالات نامناسب (سؤالات مبهم و گنگ)

- روش هاي سنجشی را که تحت عنوان "Assessment tools" می آوریم را شرح دهید.
 - هر چه درباره آموزش بالینی می دانید بنویسید.
 - مدیریت مشارکتی چه نوع مدیریتی است.
- این سؤالات آنقدر کلی و مبهم هستند که اگر اساتید مختلف آن درس هم به آن پاسخ دهند تفاوت زیادی بین پاسخ هاي آنان ها مشاهده می شود.

اصل اندازه گیری هدف های عالی آموزشی

سؤالات انشایی را به اندازه گیری هدف هاي عالی تر آموزشی و فرایندهاي پیچیده تر یادگیری محدود کنید. بنابراین سؤالات بهتر است با عباراتی نظیر "مقایسه کنید"، "تجزیه و تحلیل کنید"، "تفسیر کنید"، "نقد کنید"، "ارزشیابی کنید" نوشته شوند تا با عباراتی نظیر "تعریف کنید" "توضیح دهید" "صحبت کنید".

اصل استفاده از جدول مشخصات

رسم جدول مشخصات آزمون در شرایطی که می خواهیم تمام یا قسمتی از سؤالات آزمون انشایی باشد ضروری است. اگر قصد داشته باشیم کلیه سؤالات آزمون تشریحي باشد چون تعداد سؤال ها کمتر از آزمون هاي عینی است دقت عمل در نوشتن سؤالات بیشتر ضروری است زیرا بد بودن يك سؤال در يك آزمون تشریحي پنج سؤالی خیلی بیشتر از بد بودن يك سؤال در يك آزمون عینی 50 سؤالی به نتیجه ارزشیابی آسیب می زند. چنانچه بخشی از آزمون را سؤالات تشریحي تشکیل می دهد نیز جدول اختصاصی آزمون برای تخصیص سؤال به اهداف آموزشی ضروری است. پس از تعیین جدول، مشخص خواهد بود که هر سؤال در کدام موضوع و چه سطحی از اهداف را مورد سنجش قرار می دهد.

اصل ارزشیابی از سطوح بالای شناختی

سؤالات تشریحي را تنها به سنجش هدف هایی محدود کنید که با سایر انواع سؤالات به خوبی قابل اندازه گیری نیستند. پس از طراحی جدول مشخصات آزمون موضوعاتی که باید در سطح تحلیل، ترکیب و ارزشیابی مورد سنجش قرار گیرند موضوعات مناسب برای طراحی سؤالات تشریحي گسترده پاسخ هستند، اما برای سنجش موضوعاتی که باید در سطح

کاربستن و فهمیدن بررسی شوند سؤال های محدود پاسخ مناسب خواهند بود.

سایر اصول

در حد امکان از سوال های تازه و موقعیت های جدید استفاده کنید. سؤال هایی که موضوعات آن در کتاب و یا کلاس بحث شده اند صرفاً معلومات و حافظه فراگیران را می سنجند اما استفاده از موقعیت های جدید در طرح سؤال توان کاربرد، درک و فهم را اندازه می گیرند. بنابراین اگر برای سنجش کاربستن و خلاقیت سؤالی طراحی می شود. رعایت موارد پیش گفت ضروری است:

- کاربرد کلمات "چه کسی" "چه وقت" "کجا" بیانگر سؤالاتی است که تنها معلومات را اندازه می گیرند بنابراین از این کلمات و از این نوع سؤالات بپرهیزید.
- سؤالات مربوط به موضوعات و عقاید بحث انگیز باید طوری طرح شوند که از آزمون شونده بخواهند تا شواهد لازم را برای عقیده مورد نظر بیان کند، نه اینکه از او بخواهند تا صرفاً عقاید شخصی خود را بیان کند. این نوع سؤال ها باید بطور روشن از فراگیر بخواهند تا درباره موضوع با دلایل علمی، منطقی و مستند بحث نماید.
- به فراگیران حق انتخاب از میان تعداد سؤال ندهید کلیه آزمون شوندهگان باید به تمام سؤال ها پاسخ دهند. این امر اجازه می دهد تا با کاربرد ملاکی یکنواخت به تصحیح بپردازید و پاسخ ها را بهتر با یکدیگر مقایسه کنید. اگر آزمون شوندهگان به سؤال های مختلف جواب دهند امکان مقایسه آنها با یکدیگر میسر نخواهد بود.
- در کل با وجود تمام پتانسیل ها و مشکلات این نوع سؤالات؛ استفاده محدود از این نوع سؤالات برای سنجش سطوح بالای اهداف آموزشی و همراه نمودن این نوع سؤالات با سایر انواع سؤالات عینی، عملکردی و می توان آزمون پایا و معتبر را طراحی نمود.

➤ آزمون

- 1- کدام يك از سؤالات زیر سؤال تشریحی مناسب است؟
الف) انواع آزمون های ایستگاهی را نام برده و با هم مقایسه کنید.
ب) برای سنجش استدلال بالینی از چه روش ها می توان استفاده نمود. کدام يك از روش ها به قضاوت شما مناسب تر است.
ج) برای هدف آموزشی "کارآموزان باید بتوانند از کودکان بین 3 تا 10 سال خونگیری نمایند" يك شیوه سنجش مناسب انتخاب کنید.
د) هرم میلر را توضیح دهید.
- 2- کداميك از انواع آزمونهاي زیر سنجش تکمیلی درس فیزیولوژی بهتری است؟

(دانشجویان پزشکی، حجم درس 50 صفحه کتاب گایتون)

الف) 5 سوال تشریحی

ب) 30 سوال چند گزینه ای و يك سوال تشریحی گسترده پاسخ و 3 سوال محدود پاسخ

ج) 20 سوال چند گزینه ای و 15 سوال کوتاه پاسخ و 5 صحیح و غلط

تصحیح سؤالات تشریحی

➤ مفاهیم کلیدی

- 1- در سؤالات تشریحی شیوه تصحیح در پایایی آزمون بسیار مؤثر است.
- 2- سؤالات تشریحی را باید در یک نشست و بصورت یک سؤال برای کلیه آزمون شوندهگان تصحیح شود.
- 3- دو شیوه کلی تصحیح آزمون های تشریحی وجود دارد: روش تحلیلی و کلی.

➤ اهداف یادگیری

- شما می توانید پس از مطالعه مطالب این بخش به اهداف زیر نایل آید:
- 1- دو شیوه کلی برای تصحیح آزمون های تشریحی را بیان کند و با هم مقایسه نماید.
 - 2- در فرایند تصحیح یک آزمون تشریحی باید به چه مواردی توجه نمود آنها را بیان کند.

اصول تصحیح سؤالات تشریحی

در آزمون های عینی تصحیح آزمون کاری غیر تخصصی است و توسط منشی های گروه ها، دستگاه های مارک خوان یا اسکنرها صورت گرفته و توسط نرم افزارهای محاسبه گر نیز نمره دهی و تحلیل می شود. اما در آزمون های تشریحی در حقیقت تنها مدرس مصحح آزمون است و نظر شخص تصحیح کننده اوراق، در فرایند خواندن پاسخ ها و نمره دهی مؤثر است. دقت اندازه گیری آزمون های تشریحی به شیوه تصحیح برگه های امتحانی و دقت عمل مصححان مربوط است. بنابراین برای دقت در این فرایند باید به موارد ذیل توجه نمود:

- پاسخ های فراگیران را تنها بر اساس هدفی که سؤال بر آن اساس طراحی شده است تصحیح شود. دیده می شود که فراگیرانی که در نوشته های خود از بیانی فصیح استفاده می کنند نکات دستوری و نقطه گذاری را رعایت می کنند و مطالب را با خط زیباتری می نویسند در شرایط مساوی از نظر سطح یادگیری بیش از کسانی که نکات فوق را رعایت نمی کنند نمره می گیرند اگر چه موارد پیش گفت اهمیت خود را دارد اما جستجو برای تطابق پاسخ داده شده با هدف مورد نظر، وظیفه مصححین است.
- هنگام طراحی سؤال یک کلید پیش ساخته نیز تهیه کنید این کلید در پایایی نمره گذاری کمک زیادی به مصحح می کند. در تهیه پاسخ های نمونه تلاش کنید. اجزاء پاسخ را معین نموده و برای هر قسمت امتیاز جداگانه ای منظور نمایید و به پاسخ های فراگیران بر همین اساس نمره دهید.
- پاسخ های فراگیران به سؤالات تشریحی را سؤال به سؤال تصحیح کنید یعنی ابتدا پاسخ سؤال یک تمام اوراق را تصحیح نموده و سپس به سراغ سؤال دوم بروید. این امر علاوه بر آن که کار

- نمره گذاري را راحت تر مي کند امکان به کاربردن يك ملك واحد براي كليۀ فراگيران ايجاد مي کند.
- هنگام تصحيح آزمون به اسامي فراگيران نگاه نکنيد " اين امر در صورتي که تعداد دانشجويان درحدي باشد که شما اسامي آنها را به خاطر داشته باشيد بر نمرات اثر خواهد داشت." برداشت هاي ذهني شما از توانايي دانشجويان بر نحوه تصحيح سؤالات موثر خواهد بود.
 - تمام فرايند تصحيح يك سؤال را در يك نشست و بدون وقفه زماني انجام دهيد. اين امر بدین علت توصيه مي شود که ممکن است عوامل مختلفی بر نحوه ارزشيابي شما اثر بگذارد حالات جسمي، اتفاقات غير منتظره و ... اگر ناچار شديد تصحيح تعدادي از پاسخ هاي سؤالي را که در حال انجام هستيد به نوبت ديگري موکول کنيد در آغاز شروع مجدد تعدادي از اوراق قبلي را که تصحيح کرده ايد بازخواني کنيد تا با نحوه نمره گذاري و شيوه ارزشيابي قبلي خود آشنا شويد و سؤالات مانده را به همان سبک نمره دهی کنيد.
 - به نمرات که به ساير سؤالات داده ايد نگاه نکنيد بايد نمره دهی سؤالات مختلف مستقل از هم صورت پذيرد.
 - پس از تصحيح هر سؤال برگه ها را زير و رو کنيد. گاه توالي دو برگه که يکي بسيار خوب يا بسيار بد و ديگري متوسط، تأثيرات متفاوتي را در ذهن شما ايجاد مي کند تغيير توالي برگه ها ميزان اين تأثير را کاهش مي دهد.
 - به فراگيران بر روي برگه ها بصورت مکتوب بازخورد مناسبی را نسبت به پاسخي که نوشته اند بدهيد اين امر سبب بهبود عملکرد آنها مي شود. گاه بهتر است بصورت شفاهي با تک تک دانشجويان در مورد پاسخ هاي آنها صحبت نمود و در يك تعامل دو طرفه بازخورد به دانشجو ارائه شود.
 - چنانچه نمره آزموني که برگزار نموده اي بر شرايط آتي فراگيران اثر شگرفي دارد از دو نفر از همکاران تان بخواهيد تا پاسخ هايي را که شما تصحيح نموده ايد را مجدداً تصحيح کنند اين امر سبب افزايش پايي آزمون مي شود. در اين شرايط ميانگين نمره مصححان مختلف از نمرات هر يك از آنان دقيق تر خواهد بود.

شيوه هاي تصحيح

• روش تحليلي

اين روش براي سؤالات محدود پاسخ مناسب تر است. پاسخ نمونه به اجزاء کوچک تري تقسيم شده و براي هر جزء، نمره يا امتياز جداگانه اي در نظر گرفته مي شود. در اين روش عواملی مانند "ساختار منطقي پاسخ" "فصاحت نگارش" "ذکر دلايل" امتيازات جداگانه اي دريافت مي کند.

• روش کلی

در این شیوه کل پاسخ مطالعه می شود و مصحح بر اساس کل پاسخ فرد به سؤال و با توجه به برداشت کلی از آن یک نمره می دهد. در این روش به هیچ یک از موارد به تنهایی امتیاز داده نمی شود. بلکه همه عوامل با هم دیده می شود. بطوری که برای یک سؤال نمره دوگانه رد یا قبول می دهیم یا می توان مقیاس را تا 5 واحد به شکل زیر افزایش داد. (5 خیلی خوب، 4 خوب، 3 متوسط، 2 ضعیف، 1 خیلی ضعیف) شیوه کلی تصحیح بیشتر در آزمون های تشریحی گسترده پاسخ استفاده می شود.

• مقایسه دو روش

اگر چه این دو روش را برای آزمون های وابسته به هنجار یا ملاک می توان استفاده نمود، اما روش تحلیلی برای آزمون های مبتنی بر ملاک و روش کلی برای آزمون های وابسته به هنجار مناسب تر است. با توجه به مشکلات اجرای استفاده از این دو روش بصورت مجزا بهترین شیوه بر اساس نظر بسیاری از متخصصان تلفیق این دو روش است بطوریکه ابتدا همه پاسخ ها را با روش کلی تصحیح کنید تا یک برداشت کلی نسبت به کیفیت و گستره ی آنها کسب نمایید سپس پاسخ ها را بصورت تحلیلی تصحیح نمایید این شیوه نمره گذاری ترکیبی سبب جلوگیری از معایب دو روش شده و پایایی تصحیح نمرات افزایش یابد.

➤ آزمون

1- آقای دکتر احمدی پس از طراحی 5 سؤال تشریحی پاسخ های آنها را هم نوشت و اجزاء اصلی پاسخ را هم تعیین نمود. او با این کار سبب افزایش کدام یک از خصوصیات آزمون خود شده است.

- الف) اعتبار آزمون
ب) پایایی آزمون
ج) عینیت نمره گذاری
د) همسانی نمره گذاری

2- آزمون تشریحی توسط دانشکده بهداشت برای انتخاب ورودی های جدید برای MPH در بین پزشکان برگزار شد. سؤال مطرح شده این بود "اگر شما مسئول تخصیص منابع پزشکی در استان فارس بوده اید و صد میلیون تومان بودجه داشتید این پول را چگونه هزینه می کردید". در تصحیح این سؤال از چه شیوه ای می توان بهره برد و یا به چه مواردی باید توجه کرد.

- الف) تحلیلی
ب) کلی
ج) هر دو
د) وجود اهمیت استفاده از چندین مصحح

خلاصه

سؤالات تشریحی ابزارهایی هستند که امکان سنجش سطوح بالایی اهداف آموزشی مانند کاربستن، تحلیل، ترکیب، ارزشیابی را فراهم می‌کند. در طراحی سؤالات تشریحی باید به مواردی مانند تدوین جدول اختصاصی آزمون، روشن و شفاف بودن سؤال و جدید بودن شرایط مطرح شده در سؤال توجه نمود. در این صورت است که با وجود دشواری‌های تصحیح این نوع سؤالات ابزار مناسبی برای سنجش سطوحی از استدلال، استنتاج و نقد خواهند بود که با سایر روش‌ها قابل سنجش نخواهد بود. در تصحیح سؤالات تشریحی می‌توان با استفاده از 2 یا 3 مصحح و استفاده از شیوه‌های تحلیل یا ترکیبی پایایی نمرات و آزمون را افزایش داد.

منابع

1. سیف، علی اکبر. اندازه‌گیری، سنجش و ارزشیابی آموزشی، ویرایش چهارم، 1385.
2. شریفی، دکتر حسن پاشا. اصول روان‌سنجی و روایی آزمای، چاپ دهم، 1385.

3. Cashin W.E. Improving Essay Tests. Kansan state university.
4. Norton L.S. Essay-Writing: what really counts? Highest Education 20: 411-442.1990.
5. John A.Dent, Ronald M.Handen. Apractical guide for. Medical teachers. Second education. 2005.