

بلاغت

سنجش و اندازه‌گیری در تعلیم و

تربیت

فهرست مطالب

<u>صفحه</u>	<u>عنوان</u>
۴	فصل اول: مفهوم‌شناسی و تعاریف سنجش و اندازه‌گیری
۱۲	فصل دوم: مقیاس‌های اندازه‌گیری
۱۸	فصل سوم: دسته‌بندی آزمون‌های مورد استفاده در روانشناسی و آموزش
۲۶	فصل چهارم: دسته‌بندی ارزشیابی‌های آموزشی
۴۲	فصل پنجم: مراحل و فعالیت‌های ارزشیابی آموزشی
۴۹	فصل ششم: ارزشیابی پیشرفت تحصیلی
۵۵	فصل هفتم: انواع آزمون‌های پیشرفت تحصیلی
۶۸	فصل هشتم: روش‌های سنجش فرایندها و فرآورده‌های غیرشناختی
۷۷	فصل نهم: روش‌های سنجش مستقیم رفتار
۸۱	فصل دهم: اجرا، نمره‌گذاری و تحلیل آزمون
۹۳	فصل یازدهم: استفاده از روش‌های آماری برای تفسیر نمره‌های آزمون
۱۲۳	فصل دوازدهم: هنجارها و نیمرخ‌ها
۱۲۹	فصل سیزدهم: روایی و پایایی آزمون
۱۴۵	فصل چهاردهم: ارزشیابی معلم
۱۵۲	منابع

فصل اول:

مفهوم شناسی و تعاریف سخن و اندازه‌گیری

مقدمه

قبل از ورود به بحث اصلی و پرداختن به ابعاد مختلف این درس در آغاز سخن، مفاهیم و اصطلاحات به کار رفته در عنوان درس و مفاهیم مرتبطی چون سخن، اندازه‌گیری، آزمون، و ارزشیابی را مفهوم‌شناسی کرده و تعریف روشنی از آنها بدست آوریم؛ و به ارتباط بین آنها و نقاط مشترک و تفاوت آنها اشاره نماییم.

تعریف اندازه‌گیری^۱

در اندازه‌گیری مشخص می‌گردد که یک شخص یا یک چیز چه مقدار از یک ویژگی برخوردار است. در اندازه‌گیری ویژگی یا صفتی که اندازه‌گیری می‌شود حتماً به صورت عدد یا رقم مشخص می‌گردد. بنابراین و به عقیده‌گی اندازه‌گیری فرایند تعیین میزان برخورداری یک شخص یا شیئی از یک ویژگی خاص است. البته تعریف دقیق‌تر از اندازه‌گیری به کاربرد قواعد نیز تأکید دارد و اندازه‌گیری را دربرگیرنده قواعدی می‌داند برای نسبت دادن اعداد به

^۱measurement

اشیا به گونه‌ای که صفاتی از آنها را به صورت کمیت نشان دهد یا به صورت ساده‌تر، اندازه‌گیری اطلاق نمره یا عدد به خصوصیتی طبق قواعد و ضوابط خاص می‌باشد. البته این مسأله درباره اندازه‌گیری ویژگی‌های روانی مانند یادگیری صدق می‌کند.

تعاریفی از اندازه‌گیری آموزشی

نیتکو^۱ اندازه‌گیری را شیوه نسبت دادن یک عدد به یک صفت یا ویژگی معین یک شخص به گونه‌ای که آن عدد درجه‌ای از آن صفت یا ویژگی را که آن فرد از آن برخوردار است نمایان سازد.

نانالی^۲ در تعریف اندازه‌گیری بر کاربرد قواعد تاکید می‌کند. مقصود از کاربرد قواعد این است که روش‌های نسبت دادن اعداد باید بطور واضح و روشن بیان شوند. به عبارتی قواعد اندازه‌گیری خط مشی‌هایی هستند برای نشان دادن مقدار شئی مورد اندازه‌گیری.

بیان قواعد اندازه‌گیری از جنبه‌های مهم استاندارد کردن یا میزان کردن آزمون‌های روانی و تربیتی است تا افراد مختلفی که این وسایل اندازه‌گیری را به کار می‌برند نتایج یکسانی بدست آورند.

تعریف آزمون یا تست^۳

آزمون معمول‌ترین وسیله اندازه‌گیری ویژگی‌ها یا صفات روانی است. اندازه‌گیری یک فعالیت یا فرایند است که نیاز به وسیله‌ای دارد که به آن وسیله، وسیله اندازه‌گیری می‌گویند. کاربرد اندازه‌گیری مشخص کردن ویژگی‌های مختلف اشیاء است. اندازه‌گیری مستقیم برای اندازه‌گیری ویژگی‌های فیزیکی (جسمی) و رفتارهای آشکار که بطور مستقیم اندازه‌گیری می‌شود مانند قد یا وزن به کار برده می‌شود. اما وسایل اندازه‌گیری غیر مستقیم برای ویژگی‌ها تا صفات روانی مانند یادگیری و هوش که بطور مستقیم قابلیت دسترسی ندارند به کار می‌روند.

آزمون وسیله و روش نظام‌دار (سیستماتیک) برای اندازه‌گیری نمونه‌ای رفتار است. نکته مورد تأکید در این رابطه این است که آزمون کل یک ویژگی را اندازه‌گیری نمی‌کند. مثلاً در درس ریاضی معلم نمی‌تواند از همه مطالب درس سؤال طرح کند. لذا باید نمونه‌ای از محتوای درس را انتخاب کند و از آن سؤال طرح نماید. برای انجام این کار روش رایج و مورد استفاده برای معلم بکارگیری جدول مشخصات درس است. استفاده از این جدول مشخصات باعث می‌شود آزمون دارای اعتبار محتوایی^۴ باشد و به علاوه اگر نمونه‌های انتخاب شده در آزمون بیشتر معرف ویژگی روانی مورد نظر باشد به همان میزان نتایج اندازه‌گیری دقیق‌تر است. معروف‌ترین وسیله اندازه‌گیری

^۱Nitko

^۲Nunnally

^۳test

^۴Content Validity

در روانشناسی و آموزش و پرورش، آزمون است که به صورت مجموعه‌ای سؤال و غالباً به طور کتبی برای پاسخگویی در اختیار آزمون‌شوندگان قرار داده می‌شود.

آزمون در معنای محدود ممکن است به کاربرد مجموعه سئوالات چندگزینه‌ای گفته شود لکن در معنای وسیع آزمون مشتمل بر انواع سئوالات عینی، تشریحی و کوتاه پاسخ است. در کنار این منظور اصطلاح آزمون برای شیوه‌های دیگر اندازه‌گیری نظیر پرسشنامه‌ها کاربرد پیدا می‌کند. به عبارت بهتر در کاربرد اصطلاح آزمون برای اندازه‌گیری تمامی ویژگی‌های روانی و تربیتی عملاً هیچ گونه محدودیتی وجود ندارد. در واقع آزمودن نوعی خاص از اندازه‌گیری است. ممکن است اندازه‌گیری صورت پذیرد اما آزمونی در کار نباشد.

تعریف آزمودن^۱

بنا به تعریف، وقتی که برای اندازه‌گیری یکی از ویژگی‌های روانی و تربیتی یک فرد یا گروهی از افراد از آزمون استفاده می‌شود. به این فعالیت آزمون می‌گویند. اندازه‌گیری متغیرهای روانشناختی به وسیله تدابیر یا روش‌هایی که به منظور کسب نمونه‌هایی از رفتار طراحی شده‌اند، آزمون روانی یا روان‌آزمایی نام دارد. از آنجا که آزمون یک وسیله اندازه‌گیری است، هر وقت آزمودن صورت گیرد یعنی آزمونی به کار رود نوعی اندازه‌گیری انجام می‌شود، اما هر نوع اندازه‌گیری الزاماً آزمودن نیست. اندازه‌گیری از آزمون مفهوم گسترده‌تری دارد، زیرا آزمودن نوع خاصی از اندازه‌گیری است. ممکن است اندازه‌گیری صورت پذیرد اما آزمونی در کار نباشد؛ یعنی اندازه‌گیری می‌تواند با استفاده از وسیله دیگری به جز آزمون انجام شود. مثال:

«اندازه‌گیری رفتار پرخاشگرانه دانش آموز از راه مشاهده و با شمردن تعداد دفعات رفتار توسط معلم»

تعریف سنجش^۲

سنجش مفهومی نزدیک به اندازه‌گیری و آزمون است لکن از این دو گسترده‌تر است. به عنوان مثال در نگاه نیتکو سنجش یک اصطلاح کلی است و به صورت فرایند تعریف می‌شود و برای گردآوری اطلاعات مورد نیاز تصمیم‌گیری درباره دانش‌آموزان، برنامه‌های درسی و دیگر عناصر نظام آموزشی مورد استفاده قرار می‌گیرد. زمانی که در اندازه‌گیری به جای آزمون یا علاوه بر آن وسایل دیگری هم به کار می‌رود به آن سنجش می‌گویند. مثلاً در روانشناسی بالینی برای تعیین میزان افسردگی یک فرد هم از آزمون افسردگی و هم از مصاحبه استفاده می‌کنند.

اصطلاح سنجش در حوزه روانشناسی بالینی و آموزش و پرورش بیشتر به کار می‌رود ولی سنجش نیز نوعی اندازه‌گیری است.

عده‌ای تعریف فوق را برای سنجش کامل نمی‌دانند و معتقدند که سنجش در گفتار فارسی به دو صورت به کار می‌رود. در حالت اول سنجش نوعی اندازه‌گیری است مانند دماسنج و در حالت دوم سنجش شامل اندازه‌گیری و هم شامل قضاوت کردن و ارزشیابی است. مانند سنجیده سخن گفتن.

هنگامی اندازه‌گیری صرفاً با استفاده از آزمون انجام می‌شود آزمون اتفاق می‌افتد لکن در سنجش وسایل و فنون مختلف و متفاوتی برای گردآوری اطلاعات بکار می‌رود. مانند پرسشنامه، فهرست واری و...

در دیدگاه کرونباخ سنجش مشتمل بر بکارگیری فنون مختلف است و تاکید زیادی بر مشاهده عملکرد دارد و سنجش نوعی تحلیل بالینی و پیش‌بینی عملکرد است در صورتی که آزمون یک ابزار و یک وسیله اندازه‌گیری است. یکی دیگر از تفاوت‌های موجود مابین سنجش، اندازه‌گیری و آزمون این است که اندازه‌گیری و آزمون با کمیت (اعداد و ارقام) سروکار دارند اما سنجش الزاماً به کمیت منحصر نمی‌شود. (توصیف).

تفاوت سنجش و اندازه‌گیری و آزمون این است که هم اندازه‌گیری و هم آزمون عموماً با کمیت سر و کار دارند. اما سنجش الزاماً به کمیت منحصر نمی‌شود. به عبارتی دیگر زمانی که اندازه‌گیری یا آزمون صورت می‌پذیرد، نتیجه به صورت اعداد یا ارقام نشان داده می‌شود، اما نتیجه سنجش می‌تواند به صورت غیر کمی گزارش شود.

ارزشیابی^۱

ارزشیابی در واقع تعیین ارزش برای هر چیزی یا داوری ارزشی کردن می‌باشد، اما به طور جامع‌تر ارزشیابی به یک فرآیند نظام‌دار (سیستماتیک) برای جمع‌آوری، تحلیل و تفسیر اطلاعات گفته می‌شود. (اندازه‌گیری) به این منظور تعیین شود که آیا هدفهای مورد نظر تحقق یافته‌اند یا در حال تحقق یافتن هستند و به چه میزانی، به کار می‌رود.

یکی از ویژگی‌های مهم ارزشیابی، تعیین کیفیت است. در فرآیند ارزشیابی داوری ارزشی با توجه به کیفیت به عمل می‌آید. کیفیت را در ارزشیابی پیشرفت تحصیلی به عنوان دانش، مهارت‌ها و توانایی‌هایی که از دانش‌آموزان پس از آموزش انتظار می‌رود. تعریف کرده‌اند.

ارزشیابی پیشرفت تحصیلی

ارزشیابی پیشرفت تحصیلی مفهومی بسیار نزدیک به مفهوم سنجش یادگیری است، اما ارزشیابی آموزشی در معنایی کلی معنایی بسیار وسیع دارد. در آموزش و پرورش ارزشیابی به یک فعالیت رسمی گفته می‌شود که برای تعیین کیفیت، اثربخشی، یا ارزش یک برنامه، فرآورده، پروژه، فرایند، هدف یا برنامه درسی به اجرا در می‌آید.

مقایسه اندازه‌گیری، آزمون، سنجش و ارزشیابی

- در اندازه‌گیری صرفاً از آزمون استفاده می‌شود، اما در سنجش وسایل و فنون مختلفی برای جمع‌آوری اطلاعات به کار می‌رود.

- سنجش نسبت به آزمون و اندازه‌گیری مفهوم گسترده‌تری دارد و بیشتر در حوزه روان‌شناسی بالینی و آموزش و پرورش به کار می‌رود.

- اندازه‌گیری و آزمون عموماً با کمیت سر و کار دارند، اما سنجش الزاماً به کمیت منحصر نمی‌شود.

- ارزشیابی شامل داوری ارزشی درباره کیفیت موضوع ارزشیابی است که مبتنی بر تحلیل و تفسیر اطلاعات جمع‌آوری شده است. اطلاعات مورد نیاز ارزشیابی آموزشی به کمک اندازه‌گیری و عموماً آزمون و سنجش گردآوری می‌شوند.

پس فرق ارزشیابی و اندازه‌گیری این است که در اندازه‌گیری با استفاده از آزمون و سنجش ویژگی یا صفتی اندازه‌گیری می‌شود اما قضاوتی صورت نمی‌گیرد، ولی در ارزشیابی قضاوت یا داوری صورت می‌گیرد پس اندازه‌گیری پایه و اساس ارزشیابی درست و صحیح را فراهم می‌سازد.

ارزشیابی آموزشی و پژوهش آموزشی

هم پژوهش آموزشی و هم ارزشیابی آموزشی جزء فعالیت‌های علمی به حساب می‌آیند، زیرا هر دو قویاً بر روش‌ها و فنون بررسی تجربی متکی‌اند. هرچند که متخصصان تحقیقات و ارزشیابی آموزش، پژوهش آموزشی و ارزشیابی آموزشی را به صورت فرآیند جستجوی منظم برای مشخص کردن موقعیت نامعین تعریف می‌کنند، با وجود این بین پژوهش آموزشی و ارزشیابی آموزشی تفاوت‌هایی وجود دارد که عمده‌ترین آنها عبارتند از:

(الف) هدف ارزشیابی به طور عمده تصمیم‌گیری است ولی هدف پژوهش استنتاج است.

(ب) در ارزشیابی آموزشی بر یک پدیده آموزشی ویژه تمرکز صورت می‌گیرد و تعمیم‌پذیری نقشی ندارد در حالی که تعمیم‌پذیری نتایج تحقیق از اهمیت بالایی برخوردار است و نشان‌دهنده اعتبار بیرونی تحقیق است.

(ج) ارزشیابی آموزش با ارزش سر و کار دارد و در حالی که پژوهش خالی از ارزش است.

(د) پژوهش وابسته به نظریه است. ولی ارزشیابی الزاماً مبنای نظری ندارد. هدف اصلی پژوهش تولید دانش تازه و افزودن آن به نظریه‌های موجود است.

پژوهش ارزشیابی^۱

پژوهش ارزشیابی، ارزشیابی است که بر پایه شواهد تجربی جمع‌آوری شده به راه‌هایی که قابل تکرار باشند استوار است و در آن مسائل مربوط به روایی درونی، روایی بیرونی و روایی سازه منظور می‌شود. به عبارت دیگر پژوهش ارزشیابی عبارت از کاربرد روش‌هایی پژوهش در مسایل ارزشیابی است.

نکات مهم و کلیدی :

- اندازه‌گیری فرایند یا شیوه‌ای دقیق و نظام‌دار است.
- نانالی در تعریف اندازه‌گیری بر کاربرد قواعد تاکید می‌کند.
- قواعد اندازه‌گیری خط مشی‌هایی هستند برای نشان دادن مقدار شئی مورد اندازه‌گیری.
- از مشخصات مهم آزمون‌های استاندارد یا میزان شده این است که افراد مختلفی که این وسایل اندازه‌گیری را به کار می‌برند نتایج یکسانی بدست آورند.
- سنجش کلاسی! تمامی فرایندهای مورد نیاز برای تصمیم‌گیری در مورد یادگیری دانش‌آموزان.
- سنجش یک تحلیل جامعه و چندوجهی از عملکرد است. (ویگینز).
- سنجش نوعی تحلیل بالینی و پیش‌بینی عملکرد است.
- سنجش تنها به گردآوری داده‌ها یا اطلاعات خلاصه نمی‌شود، بلکه به جمع‌آوری داده‌ها همراه با قصد یا منظور گفته می‌شود. به نظر متخصصان معتقد بر این نگاه سنجش در آموزش و پرورش به دو منظور انجام می‌گیرد:
 ۱. تعیین و تشخیص مشکلات دانش‌آموزان؛
 ۲. تصمیم‌گیری درباره آنان
- در ارزشیابی آموزشی کیفیت به صورت تناسب داشتن موضوع مورد ارزشیابی با هدفهای مورد نظر تعریف می‌شود.
- ارزشیابی آموزشی و پژوهش آموزشی از جهت هدف و منظور باهم فرق دارند.
- در ارزشیابی آموزشی بر موقعیت موجود و تصمیماتی که درباره آن می‌توان گرفت تاکید می‌شود.
- هدف ارزشیابی به طور عمده تصمیم‌گیری است و هدف پژوهش استنتاج است.
- هدف پژوهش استنتاج است. و در پژوهش آموزشی تعمیم‌پذیری نتایج از اهمیت خاصی برخوردار است.
- مطالعات ارزشیابی عموماً به منظور حل مشکلات عملی ویژه و کمک به تصمیم‌گیری صورت می‌گیرد.
- نقطه مشترک ارزشیابی آموزشی و پژوهش آموزشی بررسی تجربی است.
- ارزشیابی با ارزش سر و کار دارد در حالیکه پژوهش خالی از ارزش است.

فصل دوم: مقیاس‌های اندازه‌گیری

مقیاس‌های اندازه‌گیری^۱

در اندازه‌گیری به ویژگی‌ها یا صفات اشخاص اعدادی نسبت داده می‌شود اما با توجه به اینکه چه چیزی را اندازه می‌گیریم و چگونه آن را اندازه‌گیری می‌کنیم اعداد ویژگی‌های متفاوتی دارند. این ویژگی‌های متفاوت اعداد را به طور مختلفی اندازه‌گیری می‌کنند و در این رابطه از مقیاس‌های مختلف اندازه‌گیری که شامل مقیاس‌های اسمی، رتبه‌ای (ترتیبی)، فاصله‌ای و نسبی یا نسبتی است، استفاده می‌گردد. این طبقه‌بندی مقیاس‌ها از اسمی به سمت نسبی از ساده به پیچیده است و هر طبقه دارای ویژگی‌های طبقه‌های قبلی به علاوه ویژگی‌های مخصوص به آن طبقه است.

تعریف مقیاس و مقیاس اندازه‌گیری

مقیاس مجموعه‌ای اعداد است که ویژگی‌های آنها بر ویژگی‌های تجربی اشیایی که اعداد به آنها نسبت داده می‌شوند منطبق است. مقیاس‌های اندازه‌گیری را به چهار سطح ۱- اسمی، ۲- ترتیبی، ۳- فاصله‌ای و ۴- نسبی طبقه‌بندی می‌کنند. در ادامه بحث به مفهوم‌شناسی هر کدام از این مقیاس‌ها و کاربردها و محدودیت‌های آنها می‌پردازیم.

^۱Scales of measurement

سطح الف) مقیاس اسمی^۱

این مقیاس پایین‌ترین و ساده‌ترین نوع اندازه‌گیری است. مقیاس اسمی به دو صورت مورد استفاده قرار می‌گیرد:

۱- برای اسم‌گذاری و ۲- برای طبقه‌بندی. اسم‌گذاری (شماره تلفن) و طبقه‌بندی مانند عدد یک برای مرد و صفر برای زن به کار می‌رود. اعداد در این مقیاس معنای کمیتی ندارند. لذا نمی‌توان هیچ یک از عملیات مجاز ریاضی را از جمله جمع، تفریق و ضرب و تقسیم را در ارتباط با آن اعداد به کار برد. مثلاً هیچ‌کس شماره پیراهن بازیکنان را با هم جمع نمی‌کند. و... و یا اینکه کسی که شماره پیراهنش ۲ می‌باشد و کسی که ۲۴ می‌باشد از لحاظ کمیتی تفاوتی با هم ندارند. پس نتیجه می‌گیریم که این مقیاس صرفاً برای نام‌گذاری و شناسایی افراد یا اشیاء مختلف به کار برده می‌شود تنها عملیات مجاز آماری تعیین نما برای این مقیاس است و لذا نمی‌توان از اعداد اسمی میانه و میانگین و واریانس گرفت.

در تعریف اندازه‌گیری فهمیدیم که استفاده از قواعد اندازه‌گیری ضروری است. لذا در مقیاس اندازه‌گیری هم باید از قواعد مذکور استفاده شود اما مقیاس اسمی فاقد قواعد اندازه‌گیری است و متخصصان اندازه‌گیری به همین دلیل استفاده از مقیاس اسمی را به عنوان مقیاس اندازه‌گیری درست نمی‌دانند.

- عملیات مجاز آماری درباره مقیاس اسمی: شمارش فراوانی، تعیین نما

- عملیات مجاز ریاضی درباره مقیاس اسمی: انجام هیچ یک از چهار عمل اصلی در این مقیاس میسر نیست.

سطح ب) مقیاس ترتیبی^۲ (رتبه‌ای)

همانطور که از نام این مقیاس استنباط می‌شود برای رتبه‌بندی افراد یا اشیاء مختلف از لحاظ دارا بودن یک صفت یا ویژگی به کار برده می‌شود. در این مقیاس به افراد و اشیا رتبه داده می‌شود و لذا اگر هدف اندازه‌گیری انتخاب بهترین یا مناسب‌ترین شیء و فرد باشد، این مقیاس مناسب است. مثلاً رده‌بندی تیم‌ها در مسابقات لیگ آزادگان برحسب تیم‌های رده اول و دوم تا دوازدهم که به هر تیم یک رتبه داده می‌شود. یا برندگان مسابقات کشتی از نفر اول تا سوم. مثال معمول آن شاگرد اول، دوم و سوم در کلاس می‌باشد.

در مقیاس ترتیبی عملیات بیشتری از مقیاس اسمی مجاز است. با این حال عملیات مورد نیاز یک مقیاس خوب اندازه‌گیری در این مقیاس هم مجاز نیست.

^۱Nominal Scale

^۲Ordinal Scale

- عملیات مجاز آماری درباره مقیاس ترتیبی: شمارش فراوانی، تعیین نما، محاسبه میانه، محاسبه درصدها، و محاسبه ضریب همبستگی رتبه‌ای اسپیرمن.

- عملیات مجاز ریاضی درباره مقیاس ترتیبی: انجام هیچ یک از چهار عمل اصلی در این مقیاس میسر نیست. در این مقیاس محاسبات ریاضی مانند جمع و تفریق، ضرب و تقسیم را نمی‌توان انجام داد. هیچ کس رتبه نفر سوم را ضرب نمی‌کند.

سطح ج) مقیاس فاصله‌ای^۱

در این مقیاس هم رتبه اشیا با توجه به یک صفت مشخص است و هم معلوم است که اشیا یا افراد از لحاظ صفت مورد اندازه‌گیری چه مقدار از یکدیگر فاصله دارند. اما هیچ نوع اطلاعاتی راجع به مقدار صفت مورد نظر برای هیچ یک از اشیا یا افراد در دست نیست چنانکه از این طریق معلوم است، تمایز این مقیاس با مقیاس ترتیبی در این است که در این مقیاس فواصل بین واحدها معلوم و برابر هستند، در صورتی که در مقیاس ترتیبی فاصله بین واحدها معلوم نیست. مثلاً نمراتی که دانش‌آموزان سال دوم ابتدایی از درس ریاضی می‌گیرند نشان می‌دهد که چه کسانی بالاترین و پایین‌ترین نمرات را گرفته‌اند. (رتبه‌بندی امکان‌پذیر است) و هم اختلاف نمرات آنها مشخص است. (فاصله بین نمرات) به علاوه اینکه برای فردی که نمره‌اش در امتحان ریاضی صفر شده است نمی‌توان گفت که او اصلاً ریاضی نمی‌داند. لذا در این مقیاس می‌گویند، مقدار مطلق صفر وجود ندارد.

چون در مقیاس فاصله‌ای فواصل بین واحدها معلوم و برابر است و مقیاس فاصله‌ای از ویژگی‌های مهم یک مقیاس خوب اندازه‌گیری برخوردار است لذا همه محاسبات آماری و غالب محاسبات ریاضی را درباره آن می‌توان انجام داد.

- عملیات مجاز آماری درباره مقیاس ترتیبی: عملیات مجاز آماری در این مقیاس عبارتند از محاسبه نما، میانه، میانگین، واریانس، انحراف معیار، ضریب همبستگی رتبه‌ای و ضریب همبستگی پیرسون.

- عملیات مجاز ریاضی درباره مقیاس ترتیبی: در این مقیاس، عملیات جمع و تفریق مجاز است ولی ضرب و تقسیم مجاز نیست.

سطح د) مقیاس نسبی^۲ (نسبتی)

در این مقیاس رتبه اشخاص با توجه به یک صفت معلوم است. فاصله بین اشخاص معلوم است و به علاوه فاصله حداقل یکی از اشخاص از یک صفر منطقی نیز معلوم است. در این مقیاس صفر مطلق وجود دارد.

^۱Interval Scale

^۲Ratio Scale

فصل دوم : مقیاس‌های اندازه‌گیری / ۱۵

مهم‌ترین مقیاس نسبتی همان مقیاس اعداد از صفر تا بی‌نهایت است. البته اندازه‌گیری متغیرهای روانی و تربیتی مانند هوش و یادگیری با این مقیاس به ندرت امکان‌پذیر است. مثلاً وزن و قد را در این مقیاس می‌توان مشخص کرد که یک چیز اصلاً وزن ندارد. در حالی که نمی‌توان گفت که کسی اصلاً هوش ندارد و نمره هوشی او صفر است.

- عملیات مجاز آماری درباره مقیاس نسبتی: تمام عملیات‌های آماری با این مقیاس امکان‌پذیر است.

- عملیات مجاز ریاضی درباره مقیاس نسبتی: تمام عملیات‌های ریاضی با این مقیاس امکان‌پذیر است.

مقیاس	وجود ترتیب در طبقات	فواصل مساوی طبقات	وجود صفر مطلق	عملیات مجاز آماری	عملیات مجاز ریاضی
اسمی	-	-	-	نما - فراوانی	هیچکدام
رتبه‌ای	+	-	-	نما - فراوانی - میانه - همبستگی اسپیرمن	هیچکدام
فاصله‌ای	+	+	-	همه عملیات آماری	جمع و تفریق
نسبی	+	+	+	همه عملیات آماری	همه عملیات ریاضی

جدول مقایسه مقیاس‌های چهارگانه اندازه‌گیری

مقیاس‌های اندازه‌گیری پیوسته و گسسته

در کنار مقیاس‌های چهارگانه فوق در یک طبقه‌بندی دیگر مقیاس‌ها را به دو دسته دیگر هم تقسیم می‌کنند.

۱. مقیاس اندازه‌گیری گسسته؛

۲. مقیاس اندازه‌گیری پیوسته؛

الف) مقیاس اندازه‌گیری گسسته

در مقیاس گسسته متغیر مورد نظر تنها می‌تواند هر یک از ارزش‌های معین را به خود اختصاص دهد. ویژگی مهم این مقیاس تساوی واحدهای شمارش آنها در طول مقیاس مثل تأهل، جنس، تحصیلات و تعداد فرزندان است. نکته کنکوری این است که هرچه وسایل اندازه‌گیری دقیق‌تر شود، نتیجه فرقی نمی‌کند.

انجام تمام عملیات اصلی ریاضی مانند جمع و تفریق و ضرب و... با این مقیاس اندازه‌گیری مجاز است. تعداد فرزندان خانواده و جنسیت مثال‌هایی از مقیاس گسسته است. مقیاس گسسته الزاماً با اعداد درست سروکار ندارد.

ب) مقیاس اندازه‌گیری مقیاس پیوسته

مقیاس پیوسته مقیاسی است که متغیر مورد نظر می‌تواند در فاصله بین نقاط مقیاس، هر تعدادی ارزش را به خود اختصاص دهد، مانند وزن. نکته مهم این است که هرچه وسیله دقیق‌تر باشد، وزن فرق می‌کند و دقیق‌تر است. از جمله ویژگی‌های دیگر مقیاس‌های پیوسته تساوی واحدهای اندازه‌گیری است.

نکته قابل توجه اینکه تنها اعداد متغیرهای پیوسته دارای حدود واقعی هستند. حدود واقعی یک متغیر پیوسته برابر است با آن عدد به اضافه یا منهای نصف واحد اندازه‌گیری، مثلاً واحد اندازه‌گیری وزن شخصی ۷۰ کیلوگرم باشد، حدود واقعی آن $69/5$ تا $70/5$ کیلوگرم است. با توجه به تعریف بالا متغیرهای پیوسته می‌توانند شامل اعداد اعشاری باشند. قاعده کلی این است که تا سه رقم اعشار عملیات ریاضی را پیش برده و تا دو رقم گرد می‌نماییم. مثلاً اگر در گرد کردن رقم بعد از رقم مورد نظر بیشتر از ۵ باشد. به رقم مورد نظر یک رقم می‌افزاییم. ولی اگر کمتر باشد، آن رقم تغییر نمی‌کند که در این رابطه توضیحات بیشتر در کتاب‌های آماری آمده است.

حدود واقعی اعداد

حدود واقعی ارزش عددی یک متغیر پیوسته و برابر است با آن عدد به اضافه منهای نصف واحد اندازه‌گیری. مثلاً حد واقعی ۶۱ برابر است با $61/5$ و $60/5$. همچنین اعدادی که دارای اعشار هستند، نصف واحد اندازه‌گیری از همان عدد اعشار اضافه و منها می‌شود.

مثال: حد واقعی عدد $5/5$ به این صورت است: $5/5$ منهای $0/5$ و به اضافه $0/5$: $5/5 - 0/5$

نکات مهم و کلیدی:

- ویژگی‌های متفاوت افراد را مقیاس اندازه‌گیری گویند.
- انجام هیچ یک از چهار عمل اصلی در مقیاس اسمی و ترتیبی میسر نیست.
- در مقیاس فاصله‌ای فواصل بین واحدها معلوم و برابر هستند، در صورتی که در مقیاس ترتیبی فاصله بین واحدها معلوم نیست.
- مشکل مقیاس اسمی از نظر دانشمندان این است که فاقد قواعد اندازه‌گیری است.
- در مقیاس فاصله‌ای مقدار مطلق صفر وجود ندارد.
- مقیاس فاصله‌ای از ویژگی‌های مهم یک مقیاس خوب اندازه‌گیری برخوردار است.
- در مقیاس نسبتی صفر مطلق وجود دارد.
- مقیاس ترتیبی شکل ضعیفی از مقیاس فاصله‌ای است. و بروگاتا و بوهر عقیده دارند که بهتر است مقیاس ترتیبی را به صورت مقیاس فاصله‌ای ناکامل تصور کنیم.
- مهم‌ترین مقیاس نسبتی همان مقیاس اعداد از صفر تا بی‌نهایت است.
- ویژگی مهم مقیاس گسسته تساوی واحدهای شمارش آنها در طول مقیاس است.
- در مقیاس گسسته هرچه وسایل اندازه‌گیری دقیق‌تر شود، نتیجه فرقی نمی‌کند.
- در مقیاس پیوسته هرچه وسیله دقیق‌تر باشد، نتیجه فرقی می‌کند. از جمله ویژگی‌های دیگر مقیاس‌های پیوسته تساوی واحدهای اندازه‌گیری است.
- حدود واقعی ارزش عددی یک متغیر پیوسته برابر است با آن عدد به اضافه یا منهای نصف واحد اندازه‌گیری.
- اندازه به دست آمده از متغیر پیوسته همیشه تقریبی است.
- اندازه‌گیری به دست آمده از مقیاس گسسته همواره دقیق و قطعی است.
- نمونه‌های مقیاس گسسته: جنسیت، وضعیت تاهل، سطح تحصیلات
- نمونه‌های مقیاس پیوسته: زمان، طول، وزن، هوش، استعداد
- عملیات ریاضی مقیاس گسسته: جمع و تفریق، ضرب و تقسیم

فصل سوم:

دسته‌بندی آزمون‌های مورد استفاده در روانشناسی و

آموزش

دسته‌بندی آزمون‌ها

همان‌طور که قبلاً گفته شد متداول‌ترین وسیله اندازه‌گیری در روانشناسی و آموزش و پرورش آزمون است. انواع آزمون‌های مورد استفاده در روانشناسی و آموزش و پرورش را از جهات مختلف می‌توان دسته‌بندی کرد.

۱- دسته‌بندی آزمون‌ها با توجه به ویژگی مورد اندازه‌گیری

- آزمون‌های توانایی
- آزمون‌های عاطفی

۲- دسته‌بندی آزمون‌ها با توجه به چگونگی تهیه آنها

- آزمون‌های میزان شده یا استاندارد شده
- آزمون‌های معلم ساخته

الف- دسته‌بندی آزمون‌ها با توجه به ویژگی مورد اندازه‌گیری

آزمون‌های مختلف را می‌توان با توجه به ویژگی‌های گوناگونی که اندازه می‌گیرند به دو دسته کلی آزمون‌های توانایی و آزمون‌های عاطفی تقسیم کرد. هرکدام از این دو دسته آزمون نیز به اجزای کوچکتری تقسیم می‌شوند.

۱- آزمون‌های توانایی^۱

آزمون‌های توانایی آموخته‌ها، مهارت‌ها و استعداد‌های افراد را می‌سنجد. دسته‌ای از این آزمون‌ها توانایی‌های شناختی^۲ افراد را می‌سنجد و تعدادی دیگر توانایی‌های روانی و حرکتی^۳ را اندازه‌گیری می‌کنند. ویژگی مهم آزمون‌های توانایی (شناختی و روانی حرکتی) این است که هنگام اجرای آنها از آزمون شونده خواسته می‌شود تا حداکثر سعی خود را به کار بندد تا بهترین نمره ممکن را به دست آورد.

آزمون توانایی‌های شناختی

آزمون توانایی‌های شناختی بیشتر در زمینه‌های تحصیلی و فعالیت‌های فکری و ذهنی مورد استفاده قرار می‌گیرد.

آزمون‌های توانایی‌های روانی - حرکتی

آزمون‌های توانایی‌های روانی - حرکتی بیشتر در زمینه‌های صنعتی و نظامی و تربیت بدنی مورد استفاده قرار می‌گیرند. آزمون‌های توانایی را به دو دسته تقسیم می‌کنند.

۱- استعداد^۴

۲- پیشرفت تحصیلی^۵

الف) آزمون‌های استعداد

آزمون‌های استعداد مقدار توانایی فرد را برای انجام کارهایی که در پیش دارد می‌سنجد و ناظر به آینده است. و ظرفیت فرد را برای یادگیری‌های متفاوت می‌سنجد. آزمون‌های هوش و استعداد‌های مختلف معرف آزمون‌های استعداد هستند که معروفترین آزمون‌های استعداد می‌باشند. هدف از کاربرد آزمون‌های هوش و استعداد‌های مختلف پیش‌بینی

^۱Ability

^۲Cognitive

^۳psychomotor

^۴Aptitude

^۵Academic(Scholastic) Acievement

موفقیت‌های همه‌جانبه افراد است. اما آزمون‌های استعداد‌های ویژه برای پیش‌بینی موفقیت یا شایستگی افراد در زمینه‌های فکری و علمی مشخص به کار می‌روند.

ب) آزمون‌های پیشرفت تحصیلی

آزمون‌های پیشرفت تحصیلی دانش و مهارت‌هایی را که فرد تا لحظه اجرای آزمون کسب کرده است، اندازه می‌گیرد و ناظر به گذشته است. آزمون‌های پیشرفت تحصیلی با موضوع یا محتوایی که اندازه‌گیری می‌شوند، مشخص می‌گردند و در نتیجه با عنوان آن موضوع‌ها نام‌گذاری می‌شوند. مانند آزمون پیشرفت ریاضی، آزمون پیشرفت فیزیک و ...

آزمون‌های استعداد و آزمون‌های پیشرفت تحصیلی را به روشنی و وضوح نمی‌توان از هم تفکیک کرد چرا که یادگیری‌های قبلی فرد پایه‌ای برای یادگیری‌های آینده وی است. بنابراین تمایز بین آزمون‌های استعداد و پیشرفت تحصیلی به قصد و منظوری مربوط می‌شود که نتایج آزمون برای آن قصد مورد استفاده قرار می‌گیرد نه به ماهیت و محتوای آن آزمون.

انواع مختلف آزمون‌های توانایی

آزمون‌های توانایی							
آزمون‌های استعداد			آزمون‌های پیشرفت تحصیلی				
آزمون‌های استعداد ویژه		آزمون‌های هوش کلی		و غیره	آزمون مهارت‌های فنی	آزمون فیزیک	آزمون ریاضی
و غیره	آزمون استعداد فنی	آزمون استعداد ریاضی	آزمون استعداد موسیقی				

۲- آزمون‌های عاطفی^۱

ویژگی‌های عاطفی و شخصیتی افراد را می‌سنجد و برخلاف آزمون‌های توانایی که حداکثر عملکرد آزمون‌شونده را اندازه می‌گیرد آزمون‌های عاطفی وضع موجود آزمون‌شونده را در موقعیت‌های طبیعی زندگی می‌سنجد و بیشتر برای امور مشاوره در حل مسائل روانی تربیتی حرفه‌ای به کار می‌رود. عیب آزمون‌های عاطفی این است که به اندازه آزمون‌های شناختی و عینی دقیق نیستند. آزمون‌های عاطفی دارای انواع زیادی هستند که مهم‌ترین آنها عبارتند از:

۱- آزمون‌های شخصیت^۲

۲- آزمون‌های نگرش^۳

۳- آزمون‌های علاقه^۱

^۱Affective
^۲Personality
^۳Attitude

الف) آزمون‌های شخصیت

شخصیت مجموعه رفتارهایی است که فرد در موقعیت‌های اجتماعی از خود نشان می‌دهد. روش‌های سنجش شخصیت باید شامل متغیرهای مختلف شناختی، روانی - حرکتی و عاطفی باشند. آزمون‌های شخصیت به طور عمده به اندازه‌گیری متغیرهایی چون انگیزش، مزاج، سازگاری و منش می‌پردازند. ابزارهای اندازه‌گیری شخصیت را می‌توان به سه دسته عمده زیر تقسیم کرد:

۱- پرسشنامه‌های گزارش شخصی

۲- پرسشنامه‌های جامعه مدنی

۳- فنون مشاهده‌ای

پرسشنامه‌های گزارش شخصی

در این نوع از ابزارهای اندازه‌گیری فرد نظر شخصی خودش را نسبت به خود بیان می‌دارد.

پرسشنامه‌های جامعه مدنی

در این نوع از ابزارهای اندازه‌گیری افراد دیگر نظر شخصی خودشان را نسبت به فرد مورد نظر بیان می‌دارند.

فنون مشاهده‌ای

در این نوع از ابزارهای اندازه‌گیری شخصیت رفتار فرد در یک موقعیت بخصوص مورد مشاهده قرار می‌گیرد. راه دیگر دسته‌بندی آزمون‌های شخصیت، توجه به روش مطالعه رفتار فرد یا نحوه ارائه محرک‌هاست. آزمون‌های شخصیت را از این بابت می‌توان به دو دسته تقسیم کرد. در یکی از این دو دسته آزمون‌ها، محرک‌ها به صورت سازمان‌یافته ارائه می‌شوند. و در دسته دیگر به صورت سازمان‌نیافته، یعنی سؤال‌های آزمون به وسیله آزمون‌شوندگان مختلف به گونه‌های متفاوت تعبیر و تفسیر می‌شوند. آزمون‌های فرافکنی از این دسته هستند.

ب) آزمون‌های علاقه و نگرش

هم علاقه و هم نگرش به دوست داشتن‌ها و دوست نداشتن‌های افراد اشاره می‌کند و هر دو شامل واحدهای شخصی نسبت به امور هستند. اما آنچه علاقه و نگرش را از هم مجزا می‌کند، موضوع این علاقه‌ها یا بی‌علاقگی‌ها است. نکته قابل ذکر تفاوت نگرش و علاقه می‌باشد. علاقه به احساس فرد نسبت به یک فعالیت اشاره می‌کند، مثلاً من به فوتبال علاقه‌مندم اما نگرش حاکی از احساس فرد نسبت به یک شیء نهاد اجتماعی یا گروه است مثلاً نگرش فرد نسبت به آثار تاریخی معروف‌ترین وسیله برای اندازه‌گیری در زمینه عاطفی پرسشنامه‌های علاقه و نگرش است.

برای سنجش علاقه‌ها و نگرش‌های افراد روش‌های مختلفی وجود دارند. معروف‌ترین وسیله اندازه‌گیری در این دو زمینه پرسشنامه‌های علاقه و نگرش است.

انواع مختلف آزمون‌های عاطفی

آزمون‌های عاطفی			
و غیره	آزمون نگرش	آزمون علاقه	آزمون شخصیت

دسته‌بندی آزمون‌ها با توجه به چگونگی تهیه آنها

آزمون‌ها آزمون‌ها با توجه به چگونگی تهیه آنها به دو دسته کلی دسته‌بندی می‌شوند:

۱. آزمون‌های میزان‌شده یا استانداردشده

۲. آزمون‌های معلم‌ساخته

الف) آزمون‌های میزان شده یا استاندارد شده

آزمون‌های استاندارد شده یا میزان شده آزمون‌هایی هستند که با استفاده از نمونه‌های بزرگ آزمون‌شوندگان و غالباً توسط مؤسسات آزمون‌سازی برای اندازه‌گیری پیشرفت تحصیلی یا هوش یا مقاصد ارزشیابی از دوره‌های تحصیلی به کار می‌روند و در آنها از ملاک نسبی استفاده می‌شود. مثال رایج در ایران آزمون‌های کنکور سراسری می‌باشد.

ب) آزمون‌های معلم‌ساخته

آزمون‌های معلم‌ساخته، آزمون‌هایی هستند که معلمان برای سنجش پیشرفت تحصیلی دانش‌آموزان خود تهیه می‌کنند و در طول دوره آموزشی و یا در پایان دوره به اجرا در می‌آورند و غالباً برای تعیین میزان توفیق یادگیرندگان در نیل به اهداف دقیق آموزشی طرح‌ریزی می‌شوند و ملاک مطلق هستند. یعنی نمره هر فرد با توجه به میزان آشنایی او با اهداف درس اندازه‌گیری می‌شود.

تفاوت آزمون‌های میزان‌شده با آزمون‌های معلم‌ساخته

- تهیه آزمون‌های میزان‌شده در مقایسه با آزمون‌های معلم‌ساخته به زمان و تخصص بیشتری نیاز دارد.
- از آزمون‌های میزان‌شده در مناطق مختلف و دانش‌آموزان مختلف استفاده می‌شود. از این رو در این آزمون‌ها ویژگی‌های اختصاصی کلاس‌ها و مدارس به خصوص نادیده گرفته می‌شود.
- آزمون‌های میزان‌شده (استاندارد) به وجود داده‌های هنجاری اشاره می‌کنند. اما آزمون‌های معلم‌ساخته نیاز به هنجار ندارند.

۴. هدف اصلی آزمون‌های میزان‌شده مقایسه یک فرد یا یک گروه یا فرد یا گروه دیگر است.

استفاده از آزمون‌های میزان‌شده و معلم‌ساخته در تصمیم‌گیری‌های آموزشی

۱. تصمیمات آموزشی مربوط به کم و کیف جریان آموزشی را باید بر مبنای آزمون‌های معلم‌ساخته گرفت.
۲. تصمیمات مربوط به ارتقای دانش‌آموزان و دانشجویان را نیز باید بر مبنای آزمون‌های معلم‌ساخته گرفت.
۳. تصمیمات تشخیصی مربوط به مشکلات یادگیرندگان و تصمیمات مربوط به رفع نواقص یادگیری آموزش را می‌توان بر مبنای هر دو نوع آزمون معلم‌ساخته و میزان‌شده آغاز کرد.
۴. تصمیمات مربوط به جایابی یا پایه‌گزینی افراد را براساس آزمون‌های میزان‌شده اتخاذ می‌کنند.
۵. تصمیمات مربوط به راهنمایی و مشاوره دانش‌آموزان براساس آزمون‌های میزان‌شده گرفته می‌شود.
۶. برای مقایسه افراد و گروه‌ها از آزمون‌های میزان‌شده استفاده می‌گردد.
۷. تصمیمات مربوط به برنامه‌ریزی درسی براساس هر دو دسته آزمون‌ها صورت می‌گیرد.
۸. تصمیمات مربوط به ارزشیابی از میزان توفیق مدارس مختلف نیز با استفاده از آزمون‌های میزان‌شده صورت می‌گیرد.

نکات کلیدی و مهم

- آزمون‌های توانایی حداکثر عملکرد آزمون شونده را اندازه می‌گیرند.
- آزمون‌های عاطفی وضع موجود آزمون شونده را در موقعیت‌های طبیعی زندگی می‌سنجند.
- آزمون‌های توانایی به طور عمده برای مقاصد پیش‌بینی موفقیت‌های تحصیلی و شغلی مورد استفاده قرار می‌گیرند.
- آزمون‌های عاطفی بیشتر به امور مشاوره در حل مسائل روانی، تربیتی و حرفه‌ای کمک می‌کنند.
- آزمون‌های عاطفی به اندازه آزمون‌های شناختی عینی و دقیق نیستند به همین دلیل بسیاری از روانشناسان کاربرد اصطلاح آزمون را برای این ابزار اندازه‌گیری درست نمی‌دانند.
- آزمون‌های فرافکنی شامل تصاویر مبهم یا لکه‌های جوهر هستند که آزمون‌شونده باید آنها را تعبیر و تفسیر کند.
- دسته‌بندی انواع آزمون‌های مورد استفاده در آموزش و پرورش در روانشناسی

گروه اول: انواع آزمون‌های با توجه به ویژگی مورد اندازه‌گیری

آزمون‌های توانایی

آزمون‌های پیشرفت تحصیلی

- شناختی
- روانی - حرکتی

آزمون‌های استعداد

- آزمون‌های هوش یا استعداد کلی
- آزمون‌های استعداد‌های ویژه

آزمون‌های عاطفی

۱. آزمون شخصیت
۲. آزمون علاقه
۳. آزمون نگرش
۴. غیره

گروه دوم: انواع آزمون‌ها با توجه به چگونگی تهیه آنها: ۱- آزمون‌های میزان‌شده و ۲- آزمون‌های معلم‌ساخته معروف‌ترین وسیله اندازه‌گیری علاقه‌ها و نگرش‌های افراد پرسشنامه‌های علاقه و نگرش است.

فصل چهارم:

دسته‌بندی ارزشیابی‌های آموزشی

دسته‌بندی ارزشیابی‌های آموزشی در قالب رویکردهای مختلف

صاحب‌نظران ارزشیابی آموزشی نسبت به اینکه ارزشیابی چیست و چه مراحل دارد اختلاف نظر دارند که این اختلاف نظرات رویکردهای ارزشیابی آموزشی را به وجود آورده است.

دسته‌بندی ارزشیابی‌های آموزشی در قالب رویکردهای ورتن و سندرز

ورتن و سندرز رویکرد ارزشیابی را به شش دسته تقسیم کرده‌اند که شامل:

- ۱- رویکرد مبتنی بر هدف،
- ۲- رویکرد مبتنی بر مدیریت،
- ۳- رویکرد مبتنی بر مصرف‌کننده،
- ۴- رویکرد مبتنی بر نظر متخصصان،
- ۵- رویکرد مبتنی بر اختلاف نظر متخصصان

۶- رویکرد مبتنی بر مشارکت‌کنندگان و طبیعت‌گرایانه است.

۱- رویکرد مبتنی بر هدف^۱

در این رویکرد هدف‌های کلی و دقیق مورد تأکید قرار می‌گیرد. این رویکرد تلاش دارد تعیین نماید چه میزان از اهداف تحقق یافته‌اند و یا در حال تحقق یافتن هستند. در این رویکرد الگوی ارزشیابی تایلر و هیلدا تابا را می‌توان ذکر کرد.

الف) الگوی ارزشیابی تایلر^۲

الگوی ارزشیابی تایلر نخستین الگوی ارزشیابی آموزشی است که تاریخ آن به ۱۹۳۰ باز می‌گردد. این الگو به الگوی تایلری معروف است. مقصود تایلر از ارزشیابی آموزشی این است که تعیین کند هدف‌های آموزشی برنامه آموزشگاه یا برنامه درسی تا چه میزانی تحقق یافته‌اند. به همین منظور تایلر در الگوی ارزشیابی خود مراحل زیر را پیشنهاد داده است:

- ۱- تعیین هدف‌های کلی (غایتها) و هدف‌های دقیق
- ۲- طبقه‌بندی غایت‌ها و هدف‌ها
- ۳- بیان هدف‌ها به صورت رفتاری، یعنی قابل اندازه‌گیری
- ۴- یافتن موقعیت‌هایی که در آنها بتوان دستیابی به هدف‌ها را نشان داد.
- ۵- تولید یا تهیه روش‌ها و فنون اندازه‌گیری
- ۶- گردآوری داده‌های مربوط به عملکرد یادگیرندگان
- ۷- مقایسه داده‌های مربوط به عملکرد با هدف‌های رفتاری

۲- رویکرد مبتنی بر مدیریت^۳

در این رویکرد، ارزشیابی به قصد کمک به مدیران و تصمیم‌گیران انجام می‌گیرد و مخاطب آن تصمیم‌گیرندگان آموزش و پرورش هستند.

الف) الگوی ارزشیابی سیپ

الگوی ارزشیابی سیپ (cip) از استافیل بیم برخاسته و رویکرد مبتنی بر مدیریت است که در آن به ارزشیابی بافت^۱

^۱Objective – Oriented approach

^۲Tylerian Evaluation Approach

^۳Management – Oriented Approach

ارزشیابی از درون‌داد، ارزشیابی از فرایند^۳ و ارزشیابی از فرآورده می‌پردازند. نام الگو از حروف اول این چهار نوع ارزشیابی گرفته شده است.

ارزشیابی زمینه‌ای

به وسیله این نوع ارزشیابی به سنجش نیازها و مسائل موجود، فرصت‌ها و امکانات بلااستفاده پرداخته می‌شود. از طریق این ارزشیابی، می‌توان در مورد مطلوبیت هدف‌های از قبل تعیین شده قضاوت کرد. همچنین به تدوین هدف‌های آموزش پرداخت. هدف ارزشیابی زمینه (بستر) عبارت است از تعریف زمینه سازمانی، تعیین جامعه مخاطب فعالیت‌های آموزشی و سنجش نیازهای آموزشی آن؛ مشخص کردن فرصت‌های موجود برای برآوردن نیازها به تشخیص مسائلی که مبنای نیازها است؛ تعیین اینکه آیا هدف‌های پیشنهادی به اندازه کافی نیازهای مورد نظر را برآورده می‌کند. روش اجرای ارزشیابی زمینه شامل تحلیل سیستمی؛ بررسی پیمایشی، بازنگری اسناد و مدارک؛ گفت و گو و مذاکره؛ مصاحبه؛ آزمون‌های تشخیصی؛ و روش دلفی است. حاصل ارزشیابی زمینه آن است که موقعیت جامعه، بستر جامعه مخاطب را آشکار می‌سازد.

ارزشیابی درون‌داد

برای ارزیابی درون دادها ما به چند مورد اشاره می‌کنیم که بطور کلی در قالب چند مولفه مطرح اند. (منابع انسانی، مادی، مالی، اطلاعاتی و زمان لازم). روش اجرای این نوع ارزشیابی شامل بررسی پیشینه فعالیت‌هایی از قبیل بازدید از برنامه‌های مشابه، اجرای طرح‌های تجربی و غیره می‌باشد.

یادگیرنده: برای وقوف به رفتار ورودی، ویژگی‌ها، میزان دانش و اطلاعات معلم: ارزیابی معلم اهداف چندی می‌تواند داشته باشد از جمله

قضاوت در مورد عملکردشان برای بهبود کار خودشان

قضاوت در مورد عملکردشان برای تصمیم‌گیری‌های استخدامی و غیره...

ارزشیابی فرایند

این نوع ارزشیابی به منظور تشخیص یا پیش‌بینی مشکلات اجرایی در جریان انجام فعالیت‌های آموزش و میزان مطلوبیت فرایند اجرای این فعالیت‌ها به عمل می‌آید. به عبارت دیگر ارزشیابی فرایند وسیله‌ای است جهت ثبت وقایع در حین اجرای فعالیت‌ها. روش اجرای ارزشیابی فرایند شامل نظارت بر اجرای فعالیت‌ها و گردآوری

داده‌های مربوط به تصمیم‌گیری‌های ضمن اجرای این فعالیت‌هاست.

ارزشیابی محصول: ارزشیابی محصول در نظام آموزشی، یعنی ارزشیابی بازده‌های کوتاه‌مدت و یا به عبارتی مقطعی هر یادگیری در نظام آموزشی می‌باشد. ارزشیابی محصول شامل برون‌داد واسطه‌ای، نمره‌های پایان ثلث یا نیمسال تحصیلی، گزارش‌های پیشرفت تحصیلی از پایه‌ای به پایه بالاتر و... می‌شود.

ارزشیابی برون‌داد

این نوع ارزشیابی به منظور قضاوت درباره مطلوبیت بازده فعالیت‌های آموزش انجام می‌شود. ارزشیابی برون‌داد به عبارت دیگر، جهت مرتبط کردن برون‌داد و پیامدهای نظام (برنامه آموزشی) با عوامل مربوط به زمینه، درون‌داد و فرایند این نظام است تا از این طریق بتوان به ارزش و مطلوبیت آنها پی برد. روش انجام ارزشیابی برون‌داد شامل مشخص کردن ملاک‌های مربوط به برون‌داد، تعریف عملیاتی برون‌داد، نشانگرهای آن و چگونگی اندازه‌گیری آنهاست. علاوه بر آن باید به گردآوری داده‌ها درباره قضاوت یاران آموزش (افراد ذی‌ربط، ذی‌نفع، و ذی‌علاقه) نسبت به مطلوبیت برون‌داد پرداخت.

ارزشیابی بازخورد

بازخورد در تمام مراحل ارزشیابی نقش کلیدی دارد به عبارت دیگر یکی از دلایل اینکه ارزشیابی صورت می‌گیرد این است که بازخورد ارائه شود. تا اینکه در فعالیتها و اقدامات بهبود حاصل شود چرا که ارزیابی صرفاً برای ارزیابی جایز نیست و بایستی منجر به بهبود گردد.

ب) الگوی ارزشیابی یوسی ال.ای: UCLA

همچنین در الگوی ارزشیابی یوسی ال.ای ارزشیابی آموزشی این چنین تعریف شده است:

«ارزشیابی عبارت است از فرایند تعیین انواع تصمیماتی که باید اتخاذ شوند و نیز انتخاب، گردآوری و تحلیل اطلاعات مورد نیاز برای اخذ این تصمیمات و گزارش کردن این اطلاعات به تصمیم‌گیرندگان».

الگوی یوسی. ال.ای از پنج نوع ارزشیابی تشکیل شده است که عبارتند از: سنجش نظام، طراحی برنامه، اجرای برنامه، بهسازی برنامه و تصدیق برنامه.

۱- سنجش نظام^۱

این مرحله از الگوی ارزشیابی یوسی. ال.ای اطلاعات لازم را در مورد وضعیت نظام آموزشی گردآوری می‌کند. این

مرحله بسیار شبیه مرحله بافت در الگوی سیپ است.

۲- طراحی برنامه^۱

در این مرحله برنامه‌های ویژه‌ای انتخاب می‌شوند که تصور می‌شود در برآوردن نیازهای آموزشی مفید هستند.

۳- اجرای برنامه^۲

در این مرحله اطلاعات مربوط به این که برنامه انتخابی به گروه مورد نظر به همانگونه که مورد نظر بوده است ارائه شده است یا نه مورد بررسی قرار می‌گیرد.

۴- بهسازی برنامه^۳

در این مرحله در مورد چگونگی کارکرد برنامه، چگونگی تحقق هدفها و اینکه بازده‌های پیش‌بینی نشده‌ای ظاهر شده‌اند یا نه. (این مرحله شبیه مرحله ارزشیابی فرایند در الگوی سیپ است).

۵- تصدیق برنامه^۴

در این مرحله اطلاعات لازم جهت قضاوت در مورد ارزش برنامه و امکان استفاده از آن در موقعیت‌های دیگر فراهم می‌آید.

۳- رویکرد مبتنی بر مصرف‌کننده^۵

در این رویکرد، ارزشیابی به تعیین میزان مطلوب بودن یک فرآورده آموزشی می‌پردازد. مثلاً مطلوبیت و فواید یک برنامه ضمن خدمت را برای معلمان درس ریاضی مورد ارزیابی قرار می‌دهد. در رویکرد ارزشیابی مبتنی بر مصرف‌کننده ارزشیابی برابر است با تعیین میزان مطلوبیت یک فرآورده آموزشی. در این رویکرد الگوهای اسکرین و الگوی ارزشیابی هدف آزاد قابل ذکر است.

الف) الگوی ارزشیابی اسکرین

اسکرین معروف‌ترین شخصیتی است که زمینه ارزشیابی آموزشی به قصد کمک به مصرف‌کنندگان کار کرده است. الگوی ارزشیابی اسکرین به تمایز بین ارزشیابی تکوینی^۶ و ارزشیابی تراکمی^۷ می‌پردازد.

^۱Program planning

^۲Implementation Program

^۳Program improvement

^۴Program certification

^۵Consumer - Oriented Approach

^۶formative evaluation

^۷summative evaluation

ارزشیابی تکوینی در حین اجرای برنامه و زمانی که برنامه در حال تکوین است با هدف رفع و اصلاح نواقص صورت می‌گیرد.

ارزشیابی تراکمی برای اینکه معلوم گردد برنامه تا چه حد با اهداف مورد نظر تطابق دارد و نیاز مصرف‌کننده را برآورده می‌سازد، صورت می‌گیرد.

الگوی ارزشیابی اسکریون برخاسته از رویکرد مبتنی بر مصرف‌کننده عمدتاً یک ارزشیابی تراکمی است. ملاک‌هایی که اسکریون برای ارزشیابی از فرآورده‌ها به کار می‌گیرد عبارتند از:

۱. شواهد حاکی از محقق شدن اهداف مهم آموزشی؛

۲. شواهد حاکی از محقق شدن اهداف مهم غیرآموزشی؛

۳. نتایج حاصل از پیگیری؛

۴. آثار جانبی پیش‌بینی نشده؛

۵. دامنه استفاده (یعنی برای چه کسانی مفید است)

۶. ملاحظات اخلاقی

۷. هزینه‌ها.

ب) الگوی ارزشیابی هدف آزاد^۱

الگوی دیگر الگوی هدف آزاد اسکریون است که معتقد است برنامه علاوه بر نتایج مورد انتظار نتایج دیگری هم دارد لذا ارزشیابی هدف آزاد تمامی نتایج یک نوآوری آموزشی را بدون هیچ‌گونه پیش‌داوری قبلی ارزیابی می‌کند. نقش عمده ارزشیابی هدف آزاد، کاستن از سوگیری و افزایش عینیت است.

۴- رویکرد مبتنی بر نظر متخصصان^۲

در این رویکرد متخصصان یک رشته یا فن به داوری اثربخشی مؤسسات، برنامه‌ها، فرآورده‌ها و فعالیت‌های آموزشی متخصصان می‌پردازند. مانند امتحانات شفاهی دوره دکتری. در این رویکرد تاکید اصلی بر کاربرد نظر متخصصان در قضاوت در مورد کیفیت فعالیت‌های آموزش و پرورش است.

ورتن و سندرز الگوهای ارزشیابی برخاسته از نظر متخصصان را در چهار طبقه تقسیم‌بندی کرده‌اند:

(۱) نظامهای بررسی تخصصی رسمی

(۲) نظامهای بررسی تخصصی غیررسمی

^۱goal -free evaluation

^۲Expertise - Oriented Approach

۳) بررسی توسط جلسات میزگرد ویژه

۴) بررسی توسط افراد ویژه.

۵- رویکرد مبتنی بر اختلاف نظر متخصصان

این رویکرد بر این فرض استوار است که باید کسانی که مورد ارزشیابی قرار می‌گیرند خود در ارزیابی دخالت کنند. این رویکرد بیشتر جنبه انسان‌گرایانه دارد. پائولوفریره از صاحب‌نظران در این رویکرد است. در ارتباط با رویکرد مبتنی بر اختلاف نظر متخصصان ورتن و سندرز بیان می‌کنند درحالی که غالب رویکردهای ارزشیابی می‌کوشند تا سوگیریها را کاهش دهند، رویکرد مبتنی بر اختلاف‌نظرهای متخصصان با دخالت دادن نظرهای مثبت و منفی در ارزشیابی سوگیریها را متعادل سازد.

در رویکرد مبتنی بر اختلاف‌نظرهای متخصصان یک یا گروهی از ارزشیاب‌ها سعی دارند مثبت‌ترین جنبه‌های یک برنامه یا موضوع را به عنوان مدافعان مورد بررسی و معرفی نمایند و در مقابل یک یا گروهی از ارزشیاب‌ها سعی دارند منفی‌ترین جنبه‌های یک برنامه یا موضوع را جلوه دهند و از ترکیب نتایج ارزشیابی دو گروه یا دو ارزشیاب نتیجه درست‌تر و عادلانه‌تری حاصل می‌شود.

۶- رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان^۱

در این رویکرد بررسی طبیعت‌گرایانه و کوشش مشارکت‌کنندگان در تعیین ارزش‌ها، نیازها و داده‌های ارزشیابی مورد تأکید قرار می‌گیرند. این رویکرد بر تجارب دست اول در موقعیت‌ها و فعالیت‌های تربیتی و آموزشی تأکید دارد.

الف) الگوی ارزشیابی سیمایی

یکی از معروف‌ترین الگوهای ارزشیابی برخاسته از رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان رویکرد رابرت ای. استیک به نام الگوی سیمایی است. در این الگوی ارزشیابی بر دو عمل اساسی ارزشیابی یعنی توصیف و داوری تأکید می‌شود. این دو عمل توصیف و داوری در سه مرحله از یک برنامه آموزشی، یعنی پیش‌آیندها، مبادله‌ها و بازده‌ها یا پی‌آیندها انجام می‌گیرد.

علاوه بر این الگو، چند الگوی معروف دیگر وابسته به رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کننده وجود دارند که عبارتند از: ارزشیابی روشنگر، ارزشیابی دموکراتیک، ارزشیابی پاسخگو و ارزشیابی طبیعت‌گرایانه.

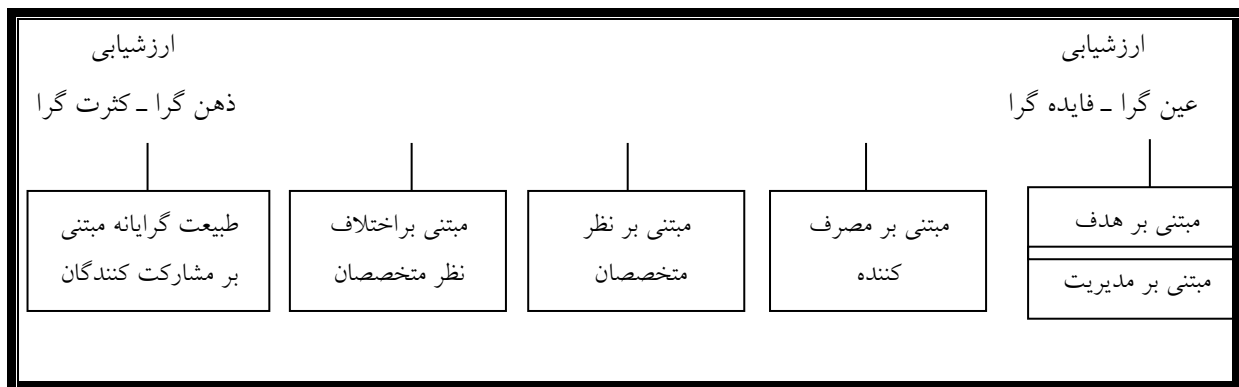
دسته‌بندی ارزشیابی‌های آموزشی در قالب تقسیم‌بندی دوگانه هاوس

^۱Naturalistic and participant-Oriented Approach

دسته‌بندی رویکردهای شش‌گانه ورتن و سندرز که در بالا به آن اشاره شد بر تقسیم‌بندی دوگانه هاوس از ارزشیابی‌های آموزشی استوار است. هاوس انواع ارزشیابی‌های آموزشی را در دو دسته عین‌گرا در مقابل ذهن‌گرا یا فایده‌گرا در مقابل شهودگرا - کثرت‌گرا قرار داده است. اگر دسته‌بندی شش‌گانه رویکردهای ارزشیابی ورتن و سندرز را با دسته‌بندی دوگانه هاوس مقایسه کنیم خواهیم دید که رویکرد مبتنی بر هدف با کرانه عینی‌گرایی یا فایده‌گرایی مطابق است و رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان بر کرانه ذهنی‌گرایی یا شهودگرایی - کثرت‌گرایی انطباق دارد. و بقیه رویکردها در میانه این دو طیف جای دارند.

در جدول زیر دو دیدگاه مربوط به دسته‌بندی ارزشیابی آموزشی مقایسه شده‌اند و جایگاه رویکردهای شش‌گانه ورتن و سندرز در مقایسه با دو بعد ارزشیابی هاوس مشخص شده‌اند.

مقایسه رویکردهای ارزشیابی ورتن و سندرز با دو بعد ارزشیابی هاوس



دسته‌بندی ارزشیابی‌های آموزشی با توجه به موضوع ارزشیابی

دسته‌بندی ارزشیابی‌های آموزشی با توجه به موضوع ارزشیابی بیانگر انواعی متفاوتی از ارزشیابی است که در ادامه بدانها می‌پردازیم.

۱. ارزشیابی از دانش‌آموزان و دانشجویان
۲. ارزشیابی از برنامه‌ها و مواد آموزشی
۳. ارزشیابی از پروژه‌ها و برنامه‌های آموزشی
۴. ارزشیابی از آموزشگاه
۵. ارزشیابی از کارکنان آموزشگاه

الف) ارزشیابی از دانش‌آموزان و دانشجویان^۱

مقصود از ارزشیابی از دانش‌آموزان و دانشجویان عمدتاً ارزشیابی پیشرفت تحصیلی یا ارزشیابی از میزان یادگیری یادگیرندگان است. عمده‌ترین ملاک قضاوت در ارزشیابی آموزشی هدفهای آموزشی از پیش تعیین شده هستند. مهم‌ترین نوع ارزشیابی آموزشی ارزشیابی پیشرفت تحصیلی است. نتایج ارزشیابی پیشرفت تحصیلی نه تنها برای قضاوت درباره عملکرد دانش‌آموزان بلکه برای داوری درباره اثربخشی کیفیت آموزش و روش‌ها و مواد آموزشی هم مفید است.

ب) ارزشیابی از برنامه‌ها و مواد آموزش^۱

گاهی اوقات ارزشیابی به منظور تعیین اثربخشی برنامه درسی یا مواد آموزشی انجام می‌شود و شامل ارزشیابی از عواملی چون روش‌های آموزشی کتاب‌های درسی مواد دیداری - شنیداری و... است. در ارزشیابی از برنامه‌های درسی و مواد آموزشی نیز عمده‌ترین ملاک، پیشرفت تحصیلی یا میزان یادگیری دانش‌آموزان و دانشجویان است اما می‌توان از ملاک‌های دیگری چون نگرش معلمان و یادگیرندگان نیز سود جست.

ج) ارزشیابی از پروژه‌ها و برنامه‌های آموزشی

منظور از پروژه‌ها و برنامه‌های آموزشی کوشش‌هایی هستند که به عنوان بخشی از برنامه‌های معمول آموزشگاهها به حساب نمی‌آیند. مفاهیم برنامه، پروژه، و برنامه درسی از هم متمایزند. پروژه یک کوشش مجزاً و احتمالاً یکباره است که با استفاده از یک نوآوری می‌کوشد تا تغییری بوجود آورد. برنامه درسی از پروژه و برنامه جامع‌تر است. برنامه درسی را به عنوان مجموعه‌ای از برنامه‌ها تعریف کرده‌اند.

د) ارزشیابی از آموزشگاه

منظور از ارزشیابی از آموزشگاه این است که تعیین شود هدف‌های پیش‌بینی شده برای آموزشگاه به چه میزان تحقق یافته‌اند و نقاط قوت و ضعف کدامند. یکی از جنبه‌های مهم ارزشیابی از آموزشگاه برنامه آزمون آموزشگاه است هرچند که این برنامه جامع‌تر و کامل‌تر باشد. اطلاعات بدست آمده نیز به همان نسبت قابل اعتمادتر و باارزش‌تر خواهند بود.

ه) ارزشیابی از کارکنان آموزشگاه

مقصود از ارزشیابی از کارکنان آموزشگاه ارزشیابی از همه کسانی است که بطور مستقیم یا غیر مستقیم مسئول بازدهی آموزشی مؤسسه آموزشی هستند شاید پیچیده‌ترین نوع ارزشیابی همین ارزشیابی از کارکنان به ویژه ارزشیابی از کار معلمان است. روش‌های ارزشیابی از معلم متنوع‌اند.

^۱evaluation of Curriculum and instructional materials

ارزشیابی معلم به عنوان دومین درون داد اصلی نظام آموزشی یک فرایند چند وجهی می‌باشد که دارای دو هدف است: الف) ارزشیابی تکوینی عملکرد معلمان از طریق ارائه داده‌های لازم و قضاوت درباره توانایی‌ها و مهارت‌های آنان و پیشنهاد برای بهبودی. ب) ارزشیابی پایانی عملکرد معلمان برای تصمیم‌گیری‌های استخدامی از قبیل ورود به خدمت، نگهداری و ارتقا. برای ارزشیابی معلمان، شیوه‌های گوناگونی به کار برده می‌شود. از جمله می‌توان:

- ۱- مصاحبه‌های استخدامی به منظور قضاوت درباره آمادگی معلمان
- ۲- آزمون دانش، نگرش و مهارت‌های آنان
- ۳- نظر همگنان (همکاران)
- ۴- مشاهده کلاس درس
- ۵- نظرخواهی از دانشجویان درباره فرایند تدریس معلم
- ۶- خودسنجی معلم
- ۷- ملاحظه رشد صلاحیت‌های حرفه‌ای و تخصصی معلم است.

دسته‌بندی ارزشیابی‌های آموزشی با توجه به ملاک مورد استفاده

ارزشیابی‌های آموزشی را می‌توان از نظر ملاک مورد استفاده از آنها نیز تقسیم نمود که از این لحاظ به دو دسته ارزشیابی وابسته به ملاک و ارزشیابی وابسته به هنجار تقسیم می‌شوند. که معروف‌ترین تقسیم‌بندی برای ارزشیابی آموزشی می‌باشد.

۱- ارزشیابی وابسته به ملاک^۱

در ارزشیابی وابسته به ملاک یا ارزشیابی ملاکی ملاک یا معیار ارزشیابی از پیش تعیین می‌شود و عملکرد یادگیرنده در آزمون با توجه به آن ملاک مورد قضاوت قرار می‌گیرند. این نوع ارزشیابی بیشتر ویژه پیشرفت تحصیلی دانش‌آموزان و دانشجویان است. گرفتن نمره ۱۰ برای قبولی در هر درس مثالی از ارزشیابی وابسته به ملاک یا ارزشیابی ملاکی است.

۲- ارزشیابی وابسته به هنجار^۲

ارزشیابی وابسته به هنجار یا ارزشیابی هنجاری به نوعی ملاک نسبی وابسته است. در این ارزشیابی پیشرفت تحصیلی گروهی دانش‌آموزان را با یکدیگر مقایسه می‌کنیم. نه با یک ملاک از پیش تعیین شده. علت نامگذاری این نوع ارزشیابی به ارزشیابی هنجاری این است که در آن ملاک نوعی هنجار گروهی است. لذا پیشرفت دانش‌آموز

^۱Criterion - referenced evaluation

^۲norm- referenced evaluation

نسبت به دانش‌آموزان دیگر تعیین می‌شود. مانند امتحان کنکور سراسری که در آن براساس رتبه‌هایی که دانشجویان کسب می‌کنند تا یک رتبه مشخص قبول و بقیه رد شوند.

در ارزشیابی هنجاری پیشرفت دانش‌آموزان برحسب رتبه درصدی یا نوعی هنجار سنی یا کلاسی و در مقایسه با گروه هنجار تعیین می‌شود.

در ارزشیابی هنجاری هدف عمده این است که تعیین شود که یک دانش‌آموز در مقایسه با دانش‌آموزان دیگر چه وضعی دارد اما در ارزشیابی ملاکی منظور این است که مشخص شود یک دانش‌آموز چه مقدار از آنچه را قرار بوده بیاموزد، آموخته است.

یکی دیگر از تفاوت‌های مهم آزمون‌های ملاکی و هنجاری در توزیع فراوانی نمرات حاصل از اجزای آنهاست. نمرات آزمون‌های ملاکی دارای کجی منفی می‌باشند.

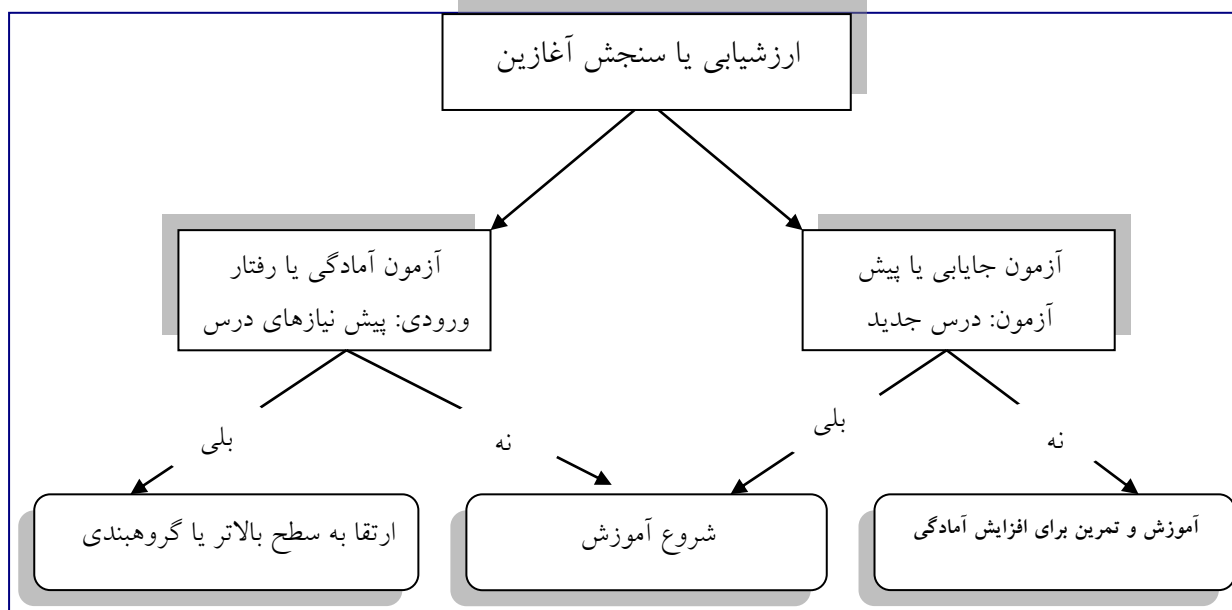
دسته‌بندی ارزشیابی‌ها با توجه به زمان و هدف استفاده از آنها

در دسته‌بندی ارزشیابی‌ها با توجه به زمان و هدف استفاده از آنها ارزشیابی‌ها را چهار دسته تقسیم‌بندی می‌کنند:

۱. ارزشیابی آغازین
۲. ارزشیابی تکوینی
۳. ارزشیابی تشخیصی
۴. ارزشیابی تراکمی

دسته اول: ارزشیابی آغازین

این ارزشیابی پیش از انجام فعالیت‌های آموزش توسط معلم به اجرا درمی‌آید با این هدف که به اندازه‌گیری رفتارهای ورودی یا مهارت‌ها و دانش‌های پیش‌آزمون می‌گویند و عمدتاً همان آزمون‌هایی که شامل هدف‌های آموزشی یادگیری درس می‌باشد را دربرمی‌گیرد هدف از اجرای پیش‌تست تعیین میزان اطلاعات یادگیرندگان از مطالبی است که قرار است به آنها آموزش داده شود. متخصصان آموزشی پیشنهاد می‌کنند که در ارزشیابی آغازین ترکیبی از آزمون رفتارهای ورودی و آزمون پیش‌تست اجرا گردد.



هدف از ارزشیابی تشخیصی یکی، قرار دادن دانش‌آموزان در جای مناسب برای شروع به یادگیری مطالب تازه درسی، و دیگری کشف علل مشکلات یادگیری آنان در رسیدن به هدف‌های آموزشی است. این نوع ارزشیابی، اگر در آغاز کار آموزشی انجام‌گیرد هدفش جایگزینی دانش‌آموزان در موقعیت‌های مناسب آموزشی و یادگیری است. این نوع جایگزینی تعیین بهترین نقطه شروع آموزشی از لحاظ تأثیرگذاری جریان آموزشی بر دانش‌آموزان است.

ضمناً، از طریق این نوع ارزشیابی، می‌توان تشخیص داد که دانش‌آموزان در یادگیری مطالب پیش‌نیاز برای درسی تازه که در دروس پیشین می‌بایست آموخته باشند توفیق کامل داشته‌اند یا از این لحاظ کم و کسرهایی دارند. اما اگر این ارزشیابی همزمان با فعالیت آموزشی اجرا شود، هدف آن یافتن علل شکست‌های دانش‌آموزان در یادگیری است.

آنجا که لازم باشد دانش‌آموزان را بر مبنای استعداد یا علایق به رشته‌ها یا گروه‌های مختلف تقسیم‌بندی کنند، از ارزشیابی تشخیصی نوع اول یعنی ارزشیابی برای جایگزینی استفاده می‌نمایند. و آنجا که هدف بهبود آموزش و یادگیری است، این نوع ارزشیابی بیشتر به منظور تشخیص و شناسایی علل و عوامل شکست دانش‌آموزان به کار می‌رود.

ارزشیابی تشخیصی مورد استفاده برای جایگزینی دانش‌آموزان که در شروع کار آموزش اجرا می‌شود عمدتاً بر نتایج آزمون‌های تراکمی متکی است، زیرا هدف آن تعیین این مطلب است که آیا دانش‌آموزان مطالب دروس پیشین را به خوبی آموخته‌اند یا نه

دسته دوم: ارزشیابی تکوینی

این نوع ارزشیابی به منظور کمک به اصلاح موضوع مورد ارزشیابی یعنی برنامه یا روش آموزشی به کار می‌رود و هدف از آن آگاه کردن تولیدکنندگان برنامه از نواقص برنامه خود و کمک به اصلاح آنهاست اگر ارزشیابی تکوینی

برای دانش‌آموزان به کار رود هدف آگاهی از میزان و نحوه یادگیری آنان برای تعیین نقاط قوت و ضعف و اصلاح آنها و تشخیص مشکلات روش آموزشی معلم است. آزمون‌های تکوینی بیشتر وابسته همه ملاک مطلق هستند. آنطور که از اسم آن پیداست، زمانی به اجرا در می‌آید که فعالیت آموزشی هنوز جریان دارد و یادگیری دانش‌آموزان در حال تکوین یا شکل‌گیری است. این ارزشیابی نه به منظور نمره دادن و صدور گواهینامه به کار می‌رود نه برای قضاوت درباره اثربخشی معلم و برنامه آموزشی، بلکه هدف اصلی آن کمک به بهبود امر یادگیری و دفع نواقص آموزشی است. در این نوع ارزشیابی آزمون‌های کوتاه تکوینی، در پایان واحدهای کوچک آموزشی، اجرا می‌شوند و به این وسیله تعیین می‌گردد که یادگیرندگان در رسیدن به حد مدرک تسلط در یادگیری هدف‌های آموزشی واحدهای معین چه وضعی دارند، تا در صورت لزوم، از طریق آموزش ترمیمی، در رفع اشکالات یادگیری به آنها کمک شود.

دسته سوم: ارزشیابی تشخیصی

نوعی ارزشیابی تکوینی که هدف آن تشخیص مشکلات یادگیری دانش‌آموزان در یک موضوع درسی است و کمک می‌کند تا با سنجش عملکرد دانش‌آموزان در یک موضوع درسی تصویر دقیقی از کج‌فهمی‌ها که منجر به اشتباهات آنان می‌شود بدست آید معمولاً در جریان آموزش انجام می‌گیرد و آزمون‌های وابسته به ملاک بهتر از وابسته به هنجار است.

دسته چهارم: ارزشیابی تراکمی

هدف ارزشیابی تراکمی تعیین مقدار یادگیری دانش‌آموزان و دانشجویان در طول یک دوره آموزش، به منظور نمره دادن و صدور گواهینامه، یا قضاوت درباره اثربخشی کار معلم و برنامه درسی، یا مقایسه برنامه‌های مختلف با یکدیگر است. این نوع ارزیابی معمولاً در آخر دوره آموزشی به عمل می‌آید. اما می‌توان از آن به دفعات بیشتر نیز در طول دوره آموزشی آن هم در پایان واحدهای درسی استفاده کرد. در این نوع ارزشیابی، هدف اصلی تعیین یادگیری‌های متراکم دانش‌آموزان در طول دوره آموزشی است و از نتایج برای اصلاح روش آموزش معلم و یا رفع اشکالات یادگیری دانش‌آموزان چندان استفاده نمی‌شود. در این ارزشیابی تمامی آموخته‌های دانش‌آموزان در طی یک دوره تعیین می‌شوند و هدف آن نمره دادن به دانش‌آموزان و قضاوت درباره اثر بخشی کار معلم و برنامه درسی است معمولاً در پایان دوره آموزشی صورت می‌گیرد این آزمون‌ها نمونه کاملی از محتوای درس و هدف‌های آموزشی را به دست می‌دهد. در مقایسه ارزشیابی تکوینی و تراکمی باید گفت که ارزشیابی تکوینی در حین اجرای برنامه درسی و با هدف کشف نقاط قوت و ضعف و اصلاح آنها و کمک به تولیدکنندگان برنامه صورت می‌گیرد. در حالی که ارزشیابی

تراکمی در پایان دوره آموزشی انجام می‌گیرد و از نتایج ارزشیابی تراکمی در تصمیمات مربوط به قطع یا گسترش برنامه استفاده می‌گردد.

دسته‌بندی ارزشیابی‌های آموزشی با توجه به ارزشیابان

در دسته‌بندی ارزشیابی‌های آموزشی با توجه به ارزشیابان به دو دسته درونی و بیرونی برخورد می‌کنیم که رابطه بسیار نزدیکی با ارزشیابی تکوینی و تراکمی دارند. منظور از ارزشیابی درونی این است که ارزشیابی توسط کارکنان تهیه و اجرای برنامه انجام می‌شود و منظور از ارزشیابی بیرونی این است که ارزشیابی به وسیله افرادی خارج از برنامه انجام می‌شود ارزشیابی درونی برای ارزشیابی تکوینی و ارزشیابی بیرونی برای ارزشیابی تراکمی مناسب‌تر می‌باشد. ارزشیابی بیرونی معادل ارزشیابی تراکمی به معنی وسیع آن است.

الف) ارزشیابی درونی^۱

ارزشیابی درونی یا همان ارزشیابی تکوینی داوری درباره کیفیت یا ارزش مواد آموزشی، روش‌های آموزشی، برنامه‌های درسی و یا برنامه‌های آموزشی در حین طراحی و اجرای آنها است. هدف ارزشیابی درونی کمک به شکل‌گیری، تجدید نظر و اصلاح و بهبود محصول آموزشی است. ارزشیابی درونی بیشتر بدین منظور بکار می‌رود که نشان دهد یادگیرندگان به هدفهای برنامه رسیده‌اند یا نه.

ب) ارزشیابی بیرونی^۲

ارزشیابی بیرونی معادل ارزشیابی تراکمی به معنای وسیع آن است. مقصود از اجرای ارزشیابی بیرونی این است که مشخص شود دانش آموزانی که با روش یا برنامه خاصی آموزش دیده‌اند نسبت به دانش‌آموزانی که با روش یا برنامه دیگری آموزش دیده‌اند چگونه است.

مکیت و مک گلین ارزشیابی درونی را با ارزشیابی از خود یا خود ارزشیابی معادل بکار می‌برند اما ارزشیابی بیرونی را به بازمینی و واری فعالیت‌های آموزشگاه می‌پندارند.

هر دو نوع ارزشیابی درونی و ارزشیابی بیرونی دارای محاسن و معایبی هستند به عنوان مثال ارزشیاب درونی خیلی بیشتر از هرکسی درباره برنامه مورد ارزیابی اطلاع دارد. اما ممکن است بی طرفانه عمل نکند اما ارزشیاب بیرونی از ارزشیاب درونی عینی‌تر و بی‌طرف‌تر است اما برای وی مشکل است که به اندازه ارزشیابی درونی از برنامه مطلع باشد.

^۱Internal evaluation

^۲External evaluation

نکات کلیدی و مهم:

- در رویکرد مبتنی بر نظر متخصصان تاکید اصلی بر کاربرد نظر متخصصان در قضاوت در مورد کیفیت فعالیت‌های آموزش و پرورش است.
- رویکرد عینی‌گرا پیرو فلسفه عینیت‌گرایی است. به عبارت بهتر در این رویکرد بیشترین خوبی آن است که به بیشترین افراد سود برساند.
- رویکرد کثرت‌گرایی بر این اندیشه استوار است که ارزش، به تاثیر برنامه بر فرد فرد افراد وابسته است نه بر اکثریت آنان.
- مهمترین نوع ارزشیابی از نگاه استافل بیم و همکاران در الگوی سیپ ارزشیابی از بافت است.
- هدف ارزشیابی از بافت در الگوی سیپ فراهم آوردن یک زمینه منطقی برای تعیین هدفهای آموزشی است.
- روش‌های ارزشیابی از بافت عمدتاً توصیفی و مقایسه‌ای هستند.
- الگوی ارزشیابی یوسی ال.ای بسیار شبیه الگوی سیپ است با این تفاوت که آنچه در سیپ فرایند نامیده می‌شود در الگوی یوسی ال.ای تغییر نامیده می‌شود.
- در رویکرد ارزشیابی مبتنی بر مصرف‌کننده ارزشیابی برابر است با تعیین میزان مطلوبیت یک فرآورده آموزشی.
- ارزشیابی هدف آزاد را اسکریون به عنوان جانشینی برای ارزشیابی مبتنی بر هدف نمی‌داند بلکه آن را مکمل ارزشیابی مبتنی بر هدف می‌داند.
- ارزشیاب درونی خیلی بیشتر از هرکسی درباره برنامه مورد ارزیابی اطلاع دارد. اما ممکن است بی طرفانه عمل نکند.
- ارزشیاب بیرونی از ارزشیاب درونی عینی‌تر و بی‌طرف‌تر است اما برای وی مشکل است که به اندازه ارزشیاب درونی از برنامه مطلع باشد.
- ارزشیابی از آموزشگاه مستلزم گردآوری اطلاعات لازم درباره همه جنبه‌های آموزشگاه است.
- کاربرد ارزشیابی تکوینی: شناسایی نقاط قوت و ضعفها، تشخیص مشکلات آموزشی معلم، تشخیص واقعی برنامه
- ارزشیابی تشخیصی به منظور تشخیص مشکلات یادگیری دانش‌آموزان مورد استفاده قرار می‌گیرد.
- ارزشیابی زمانی مورد استفاده قرار می‌گیرد که معلم با مشکلات مبرم و مکرری در یک یا چند دانش‌آموز روبرو شود که با روش اصلاحی معمول ارزیابی تکوینی قابل رفع نیستند.
- هدف سنجش تشخیصی تعیین علت‌های مشکلات مکرر یادگیری و تدوین نقشه برای اعمال اصلاحی است.
- آزمونهای مورد استفاده سنجش تشخیصی آزمونهای وابسته به هنجار ویژه می‌باشد.
- در ارزشیابی تراکمی هم از آزمونهای وابسته به ملاک و هم از آزمونهای وابسته به هنجار استفاده می‌شود.

- ارزشیابی تکوینی تصمیمات مربوط به توسعه و تولید برنامه را شامل می‌شود.
- ارزشیابی تراکمی تصمیمات مربوط به ادامه، تبلیغ و گسترش برنامه را شامل می‌شود.
- در آزمونهای تشخیصی از آزمونهای وابسته به ملاک، وابسته به هنجار و مشاهده مورد استفاده قرار می‌گیرد.
- ارزشیابی درونی بیشتر مناسب ارزشیابی تکوینی است.
- ارزشیابی بیرونی مناسب ارزشیابی تراکمی است.
- ارزشیابی درونی برابر با خود ارزشیابی می‌باشد.

فصل پنجم:

مراحل و فعالیت‌های ارزشیابی آموزشی

مراحل ارزشیابی آموزشی

صرفنظر از اینکه چه چیزی مورد ارزشیابی قرار می‌گیرد و کدام الگوی ارزشیابی به کار بسته می‌شود همیشه فرایند ارزشیابی یکسان است از آنجا که ارزشیابی برای تصمیم‌گیری استفاده می‌شود، این گونه تصمیم‌گیری‌ها در سه دسته مختلف قرار می‌گیرد:

- ۱- مرحله طراحی: چه کاری انجام خواهیم داد؟ زمان آینده.
- ۲- مرحله فرایندی: چگونه انجام می‌دهیم؟ زمان حال.
- ۳- مرحله فرآورده‌ای: چگونه انجام دادیم؟ زمان گذشته.

الف) مرحله طراحی^۱

مرحله طراحی شامل تصمیماتی است درباره فعالیتهایی که قرار است انجام گیرد و به آینده مربوط می‌باشد. در این مرحله اقدامات زیر صورت می‌پذیرد:

۱. تحلیل موقعیت
۲. تعیین و توصیف هدفها
۳. توصیف پیش‌نیازها یا رفتارهای ورودی
۴. انتخاب و تولید وسایل اندازه‌گیری
۵. توصیف راهبردها یا استراتژی‌ها
۶. انتخاب طرح پژوهش
۷. تدارک یک برنامه زمان‌بندی.

الف) تحلیل موقعیت

در مرحله تحلیل موقعیت که نخستین گام مرحله طرح‌ریزی ارزشیابی آموزشی است به جمع‌آوری سوابق اطلاعاتی و برآورد و تعیین موانع موجود پرداخته می‌شود. در این مرحله تشخیص موانع و محدودیت‌ها و یک برآورد واقع‌بینانه از امکانات برای تعیین هدف‌های ارزشیابی ضروری است. در ارزشیابی پیشرفت تحصیلی، به منظور تحلیل موقعیت موجود معلم باید از آموخته‌های قبلی دانش‌آموزان خود اطلاعات کافی کسب نماید.

ب) تعیین و توصیف هدفها

هدف‌های ارزشیابی آموزشی معمولاً در دو سطح بیان می‌شوند:

- (۱) به صورت هدف‌های کلی: مبهم‌اند، کلی‌اند و به سادگی قابل اندازه‌گیری نیستند.
 - (۲) به صورت هدف‌های جزئی: دقیق، اندازه‌پذیر و قابل تبدیل به کمیت هستند.
- هدف‌های کلی را معمولاً غایت و هدف‌های جزئی را هدف می‌نامند. هدف‌های کلی معمولاً غیرقابل اندازه‌گیری هستند. اغلب هدف‌های کلی تحلیل و به هدف‌های دقیق تبدیل می‌شوند.
- هدف‌ها به دو دسته کلی: (۱) هدف‌های فرآیندی و (۲) هدف‌های فرآورده‌ای یا تولیدی تقسیم می‌شوند. هدف‌های فرآیندی به کوشش‌ها و اقداماتی اشاره می‌کنند که در ضمن اجرای برنامه انجام خواهند گرفت. در مقابل هدف‌های فرآورده‌ای به بازده‌ها و نتایج کوشش‌ها و اقدامات اشاره می‌کنند.
- در ارزشیابی از برنامه‌های آموزشی هر دو نوع هدف‌های آموزشی فرآیندی و هدف‌های فرآورده‌ای مورد

استفاده قرار می‌گیرند لکن در ارزشیابی از پیشرفت تحصیلی غالباً هدفهای فرآورده‌ای مورد استفاده قرار می‌گیرند.

ج) توصیف پیش‌نیازها یا رفتارهای ورودی

محقق شدن هدفهای آموزشی به وجود هدفهای رفتاری^۱ وابسته است. رفتارهای ورودی همان پیش‌نیازها هستند که یادگیرنده پیش از شروع به یادگیری هدفهای تازه باید آنها را کسب کرده باشد. برای تهیه و توصیف رفتارهای ورودی از روش تحلیل تکلیف استفاده می‌شود.

در تحلیل تکلیف هریک از اهداف آموزشی مشخص می‌گردد و سپس هدفهای جزئی لازم که آنها را پوشش می‌دهند برای هر هدف کلی تدارک دیده می‌شود تا با جامعه عمل پوشیدن هر کدام از آنها هدفهای کلی هم محقق شوند.

د) انتخاب و تولید وسایل اندازه‌گیری

تهیه و تولید وسایل اندازه‌گیری کاری است که به تخصص و صلاحیت نیاز دارد پس باید آن را به متخصصان آن واگذار کرد.

ر) توصیف راهبردها یا استراتژی^۲

منظور از استراتژی یا راهبرد فعالیت یا تدبیری است که به منظور کمک به تحقق یافتن هدفهای برنامه به کار می‌رود. روش تحلیل تکلیف، مرور یا بازنگری نمونه‌ای از استراتژی‌ها هستند.

ه) انتخاب طرح پژوهش

طرح‌های پژوهشی بسیار متفاوت‌اند و غالباً به دو گونه طرح‌های آزمایشی و طرح‌های شبه آزمایشی تقسیم می‌شوند.

ی) تدارک یک برنامه زمان‌بندی

یک برنامه زمان‌بندی دربرگیرنده فهرستی از فعالیت‌های اصلی مربوط به اقدامات پیشنهادی و زمان شروع و خاتمه هر فعالیت است. روش چارتی گانت روش پیشنهادی گی برای تدارک یک برنامه زمان‌بندی است.

ب) مرحله فرآیندی یا اجرائی^۳

دومین مرحله ارزشیابی آموزشی مرحله فرآیندی است. مرحله فرآیندی شامل تصمیمات مربوط به اقدامات عملی و اجرایی است و هدف اصلی این است که با استفاده از ارزشیابی مرحله فرآیندی به هدایت فعالیت‌های آتی برای

^۱Entry behaviors
^۲Strategy
^۳Process phase

بهبود فعالیتهای آموزشی و یادگیری پرداخته شود و در واقع نوعی ارزشیابی تکوینی است و با زمان حال سروکار دارد. هدف اصلی ارزشیابی در مرحله فرآیندی هدایت فعالیتهای آتی برای بهبود فعالیتهای آموزشی و یادگیری است. از جمله نخستین اقدامات این مرحله اجرای پیش آزمون و رفتار ورودی است در دیدگاه گی مرحله فرآیندی ارزشیابی آموزشی همان ارزشیابی تکوینی است.

ج) مرحله فرآوردهای^۱

در مرحله فرآوردهای تصمیمات در پایان برنامه یعنی زمانی که فعالیتها کامل می شوند اتخاذ می گردد و در واقع ارزشیابی تراکمی می باشد با این هدف که داده‌هایی گردآوری گردد تا متخصص ارزشیابی را در تصمیم‌گیری درباره اثربخشی کل برنامه کمک کند در این مرحله تعیین می‌شود، هدف‌های از قبل تعیین شده تا چه میزان تحقق یافته‌اند.

در این مرحله تصمیمات مربوط به اثربخشی موضوع مورد ارزشیابی اتخاذ می‌گردند. نام دیگر مرحله فرآوردهای ارزشیابی تراکمی است. در این مرحله نتایج آزمون‌های پیشرفت تحصیلی، نگرش و نظایر آن مورد استفاده و تحلیل قرار می‌گیرند. هدف اصلی این مرحله جمع‌آوری داده‌هایی است که متخصص ارزشیابی را در تصمیم‌گیری‌هایش درباره اثربخشی کلی روش آموزش، برنامه درسی و... کمک نماید.

سه راه عمده استفاده از نتایج مرحله فرآوردهای:

۱. اول: نتایج بازخورد و رهنمودی برای همه ذینفعان است از جمله دانش‌آموزان و معلمان و...
۲. دوم: نتایج بازخورد و رهنمودی برای همه تصمیم‌گیرندگان خارجی است مانند والدین و مدیران و...
۳. سوم: نتایج بازخورد و رهنمودی برای گروه‌های مختلفی با توجه به نتایج ارزشیابی خواهد بود.

فعالیت‌های ارزشیابی آموزشی

فعالیت‌های ارزشیابی آموزشی شامل مراحل زیر است:

- ۱- تعیین، انتخاب، پالایش یا تغییر غایتهای برنامه و هدفهای ارزشیابی
- ۲- مشخص کردن استانداردها- هدفها
- ۳- تهیه طرح ارزشیابی مناسب
- ۴- انتخاب یا تولید روش‌های جمع‌آوری داده‌ها
- ۵- جمع‌آوری داده‌های مناسب

۶- پردازش، خلاصه کردن و تحلیل داده‌ها

۷- مقایسه داده‌ها با استانداردها

۸- گزارش و بازخورد دادن از نتایج

۹- هزینه - سود / اثربخشی

خلاصه تصمیمات و فعالیت‌های سه مرحله اصلی ارزشیابی

مرحله ارزشیابی	زمان اجرا	نوع تصمیم	فعالیت‌ها
مرحله طرح‌ریزی	پیش از اجرای برنامه و آموزش	<ul style="list-style-type: none"> • نتایج حاصل • اعمالی که باید انجام شود 	<p>۱. تحلیل موقعیت</p> <p>۲. تعیین هدف</p> <p>۳. توصیف پیش نیازها</p> <p>۴. تهیه وسایل اندازه‌گیری</p> <p>۵. توصیف استراتژیها</p> <p>۶. انتخاب طرح پژوهش</p> <p>۷. تدارک برنامه زمانبندی</p>
مرحله فرایندی	ضمن اجرای برنامه و آموزش	<ul style="list-style-type: none"> • میزان اجرایی شدن مراحل طرح‌ریزی • تغییرات لازم برای پیشرفت 	<p>۱. اجرای پیش آزمون</p> <p>۲. اجرای آزمون رفتار ورودی</p> <p>۳. سنجش مناسب بودن هدف‌ها</p> <p>۴. جمع‌آوری داده‌ها بصورت دوره ای</p> <p>۵. تحلیل اثربخشی استراتژیها</p>
مرحله فراورده ای	پس از اجرای برنامه و آموزش	<ul style="list-style-type: none"> • اثربخشی کلی آموزش و برنامه • اقدامات آتی 	<p>۱. جمع‌آوری داده‌های مربوط به اهداف؛</p> <p>۲. جمع‌آوری داده‌های مربوط به بازده‌های پیش‌بینی نشده</p> <p>۳. تحلیل و تفسیر داده‌ها</p> <p>۴. تهیه گزارش</p>

نکات کلیدی و مهم :

- تحلیل موقعیت نخستین گام مرحله طرحریزی ارزشیابی آموزشی است.
- در ارزشیابی از برنامه و پروژه، هدفهای فرآیندی ارزشیابی می‌شوند.
- هدفهای فرآیندی کمتر با رفتار یادگیرندگان بلکه بیشتر با راهبردها و فعالیتهای سرو کار دارند.
- هدفها بر اساس ربط داشتن، اندازه پذیر بودن و مهم بودن ارزیابی می‌شوند.
- در ارزشیابی پیشرفت تحصیلی، به منظور تحلیل موقعیت موجود معلم باید از آموخته‌های قبلی دانش‌آموزان خود اطلاعات کافی کسب نماید.
- در مرحله تحلیل موقعیت از مراحل طرحریزی ارزشیابی آموزشی تشخیص موانع و محدودیت‌ها و یک برآورد واقع‌بینانه از امکانات برای تعیین هدفهای ارزشیابی ضروری است.
- آموزش و ارزشیابی نظامدار مستلزم تهیه، تدوین و اندازه‌گیری پیش‌نیازها یا رفتار ورودی هستند.
- رفتار ورودی در آغاز سال تحصیلی یا در ابتدای آموزش یک درس جدید انجام می‌گیرد.
- تحلیل تکلیف
- روش تحلیل تکلیف: ۱- برای تهیه رفتار ورودی یا پیش‌نیاز، ۲- استخراج هدفهای دقیق از غایت‌های کلی و ۳- سلسله مراتب هدفهای آموزشی ضروری است.
- در دیدگاه گی مرحله فرآیندی ارزشیابی آموزشی همان ارزشیابی تکوینی است.
- نام دیگر مرحله فرآورده‌ای، ارزشیابی تراکمی است.

فصل هشتم: ارزشیابی پیشرفت تحصیلی

تعریف ارزشیابی پیشرفت تحصیلی

ارزشیابی پیشرفت تحصیلی را می‌توان به صورت زیر تعریف کرد: سنجش عملکرد یادگیرندگان و مقایسه نتایج حاصل با هدف‌های آموزشی از پیش تعیین شده به منظور تصمیم‌گیری درباره این که آیا فعالیت‌های آموزشی معلم و کوشش‌های یادگیری دانش‌آموزان یا دانشجویان به نتایج مطلوب انجامیده‌اند و به چه میزانی بنابراین در ارزشیابی پیشرفت تحصیلی دو اقدام اساسی ضروری است:

۱- تعیین هدفهای آموزشی

۲- سنجش یا اندازه‌گیری عملکرد یادگیرندگان

هدفهای آموزشی یا هدفهای یادگیری

هدفهای آموزشی را به دو دسته هدف‌های کلی که همان غایات آموزشی هستند و هدف‌های دقیق که به آنها هدف‌های رفتاری نیز گفته می‌شود تقسیم‌بندی می‌کنند هدف‌های کلی معمولاً در سطح وزارت تهیه و تدوین

می‌شود و به صورت مقاصد آرمانی دوره‌های آموزشی در اختیار معلمان و مربیان قرار می‌گیرند در مقابل این غایت‌ها نوع دیگری هدف دقیق و عینی وجود دارد که آن را غالباً خود معلمان و متخصصان آموزشی با توجه به غایت‌های کلی و با استفاده از محتوای دروس مختلف تهیه می‌کنند. امتیاز مهم هدف‌های رفتاری این است که به سهولت قابل شناسایی و اندازه‌گیری هستند.

متخصصان آموزشی و ارزشیابی توصیه می‌کنند برای نوشتن هدف‌های رفتاری باید سه ملاک مدنظر قرارگیرد:

۱. اولاً این اهداف برحسب رفتار اندازه‌پذیر دانش‌آموز نوشته شود.

۲. ثانیاً دارای شرایط یا وضعیت عملکرد باشد.

۳. ثالثاً این هدف دارای ملاک عملکرد باشد.

به‌عنوان مثال «دانش‌آموز بتواند از میان فهرستی از ماشین‌ها و وسایل دیگر که معلم در اختیار او می‌گذارد، حداقل پنج ماشین را انتخاب کند.» این هدف هر سه ویژگی اهداف رفتاری را دارا است.

طبقه‌بندی هدف‌های آموزشی

هدف‌های آموزشی بسیار متنوع‌اند و برای سهولت مطالعه آنها را به گونه‌های مختلف طبقه‌بندی کرده‌اند. معروفترین طبقه‌بندی از هدف‌های آموزشی توسط بنجامین.اس. بلوم انجام گرفته است. بلوم و همکارانش هدف‌های آموزشی را در سه حیطه طبقه‌بندی کرده‌اند.

۱. حیطه اول شناختی

۲. حیطه دوم عاطفی

۳. حیطه سوم روانی - حرکتی

حیطه اول شناختی

اهداف حوزه شناختی به جریاناتی که با فعالیت‌های ذهنی و فکری آدمی سروکار دارند، مربوط می‌شود از این لحاظ این حوزه مهمترین حوزه یادگیری به شمار می‌آید.

آزمون‌های مورد نیاز برای سنجش هدف‌های حوزه شناختی آزمون‌های توانایی شناختی هستند که غالباً به صورت کتبی تهیه و اجرا می‌شوند.

حیطه دوم عاطفی

حوزه عاطفی شامل هدف‌هایی است که تغییرات حاصل در علاقه‌ها، نگرش‌ها و ارزش‌ها و نیز رشد ارج‌شناسی و

سازگاری را نشان می‌دهد.

حیطه سوم روانی - حرکتی

حوزه روانی - حرکتی به زمینه مهارت‌های حرکتی یا حرکات بدنی معروف است به سخن دیگر هر فعالیتی که علاوه بر جنبه روانی - ذهنی دارای جنبه جسمانی باشد در این حوزه جای دارد.

بلوم حوزه شناختی را به شش طبقه کلی طبقه‌بندی کرده است که عبارتند از:

۱. دانش: شامل یادآوری (بازخوانی و بازشناسی) امور است
۲. فهمیدن: ترجمه یا برگردان - تفسیر یا خلاصه - کشف شباهتها و تفاوت‌ها
۳. کار بستن: توانایی استفاده از امور انتزاعی، قواعد و قوانین و اصول و روش‌ها و... در موقعیت‌های عینی و عملی.
۴. تحلیل: توانایی شکستن یک مفهوم به اجزای تشکیل دهنده.
۵. ترکیب: توانایی کنار هم چیدن اجزا و ایجاد چیز جدید - آفرینندگی و خلاقیت.
۶. ارزشیابی: توانایی قضاوت کمی و کیفی با توجه به ملاک‌های معین.

کراتول حوزه عاطفی را در یک طبقه‌بندی به پنج طبقه تقسیم کرده است:

۱. دریافت کردن (توجه کردن): حساسیت نسبت به وجود برخی اشیاء و پدیده‌ها.
۲. پاسخ‌دادن: توجه فعالانه به پدیده‌ها و انجام کارهایی درباره آنها.
۳. ارزش‌گذاردن: احساس یا باور حاکی از ارزشمند بودن یک چیز یا یک اندیشه و...
۴. سازمان دادن به ارزش‌ها: دسته‌بندی و نظام‌مند کردن ارزش‌ها.
۵. شخصیت پذیرفتن: تبدیل نظام ارزشی به یک سبک زندگی یا یک دیدگاه فلسفی.

حوزه روانی - حرکتی سیمپسون به هفت طبقه تقسیم می‌شود که عبارتند از:

۱. ادراک: کاربرد حواس برای هدایت کنش‌های حرکتی.
۲. آمادگی: آماده بودن برای انجام یک رشته عمل یا اعمال معین.
۳. پاسخ هدایت شده: عمل به کمک هدایت یک الگو
۴. مکانیسم یا عادت: انجام اعمال عادت‌ی و خودکار
۵. پاسخ پیچیده آشکار: ترکیب تعدادی از عادت‌ها
۶. انطباق: استفاده از مهارت‌های قبلاً آموخته شده برای انجام تکالیف تازه اما مربوط به تکالیف گذشته.

۷. ابتکار: خلق الگوهای تازه حرکتی برای حل مسائل غیر معمول یا برخورد با موقعیت‌های جدید.

جدول مشخصات

مهمترین گام در تهیه آزمون‌های پیشرفت تحصیلی تهیه جدول مشخصات برای موضوعی است که قرار است آزمون آن تهیه گردد. جدول مشخصات یک جدول دویبعدی است با تعدادی ردیف و ستون متناسب با محتوا و اهداف آموزشی درس در ردیف بالای جدول که بعد محتوا نام دارد، اجزای مطالب درسی نوشته می‌شود و در ستون کنار آن که بعد هدف نامیده می‌شود، اهداف آموزشی قرار می‌گیرد. معمولاً در بعد هدف جدول، هدف‌های آموزشی بر طبق طبقه‌بندی‌های اهداف آموزشی دسته‌بندی می‌شوند. جدول مشخصات هم برای تهیه طرح درس و هم برای تهیه آزمون مفید می‌باشد.

استفاده از جدول مشخصات برای تهیه آزمون، نشان‌دهنده اعتبار محتوایی آزمون می‌باشد.

انواع جدول‌های مشخصات

هم در تهیه آزمون‌های وابسته به ملاک و هم در تهیه آزمون‌های وابسته به هنجار استفاده از جدول مشخصات ضروری است. با این حال تفاوت عمده بین این دو نوع جدول مشخصات در آن است که در بعد هدف جدول مشخصات آزمون‌های وابسته به هنجار، معمولاً بجای هدف‌های دقیق آموزشی، اصطلاحات و عبارات‌های کلی‌تری به کار می‌روند که معرف طبقات هدف‌های آموزشی یا انواع یادگیری هستند.

علاوه بر این دو نوع جدول، نوع دیگری از جدول مشخصات وجود دارد که ویژه ارزشیابی تکوینی است.

روش پوفام برای تهیه طرح آزمون وابسته به ملاک

آزمون پوفام دارای بخش‌های زیر است:

(۱) توصیف کلی

(۲) سؤال نمونه

(۳) ویژگی‌های محرک

(۴) ویژگی‌های پاسخ

(۵) توضیحات اضافی

وسایل کتبی اندازه‌گیری پیشرفت تحصیلی

همان‌طور که قبلاً اشاره شد مهمترین وسیله اندازه‌گیری پیشرفت تحصیلی، آزمون است. به آزمون‌های پیشرفت

تحصیلی کتبی آزمون‌های مداد - کاغذی هم می‌گویند و آنها را به صورت زیر دسته‌بندی می‌کنند این آزمون‌ها را در ابتدا می‌توان به دو دسته عینی و غیرعینی (ذهنی) تقسیم کرد و سپس به دسته‌های کوچکتر تبدیل نمود. آزمون‌های عینی آزمون‌هایی هستند که هم سؤالات و هم جواب سؤالات را در اختیار آزمون‌شونده قرار می‌دهند و آزمون‌شوندگان دوباره بر روی جواب‌های داده شده اعمالی انجام می‌دهند یا تصمیماتی اتخاذ می‌کنند به همین سبب این آزمون‌ها را آزمون‌های بسته پاسخ نیز می‌نامند. آزمون‌های عینی به طور عمده از سه نوع اصلی چند گزینه‌ای، صحیح - غلط و جور کردنی تشکیل می‌گردد.

آزمون‌شونده در جواب دادن به سؤالات چند گزینه‌ای جواب صحیح را از میان تعدادی جواب پیشنهادی برمی‌گزیند، در سؤالات صحیح - غلط درست یا غلط بودن سؤالات را تعیین می‌کند و در سؤالات جورکردنی تعدادی سؤال را با تعدادی جواب جور می‌کند در تصحیح برگه‌های آزمون‌های عینی نظر شخصی مصحح هیچ دخالتی ندارد و از این رو به این آزمون‌ها آزمون‌های عینی می‌گویند.

آزمون‌های ذهنی که به آزمون‌های تشریحی یا انشایی مشهور می‌باشند، آزمون‌هایی هستند که در آنها سؤالات در اختیار آزمون‌شونده گذاشته می‌شود و او جواب سؤالات را خود آماده می‌کند و در برگه امتحانی می‌نویسد به همین سبب به آنها آزمون‌های بازپاسخ گفته می‌شود. علت نام‌گذاری این نوع آزمون‌ها به آزمون‌های ذهنی آن است که در تصحیح جواب‌های آنها ممکن است نظر شخصی مصحح نیز دخالت کند این نوع آزمون‌ها را نیز می‌توان به دو دسته آزمون‌های گسترده پاسخ و آزمون‌های محدود پاسخ تقسیم کرد. در آزمون‌های گسترده پاسخ آزمون‌شونده برای پاسخ هیچ محدودیتی ندارد اما در آزمون‌های محدود پاسخ آزمون‌شونده در دادن پاسخ به سؤالات آزادی کامل ندارد و سازنده آزمون پاسخ‌دهنده را ملزم می‌کند تا پاسخ‌های خود را در چارچوب شرایطی محدود سازد. نوع دیگری از آزمون پیشرفت تحصیلی وجود دارد که حد وسط انواع آزمون عینی و ذهنی قرار دارد که به این نوع آزمون‌های کوتاه پاسخ می‌گویند. در این آزمون‌ها سؤالات بسیار شبیه به سؤالات آزمون‌های تشریحی هستند، زیرا آزمون‌شونده خود باید پاسخ‌ها را تعیین کند اما سؤالات بسیار دقیق و مشخص هستند به نحوی که در تصحیح جواب‌های آنها نظر شخصی مصحح دخالت چندانی نمی‌کند.

فصل پنجم: انواع آزمون‌های پیشرفت تحصیلی

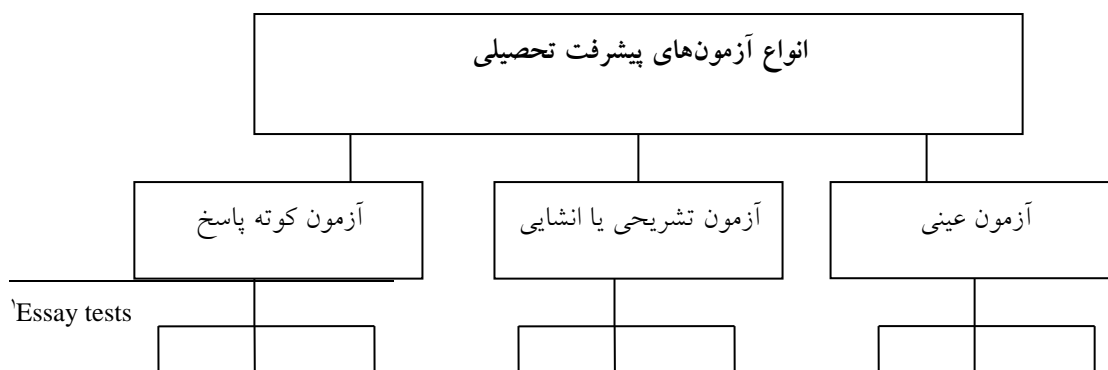
بخش نخست: آزمون‌های تشریحی یا انشایی

آزمون‌های تشریحی یا انشایی^۱

آزمون‌های تشریحی یا انشایی معمول ترین نوع آزمون‌های کلاسی است. آزمون‌های تشریحی (انشایی) را با توجه به آزادی عمل آزمون‌شونده در پاسخ دادن به سوال‌های آزمون به دو دسته تقسیم می‌کنند:

۱. گسترده پاسخ

۲. محدود پاسخ



الف) گسترده پاسخ^۱

در آزمون گسترده پاسخ هیچ‌گونه محدودیتی برای آزمون‌شونده منظور نمی‌شود و او عملاً آزاد است تا هرطور که مایل باشد خود را پیروانند و سازمان دهد. سؤال‌های آزمون‌های تشریحی گسترده پاسخ برای سنجش هدف‌های تحلیل، ترکیب و ارزشیابی مناسب‌ترین سؤال‌ها هستند.

ب) محدود پاسخ^۲

در آزمون‌های تشریحی محدود پاسخ پاسخگو آزادی کامل ندارد. در آزمون‌های محدود پاسخ هم از لحاظ زمان پاسخ‌دهی و هم از نظر مقدار پاسخ محدودیت‌هایی وجود دارد. سؤال‌های آزمون تشریحی محدود پاسخ برای اندازه‌گیری توانایی یادگیرندگان در سطوح فهمیدن، کار بستن و تحلیل به کار می‌روند. محدود ساختن آزمون‌شوندگان به یک موضوع مشخص تصحیح پاسخ‌های آنها را هماهنگ‌تر و آسان‌تر می‌کند. اما مانع ترکیب اندیشه‌ها و بیان آنها به صورت منسجم می‌شود. این نوع آزمون‌ها برای سنجش توانایی ترکیب و خلاقیت و ارزشیابی زیاد مناسب نیستند.

محاسن و معایب آزمون‌های تشریحی

مدافعان این نوع آزمون‌ها محاسن آنها را به شرح زیر شمرده‌اند:

- ۱- تهیه آزمون‌های تشریحی از آزمون‌های عینی آسانتر است.
- ۲- این نوع آزمون تنها وسیله موجود سنجش توانایی آزمون‌شونده در پروراندن جواب سؤال‌ها و بیان آنهاست.
- ۳- این آزمون‌ها توانایی پاسخ دادن به سؤال‌ها را می‌سنجد نه توانایی انتخاب پاسخ‌ها را.
- ۴- آنها موقعیت‌های واقعی‌تری را از آزمون‌های عینی به آزمون‌شوندگان عرضه می‌کنند.
- ۵- آنها بر روش مطالعه یادگیرندگان تأثیر مثبت به جای می‌گذارند.

در مقابل مخالفان دلایل زیر را علیه استفاده از این آزمون‌ها ارائه می‌دهند.

- ۱) این آزمون‌ها نمونه کوچکی از محتوای درس و هدفهای آموزش را اندازه می‌گیرند.
- ۲) تصحیح برگه‌های امتحانی این گونه آزمون‌ها نمی‌تواند با دقت و به طور عینی انجام گیرد.
- ۳) تصحیح برگه‌های امتحانی این آزمون‌ها بسیار وقت گیر است.

قواعد تهیه سؤال‌های تشریحی

- ۱- در نوشتن صورت سؤال‌های آزمون یک درس، با رسم جدول مشخصات آن درس، دقت کنید که سؤال‌ها به طور مستقیم به هدف‌های آموزشی مربوط شوند.
- ۲- سؤال‌های تشریحی را تنها به اندازه‌گیری هدف‌هایی محدود کنید که با سایر انواع آزمون‌ها به خوبی قابل اندازه‌گیری نیستند.
- ۳- صورت سؤال‌های تشریحی را با عبارات و کلمات واضح و روشن بنویسید و از کلی‌گیری و ابهام در بیان بپرهیزید.
- ۴- از کاربرد کلمات «چه کسی»، «چه وقت»، «کجا» و جز اینها بپرهیزید.
- ۵- تا حد امکان از سؤال‌های تازه و موقعیت‌های جدید استفاده کنید.
- ۶- سؤال‌های مربوط به موضوعات و عقاید بحث‌انگیز باید طوری طرح شوند که از آزمون‌شونده بخواهند تا شواهد لازم برای مستند کردن عقاید انتخابی را بیان کنند، نه اینکه از او بخواهد تا صرفاً عقاید شخصی خود را شرح دهد.
- ۷- به آزمون‌شوندگان حق انتخاب چند سؤال از میان تعدادی سؤال را ندهید.
- ۸- برای پاسخ دادن به سؤال‌ها، زمان کافی در نظر بگیرید و زمان هر سؤال را جداگانه مشخص کنید.
- ۹- با نوشتن سؤال‌هایی که جواب کوتاه نیاز دارند، تعداد آنها را افزایش دهید و از مشکل ضعف نمونه‌گیری آزمون‌های تشریحی بکاهید.
- ۱۰- عواملی را که در ارزشیابی آزمون‌های تشریحی دخالت خواهید داد از پیش تعیین کنید و آنها را به اطلاع آزمون‌شوندگان برسانید.

قواعد تصحیح پاسخ‌های سؤال‌های تشریحی

- ۱- پاسخ سؤال‌ها را تنها براساس هدفی که در سؤال گنجانیده شده است تصحیح کنید.
- ۲- با نوشتن یک پاسخ نمونه یا الگو برای هر سؤال به عنوان کلید، از دخالت عوامل نامربوط جلوگیری کنید.
- ۳- پاسخ را سؤال به سؤال تصحیح کنید نه ورقه به ورقه.
- ۴- هنگام تصحیح ورقه‌های امتحانی از شناسایی نام صاحبان آنها خودداری کنید.
- ۵- در صورت امکان از یکی دو نفر از همکارانتان بخواهید تا سؤال‌هایی را که شما تصحیح کرده‌اید تصحیح کنند.
- ۶- تمام پاسخ‌های آزمون‌شدگان به یک سؤال را در یک نشست و بدون وقفه زمانی تصحیح کنید.
- ۷- به نمرات سؤال‌هایی که قبلاً تصحیح کرده‌اید، نگاه نکنید.

۸- بر روی برگه‌های آزمون اشتباهات دانش‌آموزان را تصحیح کنید و اظهار نظرهای خود را بنویسید.

۹- از روش‌های متنوع نمره‌گذاری استفاده کنید.

روش‌های نمره‌گذاری آزمون‌های تشریحی

۱. روش تحلیلی

۲. روش کلی

۳. روش ویژگی‌های اصلی

روش تحلیلی^۱

در روش تحلیلی نمره‌گذاری که به آن روش امتیازبندی نیز گفته می‌شود، پاسخ نمونه به اجزای کوچکتری تقسیم می‌شود و برای هر جزء نمره یا امتیاز جداگانه‌ای در نظر گرفته می‌شود. روش تحلیلی برای آزمون‌های وابسته به ملاک مناسب‌تر است. در این روش عواملی چون قدرت بیان، ساختمان منطقی پاسخ و ذکر دلایل امتیازهای جداگانه‌ای دریافت می‌کنند.

روش کلی^۲

در روش کلی یا سراسری نمره‌گذاری که به آن روش درجه‌بندی نیز گفته می‌شود، پاسخ نمونه بصورت یک معیار به کار می‌رود. در این روش هیچ عامل واحدی امتیاز جداگانه‌ای دریافت نمی‌کند بلکه کل پاسخ و همه عوامل با هم نمره دریافت می‌کنند. به نکات زیر در مورد دو روش فوق توجه کنید:

- ایبل نمره‌گذاری به روش کلی و سراسری را آسان تر و سریع‌تر از روش تحلیلی می‌داند.
- گی در مورد دو روش کلی یا سراسری و روش تحلیلی بیان می‌دارد که هرچند هم برای آزمون‌های وابسته به ملاک و وابسته به هنجار می‌توان از روش کلی و روش تحلیلی استفاده کرد اما روش تحلیلی برای آزمون‌های وابسته به ملاک مناسب‌تر است و روش کلی یا سراسری برای آزمون‌های وابسته به هنجار مناسب‌تر است.
- ویرسما و جوزز بهترین روش نمره‌گذاری را ترکیبی از دو روش فوق می‌دانند.

روش ویژگی‌های اصلی

در این روش معلم یا مصحح ویژگی‌های اصلی پاسخ دانش‌آموز یا دانشجو به هر سؤال را سنجش می‌کند.

پرسش شفاهی^۱

یکی از تدابیر متداول و موثر معلمان در ارزشیابی‌های غیررسمی از یادگیری دانش‌آموزان، به منظور دادن بازخورد به آنان، تشویق و ترغیب آنان در یادگیری و رفع مشکلات و معایب یادگیری آنان استفاده از پرسش‌های شفاهی است. از پرسش شفاهی می‌توان برای بهبود روش‌های یادگیری دانش‌آموزان و شیوه‌های آموزشی معلم استفاده کرد. کاربرد پرسش‌های شفاهی بیشتر برای مقاصد تشخیصی است.

بخش دوم: آزمون‌های کوتاه پاسخ

آزمون‌های کوتاه پاسخ

آزمون کوتاه پاسخ حدوسط آزمون‌های تشریحی و عینی است. آزمون‌های کوتاه پاسخ از تعدادی سؤال مختصر تشکیل می‌شود که پاسخ‌های آنها به یک عدد، علامت، کلمه، عبارت یا جمله محدود می‌شود این سؤال‌ها بیشتر برای سنجش طبقه دانش مورد استفاده قرار می‌گیرند، با این حال صورت این گونه سؤالات را می‌توان طوری نوشت که هدف‌های طبقه فهمیدن و کار بستن را نیز سنجش کنند.

آزمون‌های کوتاه پاسخ را می‌توان به سه دسته طبقه‌بندی کرد:

۱. نوع اول پرسشی
۲. نوع دوم کامل کردنی
۳. نوع سوم تشخیصی یا تداعی

محاسن و معایب آزمون‌های کوتاه پاسخ

از امتیازهای مهم این آزمون‌ها این است که تهیه این نوع سؤالات ساده است و تصحیح پاسخ‌های آنها نیز دقیق‌تر و عینی‌تر از پاسخ‌های سؤال‌های تشریحی است از این رو بهتر است که از این آزمون‌ها برای سنجش هدف‌ها یا طبقات بالا استفاده نشود (تحلیل، ترکیب، کاربرد و ارزشیابی). همچنین این آزمون‌ها نسبت به آزمون‌های عینی چندگزینه‌ای، صحیح - غلط و جورکردنی اطلاعات تشخیصی بیشتری را در اختیار معلمان می‌گذارند. امتیاز دیگر آزمون‌های کوتاه پاسخ این است که در آنها حدس کورکورانه که از معایب عمده آزمون‌های عینی است، وجود ندارد.

همچنین تصحیح و نمره‌گذاری پاسخ‌های این آزمون‌ها به دقت و سرعت آزمون‌های عملی عینی نیست و این منجر به کاستن از میزان روایی و پایایی آزمون می‌گردد.

قواعد تهیه سوالهای کوتاه پاسخ

- ۱- برای آزمون خود جدول مشخصات تهیه کنید و سؤالات آزمون را با توجه به جدول مشخصات بنویسید.
- ۲- هر سؤال باید موضوع مهمی را شامل شود.
- ۳- صورت سؤال را کاملاً روشن بنویسید به گونه‌ای که به پاسخ مشخص و معینی نیاز داشته باشد.
- ۴- صورت سؤال را از روی مطالب کتاب عیناً نقل نکنید.
- ۵- در سؤال‌هایی که پاسخ‌های آنها اعداد هستند، واحد مقیاس و میزان دقتی را که در محاسبات باید رعایت شوند مشخص کنید.
- ۶- در سؤال‌های کوتاه پاسخ کامل کردنی، تنها عبارات و کلمات مهم را حذف کنید.
- ۷- در سؤال‌های کامل کردنی که برای ارزشیابی پیشرفت تحصیلی به کار می‌روند تعداد زیادی جای خالی منظور نکنید.
- ۸- جای خالی سؤال‌های کامل کردنی را تا آنجا که ممکن است در قسمت پایانی سؤال قرار دهید.
- ۹- جای خالی سؤال‌های کامل کردنی، را همواره یک اندازه تعیین کنید.
- ۱۰- تا حد امکان به جای سؤال‌های کامل کردنی، از سؤالات پرسشی استفاده کنید.
- ۱۱- در سؤال‌های کامل کردنی، از کاربرد اشارات دستوری و موارد دیگری که جواب سؤال را مشخص می‌کنند خودداری کنید.
- ۱۲- صورت سؤال را عین جملات کتاب ننویسید.

بخش سوم: آزمون‌های عینی: صحیح - غلط و جورکردنی

آزمون‌های عینی

همان‌طور که قبلاً اشاره شد آزمون‌های عینی آزمون‌هایی هستند که در آنها هم سؤال‌ها و هم جواب سؤال‌ها در اختیار آزمون‌شونده قرار می‌گیرد و آزمون‌شونده درباره جواب‌های داده شده تصمیماتی می‌گیرد یا اعمالی انجام می‌دهد. ویژگی مهم آزمون‌های عینی این است که تصحیح جواب آنها به طور کاملاً دقیق و عینی انجام می‌گیرد و نظر شخصی مصحح هیچ‌گونه دخالتی ندارد و وجه تمایز دیگر آزمون‌های عینی با آزمون‌های تشریحی این است که این آزمون‌ها توانایی بازشناسی یا تشخیص را می‌سنجد در حالی که آزمون‌های تشریحی عموماً توانایی بازشخوانی آزمون‌شوندگان را اندازه می‌گیرند.

آزمون‌های عینی از سه دسته اصلی تشکیل می‌شوند:

۱. صحیح - غلط
۲. جورکردنی
۳. چندگزینه‌ای

الف) آزمون‌های صحیح - غلط

در آزمون‌های صحیح - غلط تعدادی سوال یا جمله در اختیار آزمون‌شونده گذاشته می‌شود و او صحیح یا غلط بودن آنها را تعیین می‌کند. آزمون‌های صحیح - غلط ساده‌ترین صورت سؤال‌های عینی هستند و همین‌طور پوفام بیان می‌دارد که آزمون‌های صحیح - غلط متداولترین نوع سوال‌های عینی هستند و پرورشکاران از زمان سقراط از چنین روشی استفاده می‌کردند.

انواع آزمون‌های صحیح - غلط

آزمون‌های صحیح - غلط انواع مختلفی دارد که عبارتند از:

۱. نوع صحیح - غلط
۲. نوع بله - نه
۳. نوع اصلاحی
۴. نوع خوشه‌ای

۱) نوع صحیح - غلط

نوع صحیح - غلط یکی از انواع آزمون‌های صحیح - غلط است که در آن یک جمله در اختیار فرد گذاشته می‌شود و او صحت و سقم آن را برمی‌گزیند.

۲) نوع بله - نه

نوع بله - نه یکی دیگر از انواع آزمون‌های صحیح - غلط است که در آن جملات به صورت استفهامی بیان می‌شود.

۳) نوع اصلاحی

نوع اصلاحی یکی دیگر از انواع آزمون‌های صحیح - غلط است که در آن آزمون‌شونده تشخیص می‌دهد که غلطی وجود دارد و آن را اصلاح می‌کند.

۴) نوع خوشه‌ای

نوع خوشه‌ای یکی دیگر از انواع آزمون‌های صحیح - غلط است که به آن چند صحیح غلط گفته می‌شود تنه سوال به صورت جمله ناتمام نوشته می‌شود و به دنبال آن تعدادی جواب می‌آید.

محاسن و معایب آزمون‌های صحیح - غلط

هرچند که آزمون‌های صحیح - غلط از لحاظ تهیه ساده هستند و عینی بودن تصحیح جواب و تعداد زیاد آنها در هر جلسه امتحان از محاسن این آزمون‌ها می‌باشد ولی این آزمون‌ها اغلب برای سنجش اطلاعات جزئی و کم‌ارزش به کار می‌رود حدس کورکورانه تاثیر زیادی دارد و توانایی‌های طبقه‌های بالا را نمی‌سنجد البته ایبل این ایرادها را متوجه

طراحان سؤالات صحیح غلط می‌داند نه نوع سؤالات صحیح - غلط.

قواعد تهیه سؤال‌های صحیح - غلط

- ۱- سؤال‌ها را تا حد امکان مختصر، ساده و روشن بنویسید.
- ۲- جمله‌هایی را به کار برید که درست یا غلط بودن آنها برای کسانی که موضوع سؤال را یاد گرفته‌اند، آشکار باشد.
- ۳- از کاربرد کلماتی چون همه، بعضی وقت‌ها، غالباً، هرگز و از این قبیل پرهیز کنید.
- ۴- سعی کنید طول سؤال‌های غلط و سؤال‌های صحیح هم اندازه باشد.
- ۵- سعی کنید تعداد سؤال‌های غلط و سؤال‌های صحیح یک اندازه باشد.
- ۶- طول سؤال‌های صحیح و سؤال‌های غلط را هم اندازه بنویسید.
- ۷- سؤال‌های صحیح - غلط را از روی جملات کتاب بنویسید.
- ۸- سعی کنید هر سؤال به یک هدف آموزشی مهم مربوط باشد و از گنجانیدن مطالب بی‌اهمیت در سؤال پرهیز کنید.

ب) آزمون‌های جورکردنی

این آزمون‌ها شامل تعدادی سؤال می‌باشند که هر یک از آنها شامل دو ستون است. یک ستون معرف پرسش‌ها و ستون دیگر نشان دهنده پاسخ‌ها است. وظیفه آزمون‌شونده آن است که پاسخ‌ها را با پرسش‌های مربوط جور کند. این آزمون‌ها بیشتر برای سنجش توانایی انتقال یادگیری، توضیح و درک مطلب مناسب‌اند.

محاسن و معایب آزمون‌های جورکردنی

آزمون‌های جورکردنی می‌توانند نمونه نسبتاً بزرگی از محتوای درس و هدف‌های آموزشی را فراهم نمایند. همچنین تصحیح آنها با عینیت و سهولت بیشتری صورت می‌گیرد. این آزمون‌ها توانایی تشخیص بین امور را به خوبی می‌سنجد. سؤال‌های جورکردنی به دلایلی مورد انتقاد قرار می‌گیرند از جمله اینکه برخی راهنمایی‌های این سؤال‌ها مبهم هستند. بعضی پرسش‌ها و جواب‌ها بیش از حد طولانی، ناهمگون و مبهم هستند. بعضی از سؤالات این آزمون‌ها بر یادگیری طوطی‌وار تأکید دارند.

قواعد تهیه سؤال‌های جورکردنی

- ۱- پرسش‌ها و پاسخ‌های متجانس یا همگون انتخاب کنید.
- ۲- طول فهرست‌های پرسش و پاسخ را کوتاه انتخاب کنید و مطالب کمتری را در هر یک از پاسخ‌ها قرار دهید.
- ۳- در راهنمای سؤال، اطلاعات لازم را درباره نحوه مقایسه و جور کردن پرسش‌ها و پاسخ‌ها در اختیار دانش‌آموزان قرار دهید.
- ۴- همه پرسش‌ها و پاسخ‌های یک سؤال را در یک صفحه قرار دهید.

- ۵- در صورت امکان فهرست پاسخ‌ها را به طور منطقی مرتب کنید.
- ۶- در فهرست پرسش‌های سؤال‌های جور کردنی از کاربرد جملات نیمه تمام خودداری کنید.
- ۷- پرسش‌ها را با شماره و پاسخ‌ها را با حروف مشخص کنید.
- ۸- هر یک از پاسخ‌ها باید برای تمام پرسش‌های یک سؤال درست جلوه کند.
- ۹- از نوشتن سؤال‌هایی که پرسش‌ها و پاسخ‌های آنها کاملاً با هم جور هستند بپرهیزید.

ج) آزمون‌های چند گزینه‌ای

این آزمون‌ها شامل تعدادی سؤال است که هریک از آنها از یک قسمت اصلی و تعدادی گزینه تشکیل می‌شود و دانش‌آموز پاسخ را از میان گزینه‌ها انتخاب می‌کند هر سؤال چند گزینه‌ای شامل سه قسمت می‌باشد:

۱. قسمت اصلی یا تنه سؤال

۲. گزینه درست یا پاسخ سؤال

۳. گزینه‌های انحرافی

۱) قسمت اصلی یا تنه سؤال: که متن اصلی سوال را تشکیل می‌دهد و در برگرفته موضوعی است که سؤال باید آن را اندازه‌گیری کند.

۲) گزینه درست یا پاسخ سؤال: گزینه‌ای که پاسخ صحیح را دربر دارد.

۳) گزینه‌های انحرافی: نقش این گزینه‌ها منحرف کردن آزمون‌شوندگانی است که پاسخ درست را نمی‌دانند سؤالات چند گزینه‌ای سه نوع می‌باشد:

نوع اول: نوع تنها گزینه درست - که یک گزینه درست و بقیه غلط می‌باشند.

نوع دوم: نوع بهترین گزینه - تمام گزینه‌ها تا حدی درست می‌باشد ولی باید از همه درست‌تر را انتخاب نمود.

نوع سوم: نوع منفی - که در آن تنه سؤال به صورت منفی بیان می‌گردد و به غیر از یک گزینه بقیه گزینه‌ها غلط می‌باشد اما ویژگی اصلی سؤال چند گزینه‌ای تنه سؤال و تعداد گزینه می‌باشد.

محاسن و معایب آزمون‌های چند گزینه‌ای

با آزمون‌های چندگزینه‌ای می‌توان بازده‌های مهم یادگیری مانند توانایی استدلال و قضاوت را سنجید محتوای زیادی از درس را از آزمون شوندگان پرسید امکان حدس کورکورانه در آن کم می‌باشد.

پاسخ سؤالات به سادگی و عینیت کافی قابل اندازه‌گیری است و گزینه‌های انحرافی منبع بسیار مناسبی برای تشخیص دانش‌آموزان خواهد بود.

اما ساختن آزمون‌های چند گزینه‌ای و یافتن چند گزینه انحرافی دشوار است معلمان برای سنجش مطالب کم اهمیت از آن استفاده می‌کنند و دانش‌آموزان قوی که بیشتر از شاگردان معمولی قادر به پیدا کردن اشکالات پیچیدگی‌ها و نظرگاه‌های واگرا هستند با دادن جواب‌هایی که مورد نظر طراح سؤال نیست معمولاً اشتباه جواب می‌دهند.

قواعد تهیه سؤالات چند گزینه‌ای

در تهیه سؤالات چندگزینه‌ای نکات زیر قابل توجه می‌باشد:

- الف) هر سؤال باید یک موضوع مهم با یک هدف آموزشی را اندازه بگیرد.
- ب) مطالب اصلی سؤال به طور کامل در تنه سؤال نوشته شود.
- ج) برای اندازه‌گیری فرایندهای پیچیده ذهنی از موقعیت‌های جدید نه مثال‌های کلاسی استفاده کنید.
- د) در سؤالات منفی کلمات منفی را برجسته جلوه دهید.
- ه) در نوشتن سؤالاتی که در آنها متن سؤال و گزینه‌ها منفی هستند یعنی منفی مضاعف احتراز کنید.
- و) سؤال‌ها مستقل از یکدیگر باشد تا اگر کسی پاسخ خاصی را نمی‌داند سوال دیگری دنباله آن نباشد.
- ز) دو گزینه متضاد را که یکی از آنها پاسخ درست است به کار نبرید زیرا احتمال حدس زدن صحیح را افزایش می‌دهد.
- ح) محل گزینه صحیح را به صورت تصادفی انتخاب کنید.
- ط) سؤال‌ها را طوری بنویسید که در حد درک و فهم آزمون‌شوندگان باشد.
- ی) تنها در صورت لزوم از سؤالات چندگزینه‌ای استفاده کنید.
- ک) در آزمون‌های تکوینی گزینه نمی‌دانم را به کار ببرید زیرا می‌تواند در تشخیص مشکلات یادگیری دانش‌آموزان موثر باشد.

نکات کلیدی و مهم :

- امتیاز مهم هدف‌های رفتاری این است که به سهولت قابل شناسایی و اندازه‌گیری هستند.
- سؤال‌های آزمون‌های تشریحی گسترده پاسخ برای سنجش هدف‌های بالاترین سطح حوزه شناختی مناسب‌ترین سؤال‌ها هستند.
- آزمون کوتاه پاسخ حدوسط آزمون‌های تشریحی و عینی است.
- آزمون‌های صحیح - غلط ساده‌ترین صورت سؤال‌های عینی هستند.
- آزمون‌های صحیح - غلط متداولترین نوع سوال‌های عینی هستند.
- یکی از تدابیر متداول و موثر معلمان در ارزشیابی‌های غیررسمی از یادگیری دانش‌آموزان، به منظور دادن بازخورد به آنان، تشویق و ترغیب آنان در یادگیری و رفع مشکلات و معایب یادگیری آنان استفاده از پرسش‌های شفاهی است.
- گی در مورد دو روش کلی یا سراسری و روش تحلیلی بیان می‌دارد که هرچند هم برای آزمون‌های وابسته به ملاک و وابسته به هنجار می‌توان از روش کلی و روش تحلیلی استفاده کرد اما روش تحلیلی برای آزمون‌های وابسته به ملاک مناسب‌تر است و روش کلی یا سراسری برای آزمون‌های وابسته به هنجار مناسب‌تر است.
- ویرسما و جورز بهترین روش نمره‌گذاری را ترکیبی از دو روش فوق می‌دانند.

روش‌های نمره‌گذاری آزمون‌های تشریحی

- روش تحلیلی
- روش کلی
- روش ویژگی‌های اصلی
- آزمون‌های تشریحی یا انشایی معمول‌ترین نوع آزمون‌های کلاسی است
- نیتکو در اینباره اشاره می‌کند که تمایل معلمان بر این است که سوال‌های صحیح را با طول و تفصیل بیشتری بنویسند و این نکته را اغلب پاسخ‌دهندگان کشف می‌کنند.
- آزمون‌های عینی آزمون‌هایی هستند که در آنها سوال‌ها و همین‌طور جواب سوال در اختیار آزمون‌شوندگان قرار می‌گیرند.
- آزمون‌های صحیح غلط از زمان سقراط مورد استفاده پرورشکاران بوده است.
- حساسیت آزمون یعنی اینکه تفاوت‌های جزئی را نشان دهد.
- معمولاً در تصحیح آزمون‌های انشایی، ورقه‌های اول نمره کمتری دریافت می‌کنند.

- نمره‌گذاری آزمونهای تشریحی یا انشایی به صورت سوال به سوال باعث می‌شود که دقت و ثبات نمره‌گذاری افزایش یابد.
- در تصحیح اوراق دانش‌آموزان اثر پلیگاسیون عبارت است از اثر انتظارات معلم بر موفقیت دانش‌آموزان.
- روش تحلیل نمره‌گذاری در آزمونهای تشریحی نسبت به سایر روشها مناسبترین و بهترین روش است.
- از موارد استفاده پرسش شفاهی کاربرد آن با دانش‌آموزان معلول جسمی است.
- بزرگترین ضعف امتحانات شفاهی، ذهنی بودن و بی‌اعتباری آن است.
- آزمونهای کوتاه پاسخ برای اندازه‌گیری هدفهای ناختمی سطح پایین به ویژه دانش یا معلومات مفید است.
- نقش گزینه‌های انحرافی، منحرف کردن از شون‌گانی استکه پاسخ درست را نمی‌دانند.
- تجانس گزینه‌ها در سوالات چندگزینه‌ای یکی از راههای کنترل سطح دشواری سوال است.
- آزمون مهمترین وسیله اندازه‌گیری پیشرفت تحصیلی است.

فصل هشتم:

روش‌های سنجش فرایندها و فرآورده‌های غیرشناختی

روش‌های سنجش فرایندها و فرآورده‌های غیرشناختی

برای سنجش اهداف حوزه شناختی آزمون‌های کتبی یا مداد کاغذی شهرت و کاربرد زیادی دارد ولی برای سنجش اهداف حوزه عاطفی و روانی حرکتی این آزمون‌ها مناسب نیستند.

برای سنجش اهداف حوزه روانی حرکتی از دو دسته آزمون استفاده می‌گردد.

۱. آزمون‌های عملکردی

۲. روش‌های مشاهده‌ای

الف) آزمون‌های عملکردی

آزمون‌های عملکردی که به آنها آزمون‌های واقعی یا اصیل نیز گفته می‌شود فرایندها و فرآورده‌های یادگیری دانش‌آموزان را به طور مستقیم می‌سنجد. این آزمون‌ها با «مهارت» سر و کار دارند مهارت در استفاده از فرایندها و شیوه‌های اجرایی و نیز مهارت در تولید فرآورده‌ها. آزمون‌های عملکردی شامل آزمون کتبی، عملکردی، آزمون

شناسایی، انجام عملکرد در موقعیتهای شبیه‌سازی‌شده و نمونه کار می‌باشد که آزمون کتبی عملکردی بیشترین فاصله را با عملکرد در زندگی واقعی و نمونه کار کمترین فاصله را با عملکرد واقعی دارد.

فرایند در مقایسه با فرآورده

آزمون‌های عملکردی با فرایند یا شیوه اجرا، فرآورده، و یا ترکیبی از این دو سر و کار دارد. ماهیت عملکرد مورد سنجش مشخص می‌کند که تأکید بر چه قسمتی باید باشد. بعضی از عملکردها به محصولات یا فرآورده‌های ملموس و عینی نمی‌انجامد، مانند کار با وسایل آزمایشگاهی یا سخنرانی کردن. اینگونه فعالیت‌ها را باید ضمن انجام دادن سنجش کنیم و لذا با شیوه اجرا سر و کار داریم. در بعضی زمینه‌های عملکردی دیگر، محصول یا فرآورده مورد تأکید است و توجه چندانی به فرایند یا شیوه اجرا نمی‌شود. در ارزشیابی از یک مقاله تحقیقی بیشتر فرآورده مورد توجه و ارزشیابی قرار می‌گیرد.

فرآورده چیزی است که یاد گیرنده تولید می‌کند و فرایند جریانی است که او مورد استفاده قرار می‌دهد تا به فرآورده دست یابد.

به آزمون‌های عملکردی آزمون‌های واقعی یا اصیل نیز گفته می‌شود. صاحب‌نظران بر اساس پژوهش‌های خود ویژگی‌های سنجش عملکرد را در چهار دسته خلاصه می‌کنند:

۱. تأکید بر کار بست ؛
۲. تأکید بر سنجش مستقیم ؛
۳. تأکید بر استفاده از وسایل واقعی ؛
۴. تأکید بر ترغیب و تشویق تفکر باز.

انواع آزمون‌های عملکردی

آزمون‌های عملکردی یا آزمون‌های واقعی یا اصیل دارای انواع بسیار متفاوت و متنوعی هستند که به چهار نوع از آنها اشاره می‌کنیم.

۱. آزمون کتبی عملکردی
۲. آزمون شناسایی
۳. انجام عملکرد در موقعیت‌های شبیه‌سازی شده
۴. نمونه کار

الف) آزمون کتبی عملکردی

از آزمون‌های کتبی عملکردی در سنجش عملکرد استفاده می‌شود. تفاوت عمده بین آزمون کتبی عملکردی و آزمون‌های کتبی در این است که در آزمون کتبی عملکردی عمدتاً بر کاربست دانش و مهارت در موقعیت‌های عملی یا شبیه‌سازی شده با موقعیت‌های عملی تأکید می‌شود. در این گونه آزمون‌های عملکردی، یا بازده‌های پایانی یادگیری سنجش می‌شوند یا مراحل میانی عملکردی که برای رسیدن به بازده‌های مطلوب پایانی ضروری هستند مانند استفاده درست از ابزارها و دستگاهها.

ب) آزمون شناسایی

منظور از آزمون شناسایی روشی است برای سنجش توانایی یادگیرنده در تشخیص ویژگی‌ها، محاسن، معایب و موارد استفاده امور مختلف. در آزمون‌های شناسایی از آزمون‌شونده خواسته می‌شود تا یک ابزار را شناسایی کند یا کار آن را توصیف کند. مثلاً شناسایی اجزا موتور اتومبیل.

ج) انجام عملکرد در موقعیت‌های شبیه‌سازی شده

در آزمون عملکرد در موقعیت‌های شبیه‌سازی شده از یادگیرنده خواسته می‌شود تا در یک موقعیت شبیه‌سازی یا خیالی همان اعمالی را انجام دهد که در موقعیت‌های واقعی ضروری هستند. برای نمونه، در تربیت بدنی، ضربه زدن به یک توپ خیالی.

د) نمونه کار

در روش نمونه کار از یادگیرنده خواسته می‌شود تا اعمالی را انجام دهد که معرف عملکرد واقعی مورد سنجش هستند. این روش نزدیکترین روش سنجش به عملکرد واقعی یادگیرنده در محیط‌های طبیعی است.

مراحل تهیه آزمون‌های عملکرد

۱. مشخص کردن بازده‌های موردنظر عملکرد
۲. انتخاب واقع‌بینانه موقعیت آزمون
۳. استفاده از راهنمایی‌ها و دستورالعمل‌ها
۴. استفاده از روش‌ها و فنون گوناگون مشاهده

الف) مشخص کردن بازده‌های موردنظر عملکرد

ملاک یا معیار عملکرد را در هدفهای عملکردی می‌توان با توجه به یکی از موارد دقت عملکرد، سرعت عملکرد، توالی درست مراحل، مهارت عملکرد و با توجه به رعایت ایمنی مشخص می‌کرد.

ب) انتخاب واقع‌بینانه موقعیت آزمون

ج) از راهنمایی‌ها و دستورالعمل‌هایی که به روشنی موقعیت آزمون را مشخص می‌کنند، استفاده نماید.

د) از روش‌ها و فنون گوناگون مشاهده استفاده نماید.

کار پوشه (پورت فولیو)

نوع دیگری از آزمون‌های عملکردی که به تازگی شهرت زیادی کسب کرده است کارپوشه نام دارد. کارپوشه نوعاً مجموعه‌ای از کارهای یک دانش‌آموز است که کوشش، پیشرفت، و موفقیت تحصیلی او را در یک زمینه خاص نشان می‌دهد. از کار پوشه به قصد برآوردن دو هدف عمده استفاده می‌شود: (۱) ارائه بهترین کارها و (۲) نشان دادن رشد تحصیلی یادگیرنده.

ب) روش‌های سنجش مشاهده‌ای

روش دیگر برای سنجش حیطه روانی - حرکتی، روش‌های سنجش مشاهده‌ای می‌باشد و شامل موارد زیر است:

۱. فهرست واری

۲. مقیاس درجه‌بندی

۳. واقعه‌نگاری

الف) روش فهرست واری

روش فهرست واری یک رویه نظام‌دار برای گزارش قضاوت‌های مشاهده‌گر یا مشاهده‌گران است با این روش می‌توان معلوم کرد که در عملکرد شخص مورد سنجش ویژگی‌های مورد نظر موجودند یا نه. البته در فهرست واری تنها بودن یا نبودن یک ویژگی مورد نظر ثبت می‌شود.

از روش فهرست واری هم می‌توان برای سنجش فرایند عملکرد استفاده کرد و هم برای فرآورده آن. هنگام تهیه فهرست واری به منظور ارزشیابی از فرایند عملکرد تأکید باید بر رفتارهای فرد باشد. هنگام تهیه فهرست واری برای ارزشیابی از محصول یا فرآورده رفتار بر جنبه‌ها یا ویژگی‌های قابل مشاهده محصول مورد سنجش تأکید می‌شود.

ب) روش مقیاس درجه‌بندی

این روش شبیه به فهرست واری است اما به عوض تعیین بود یا نبود رفتار یا ویژگی مورد ارزشیابی درجه‌بندی یا مقدار رفتار یا ویژگی تعیین می‌گردد.

بزرگترین مشکل مقیاس درجه‌بندی، کاربرد اصطلاحات و صفات کلی و مبهم در آنها است به سخن دیگر در مقیاس درجه‌بندی کیفیت مورد ارزشیابی قرار می‌گیرد. مقیاس درجه‌بندی انواع مختلفی دارد که عبارتند از:

۱- مقیاس درجه‌بندی عددی

مقیاس درجه‌بندی عددی که در آن ارزشیاب با علامت‌گذاری بر روی یک عدد میزان موجود بودن رفتار مورد ارزشیابی را مشخص می‌کند.

۲- مقیاس درجه‌بندی نگاره‌ای

در این مقیاس درجه‌بندی، به جای استفاده از اعداد از خطوط افقی استفاده می‌شود و ارزشیاب داوری خود را درباره ویژگی مورد مشاهده با گذاشتن علامت بر روی خط مشخص می‌کند تا چه اندازه در بحث شرکت می‌کند؟

۳- مقیاس درجه‌بندی نگاره‌ای توصیفی

در این مقیاس نقاط روی مقیاس توصیف می‌شوند و این توصیف‌ها به صورت عبارت‌های رفتاری که دانش‌آموز انجام می‌دهد، بیان می‌شوند.

عوامل مؤثر بر بهبود روش مقیاس درجه‌بندی

- ۱- سعی کنید ویژگی‌هایی را که از لحاظ آموزشی مهم هستند برگزینید.
- ۲- ویژگی‌های مورد ارزشیابی را به گونه‌ای بیان کنید که مستقیماً قابل مشاهده باشند.
- ۳- هم ویژگی‌های مورد ارزشیابی و هم نقاط روی مقیاس را به روشنی تعریف کنید.
- ۴- شماره درجات مقیاس را به گونه‌ای بین سه تا هفت برگزینید و به ارزشیابان اجازه دهید تا در بین این نقاط نیز درجه‌بندی انجام دهند.
- ۵- از ارزشیابان بخواهید هر زمان به سوالی برمی‌خورند که قادر به قضاوت درباره ویژگی مطرح شده در آن نیستند به آن سوال جواب ندهند.
- ۶- در صورت امکان از چند نفر بخواهید تا رفتارها یا محصولات مورد نظر را مشاهده و ارزشیابی کنند.

روش واقع‌نگاری

روش واقع‌نگاری یا روش ثبت رویداد به توصیف‌های واقعی از رویدادها و اتفاقات معنی‌داری که معلم در نتیجه مشاهده از زندگی دانش‌آموزان به دست می‌آورد گفته می‌شود. در استفاده از روش واقع‌نگاری هیچ نوع رفتار از پیش تعیین‌شده‌ای سنجش نمی‌شود. بنابراین، این روش بیشتر جنبه تحلیلی دارد تا ارزشیابی. روش واقع‌نگاری یک روش مشاهده مستقیم رفتار است و به همین دلیل برای بیشتر هدف‌های آموزشی حوزه‌های عاطفی و روانی - حرکتی قابل استفاده است.

عوامل موثر در استفاده بهینه از روش واقع‌نگاری

- ۱- از پیش تعیین کنید که چه چیزی را مورد مشاهده قرار می‌دهید، اما آماده ثبت رفتارهای غیرمعمول نیز باشید.
- ۲- آن مقدار از موقعیت را مشاهده کنید که رفتار را برای شما معنی‌دار سازد.
- ۳- بلافاصله پس از مشاهده واقعه یا رویداد آن را ثبت کنید.
- ۴- هر یک مورد ثبت واقعه را به توصیف مختصری از یک رویداد محدود نمایید.
- ۵- توصیف واقعیتهای و تفسیر خودتان از آنها را از هم جدا کنید.
- ۶- هم موارد مثبت و هم موارد منفی رفتار را ثبت کنید.
- ۷- پیش از قضاوت و نتیجه‌گیری درباره رفتار یک دانش‌آموز، تعداد کافی گزارش واقع‌نگاری تهیه نمایید.
- ۸- در نوشتن گزارش واقع‌نگاری تمرین کنید.

سنجش اهداف عاطفی

وسایل سنجش اهداف عاطفی از آزمون‌شونده می‌خواهند هر آنچه را احساس می‌کند بگویند. بنابراین مشکل اساسی روش‌های سنجش اهداف عاطفی این است که باید موقعیت‌هایی ایجاد گردد که در آن آزمون‌شونده رفتار عاطفی مورد علاقه خود را بروز دهد.

معروف‌ترین و مهم‌ترین فنون سنجش اهداف حیطه عاطفی عبارتند از:

- ۱) پرسشنامه‌های بسته پاسخ یا گزینه بایست: بسیار شبیه به آزمون‌های عینی است زیرا در هر دو آزمون‌شونده باید پاسخی را از میان پاسخ‌های تعیین شده برگزیند. با این سوالات می‌توان اهداف مختلف حوزه عاطفی را سنجید. مثلاً دوست دارید در یک انجمن کتاب عضو شوید؟

خیلی زیاد □ زیاد □ متوسط □ کم □ خیلی کم □

- ۲) پرسشنامه‌های باز پاسخ: شبیه به آزمون‌های تشریحی در حوزه شناختی است در هر دو آزمون‌شونده سؤالاتی را دریافت می‌کند و سپس در پاسخ نظر خود را توضیح می‌دهد مثال اگر آخر این هفته سه روز تعطیل بود.....
- ۳) روش تفکیک معنایی: در این روش تعدادی صفت دو قطبی (زشت زیبا) (ضعیف و قوی) بر روی یک پیوستار به صورت جفت بیان می‌گردد و پاسخ‌دهنده با گذاشتن علامت نگرش خود را بیان می‌کند. مثلاً از دانش‌آموزان درباره معلم پرسیده شده است.

- ۴) روش یادداشت روزانه: دانش‌آموزان می‌توانند احساس‌های شخصی خود را درباره تجارب شخصی و فعالیت‌های کلاسی به صورت یادداشت روزانه ابراز نمایند. در این رابطه معلمان باید به شاگردان اطمینان دهند که یادداشت‌های خصوصی آنان را با کسی درمیان نخواهند گذاشت.

نکات کلیدی و مهم :

- آزمونهای عملکردی فرآیند و فرآورده‌ی یادگیری را به صورت مستقیم سنجش می‌کند.
- آزمونهای عملکردی با مهارت سر و کار دارند.
- آزمون کتبی عملکردی بیشترین فاصله را با عملکرد در زندگی واقعی دارد.
- در آزمونهای عملکردی تهیه و اجرای آن دقت بیشتری لازم است و نمره‌گذاری آن دشوار است.
- کارپوشه: مجموعه‌ای از کارهای یک دانش‌آموز است که کوشش، پیشرفت و موفقیت تحصیلی او را در یک زمینه خاص نشان می‌دهد.
- مطالب مربوط به کارپوشه جنبه تراکمی دارد.
- هدف عمده کارپوشه: - ارائه بهترین کارها - نشان دادن رشد تحصیلی
- نمره‌گذاری روش کارپوشه هم از روش تصحیحی تحلیلی هم روش کلی می‌توان استفاده کرد.
- در روش کارپوشه معلم ضمن ارزشیابی به آموزش نیز اقدام می‌کند.

فصل نهم:

روش‌های سنجش مستقیم رفتار

روش‌های سنجش مستقیم رفتار

روانشناسان دارای رویکرد رفتاری عمدتاً از روش‌های سنجش مستقیم استفاده می‌کنند. آنها به طور مستقیم رفتارهای آشکار افراد را مشاهده و ثبت و ضبط می‌نمایند.

بعدهای رفتار

در سنجش مستقیم رفتار، ویژگی‌ها یا بعدهای رفتار مشاهده و ثبت می‌شوند. بعدهای مختلف رفتار عبارتند از:

- شکل ظاهری رفتار - یعنی اینکه رفتار در چه شکل یا صورتی رخ می‌دهد.
- شدت یا نیروی رفتار - یعنی اینکه رفتار با چه شدتی رخ می‌دهد.
- نرخ یا فراوانی رفتار - یعنی تعداد دفعات وقوع رفتار.
- طول مدت رفتار - یعنی زمانی که انجام دادن یک کار یا یک پاسخ به طول می‌انجامد.
- دوره نهفتگی رفتار - یعنی مدتی که به طول می‌انجامد تا شخص انجام کار یا رفتاری را که از او خواسته شده است شروع کند.
- مکان یا موقعیت انجام رفتار - یعنی جایی که رفتار در آن رخ می‌دهد.

انواع روش‌های سنجش مستقیم رفتار

۱- سنجش محصولات ماندنی رفتار یا سنجش فرآورده رفتار

۲- سنجش رویدادهای گذاری رفتار یا سنجش فرآیند رفتار

الف) سنجش محصولات ماندنی رفتار یا سنجش فرآورده رفتار

بعضی رفتارها از خود شواهد عینی شامل محصولات یا تولیدات ماندنی به جای می‌گذارند. مثلاً وقتی که دانش‌آموزان درس ریاضی مسائل آن درس را حل می‌کنند و جواب‌های خود را بر روی برگه امتحانی می‌نویسند نوعی محصول ماندنی یا پایدار تولید می‌کنند. محصولات ماندنی رفتار بازده‌های یادگیری هستند، از این رو به سنجش محصولات ماندنی رفتار، ثبت بازده نیز گفته می‌شود. روش سنجش محصولات ماندنی رفتار یک روش بعد از واقعه است، یعنی اندازه‌گیری زمانی صورت می‌پذیرد که رفتار انجام شده و به پایان رسیده باشد.

روش سنجش محصولات ماندنی رفتار برای تعیین ویژگی‌ها یا بعدهای رفتاری نرخ رفتار، شکل ظاهری رفتار و شدت رفتار مفید است.

ب) سنجش رویدادهای گذاری رفتار یا سنجش فرآیند رفتار

بسیاری از رفتارهایی که روان‌شناسان یا معلمان با آنها سروکار دارند چنین آثاری از خود به جای می‌گذارند. به همین سبب، سنجش این گونه رفتارها بسیار دشوارتر است. آرام نشستن، تنبلی کردن، لبخند زدن، شرکت در بحث گروهی و از این قبیل، رفتارهای گذرا هستند و آثاری از خود به جای نمی‌گذارند که بعداً بتوان آنها را اندازه‌گیری کرد.

برای اندازه‌گیری این نوع رفتارها، لازم است هنگامی که دارند اتفاق می‌افتند به مشاهده آنها پردازیم. به همین جهت، روش‌های سنجش رویدادهای گذرا را روش‌های ثبت مشاهده‌ای نیز می‌نامند. روش‌های سنجش رفتارها گوناگون‌اند که مهمترین آنها از این قرارند: ثبت رویداد یا فراوانی، ثبت طول مدت، ثبت دوره نهفتگی، ثبت فاصله‌ای و ثبت نمونه‌گیری زمان.

ثبت فراوانی یا ثبت رویداد رفتار

اگر دفعات وقوع یک رفتار یا یک پاسخ را در یک فاصله زمانی معین شمارش و ثبت کنیم از روش ثبت رویداد استفاده کرده‌ایم. روش ثبت فراوانی یا نرخ رفتار متداول ترین روش ثبت مشاهده‌ای است. فراوانی رفتار را با استفاده از تدابیر مختلفی همچون استفاده از مداد و کاغذ، گچ و تابلوی کلاس، شمارشگر مچی، چرتکه مچی و حسابگر الکترونیکی می‌توان اندازه گرفت.

ثبت طول مدت رفتار

در برخی از رفتارها ثبت طول مدت بیشتر از فراوانی اهمیت دارد. به عنوان مثال رفتارهایی چون زمان توجه کردن کودک، ترک کردن صندلی، قشقرق راه انداختن، و مانند اینها. برای سنجش اینگونه رفتارها از روش ثبت طول مدت رفتار استفاده می‌شود. برای ثبت طول مدت وقوع یک رفتار می‌توان از وسایل مختلفی مانند ساعت دیواری، ساعت مچی یا کرنومتر استفاده کرد.

ثبت دوره نهفتگی یا درنگ رفتار

روش ثبت دوره نهفتگی تعیین می‌کند که چه زمانی طول می‌کشد تا شخص رفتاری را آغاز کند. از این روش برای سنجش بعد نهفتگی رفتار استفاده می‌شود.

ثبت فاصله زمان یا روش ثبت فاصله‌ای

بهترین روش سنجش یک رفتار آن است که تمامی موارد آن را به طور مداوم مورد مشاهده قرار دهیم و جزئیات وقوع آن را ثبت نماییم. به روش ثبت تمامی موارد یک رفتار در یک مدت زمان مشخص روش ثبت مداوم می‌گویند. ولی این روش به دلیل وقت‌گیر بودن و پرهزینه بودن چندان مورد استفاده قرار نمی‌گیرد و به جای آن معمولاً از روش ثبت فاصله زمان یا روش ثبت فاصله‌ای استفاده می‌شود. از این روش برای رفتارهایی استفاده می‌شود که آغاز و انجام قابلیت تمایزی ندارند. مثلاً فرض کنید کودکی سر و صداهای مزاحمی از خود تولید کند، مانند داد زدن که شمردن تعداد دفعات این گونه رفتارها یا اندازه‌گیری طول مدت وقوع آنها کار دشواری است.

ثبت نمونه‌گیری زمان

در روش ثبت نمونه‌گیری زمان، وقوع یک رفتار در فاصله‌های کوتاه مشاهده و ثبت می‌شود. معمولاً از روش ثبت فاصله‌ای و روش نمونه‌گیری زمان با هم استفاده می‌شود، زیرا امکان دارد که در ضمن مشاهده و ثبت رفتار با روش ثبت فاصله‌ای قسمتی از رفتار مورد مشاهده از نظر بیفتد. همچنین از ترکیب این دو روش زمانی استفاده می‌شود که مشاهده‌گر بخواهد رفتار دانش‌آموز را همزمان ثبت کند.

روش‌های ثبت فاصله‌ای و نمونه‌گیری زمان برای سنجش بعد فراوانی رفتار مفیدند، اما نمی‌توان از آنها برای اندازه‌گیری نرخ رفتار استفاده کرد.

نکات کلیدی و مهم

- روشهای سنجش مستقیم را ابتدا به دو دسته کلی سنجش محصولات ماندنی رفتار و سنجش رویدادهای گذرای رفتار تقسیم کرده‌اند.
- روش ثبت نمونه گیری زمان برای ثبت رفتارهایی که با فراوانی متوسط رخ می‌دهند مفید است.
- روش ثبت فاصله‌ای برای سنجش «بود و نبود» رفتار مناسب است.
- روش ثبت دوره نهفتگی تعیین می‌کند که چه مقدار زمان طول می‌کشد تا شخص رفتاری را آغاز کند.
- معمولا از روش ثبت فاصله‌ای و روش نمونه گیری زمان با هم استفاده می‌شود.
- به سنجش محصولات ماندنی رفتار ثبت بازه گویند.

فصل دهم:

اجرا، نمره‌گذاری و تحلیلی آزمون

شرایط اجرای آزمون‌های کتبی

گام نخست در اجرای آزمون‌های پیشرفت تحصیلی آماده‌سازی و فراهم آوردن شرایط مناسب برای اجرای آنهاست. تهیه دفترچه آزمون و پاسخنامه، ترتیب پشت‌سرهم قرار دادن سؤال‌های مختلف، و در نظر گرفتن زمان و مکان مناسب برای اجرای آزمون نیاز به دقت فراوان دارد.

ترتیب قرار دادن سؤال‌های آزمون بدنبال یکدیگر

رعایت کردن اصول مربوط به ترتیب قرار دادن سؤال‌های آزمون بدنبال یکدیگر موجب می‌شود که آزمون شونده نظم فکری داشته باشد و راحت‌تر به سؤالات پاسخ دهد. گراندولین در رابطه با قرار دادن سؤال‌های یک آزمون به دنبال یکدیگر چند پیشنهاد زیر را ارائه داده است:

(۱) سؤال‌ها را با توجه به نوع آنها دسته‌بندی کنید و پشت‌سرهم قرار دهید.

در این مورد معمولاً ترتیب زیر پیشنهاد می شود:

۱. سؤالات صحیح - غلط
۲. سؤالات جورکردنی
۳. سؤالات چندگزینه‌ای
۴. سؤالات کوتاه پاسخ
۵. سؤالات تشریحی

(۲) سؤال‌ها را از ساده به دشوار مرتب کنید.

مرتب کردن سؤال‌ها را از ساده به دشوار موجب می شود دانش آموز از سؤالات آسان شروع به پاسخگویی کند و این کار موجب اعتماد به نفس در وی می شود. در این روش هم ابتدا سؤالات برحسب نوع دسته بندی می شود و سپس از آسان به مشکل چیده می شوند.

(۳) سؤال‌ها را به ترتیب سازمان اصلی مطالب به دنبال هم مرتب کنید.

مقصود از مرتب کردن سؤال‌ها به ترتیب سازمان اصلی مطالب به دنبال هم این است که ترتیب سؤالها هماهنگ با ترتیب مطالب اصلی درس باشد. رعایت این نکته موجب نظم فکری می شود. آزمون‌هایی که سؤالات آنها درهم و برهم و خوانا نیستند موجب دلسردی آزمون‌شونده می شوند.

نوشتن راهنما و دستورالعمل

برای هر آزمون نوشتن یک راهنما و دستورالعمل لازم است راهنمای آزمون باید مختصر و مفید باشد و به طور آشکار به آزمون‌شونده بگوید که چه کاری را انجام دهد. مهرنر و ولهمان گفته‌اند که راهنمای آزمون باید اطلاعات زیر را در اختیار آزمون‌شوندگان قرار دهد:

(۱) زمان لازم برای هر قسمت

(۲) ارزش هر سؤال

(۳) مجاز یا غیرمجاز بودن حدس زدن

در نوشتن راهنما برای آزمون باید علاوه بر راهنمای کلی برای هریک از آزمون هم یک راهنمای جداگانه نوشته شود. در راهنمای آزمون‌های عینی نحوه نمره گذاری پاسخها را در رابطه با حدس زدن و یا حدس نزدن جوابها مشخص کنید. در این رابطه مهرنر و ولهمان بر این عقیده‌اند در تصحیح پاسخها نباید هیچ گونه نمره‌ای از دانش‌آموزان برای حدس زدن کسر شود.

اجرای آزمون

شرایط اجرای آزمون‌های پیشرفت تحصیلی باید به گونه‌ای باشد که از هر لحاظ آسایش جسمی و فکری آزمون‌شوندگان را تأمین کند. پژوهش‌های انجام شده نشان داده‌اند که اندکی اضطراب سبب افزایش دقت و کوشش آزمون‌شوندگان می‌شود اما اضطراب زیاد بر عملکرد آنان تأثیر منفی دارد.

نمره‌گذاری

پس از تهیه و اجرای آزمون‌های پیشرفت تحصیلی نوبت به تصحیح برگه‌های آزمون و نمره‌گذاری آنها می‌رسد در نمره‌گذاری آزمون‌های عینی دو روش عمده وجود دارد:

۱- احتساب کلی پاسخ‌های درست بدون کسر نمره برای حدس زدن

۲- کسر مقداری از نمره برای جبران حدس زدن

الف- احتساب کلی پاسخ‌های درست بدون کسر نمره برای حدس زدن

روش اول یعنی احتساب کلی پاسخ‌های درست بدون کسر نمره برای حدس زدن یک روش متداول و ساده است ولی اگر معلم بخواهد روش دوم را مورد استفاده قرار دهد باید دانش‌آموزان را به این مسأله آگاه کند که در ازای پاسخ‌های غلطی که می‌دهند مقداری از نمره آنها کم خواهد شد. برای کسر کردن نمره چند روش وجود دارد که متداول‌ترین و معروف‌ترین آنها استفاده از فرمول زیر است:

$$R = \frac{W}{N-1} = \text{نمره اصلاح شده}$$

R= تعداد پاسخ‌های درست

W= تعداد پاسخ‌های غلط

N= تعداد گزینه‌های هر سؤال

تحلیل سؤالات آزمون

هدف از تحلیل سؤالات آزمون، واریسی یکی یکی سؤال‌ها و تعیین میزان دقت و نارسایی‌های آنهاست در تحلیل سؤال‌های آزمون نقاط قوت و ضعف یک آزمون و کیفیت تک تک سؤال‌های آن تعیین می‌شود.

مراحل تحلیل سؤال

اطلاعات لازم مورد نیاز برای تحلیل سؤال‌های یک آزمون پاسخ‌هایی هستند که آزمون‌شوندگان به هر سؤال داده‌اند بنابراین باید تعیین شود که در هر سؤال چند نفر گزینه‌های درست را انتخاب کرده‌اند، هر یک از گزینه‌های انحرافی چند نفر را به خود جلب کرده است، و چند نفر آن را بی‌جواب گذاشته‌اند. برای این منظور اطلاعات مخصوص هر سؤال روی یک کارت آورده می‌شود و با مشخص کردن گروه بالا (کسانی که خوب جواب داده‌اند) و گروه پایین (کسانی که ضعیف عمل کرده‌اند) به بررسی می‌پردازیم. برای تعیین گروه‌های بالا و پایین برگه‌های آزمون را به ترتیب نمره‌ای که گرفته‌اند از کوچک به بزرگ مرتب کرده سپس از بالاترین شروع می‌کنیم و تعداد برگه‌هایی را که می‌خواهیم در گروه بالا قرار دهیم انتخاب می‌کنیم و به همین ترتیب برای انتخاب گروه پایین از پایین‌ترین نمره شروع می‌کنیم. متخصصان آزمون‌سازی پیشنهاد کرده‌اند که اگر تعداد دانش‌آموزان در آزمون بیشتر از ۲۰ و یا ۴۰ نفر هستند ۱۰ برگه بالا و ۱۰ برگه پایین انتخاب کنید. اگر تعداد دانش‌آموزان ۲۰ نفر یا کمتر است برگه‌های آزمون را کلاً به دو دسته تقسیم کنید و در شرایطی که تعداد آزمون‌شوندگان بیش از ۴۰ نفر باشد، بهترین رقم برای گروه بالا و پایین ۲۷٪ کل برگه‌هاست. این رقم برای شرایطی که نمرات دارای توزیع بهنجار باشند پیشنهاد شده است. اما برای آزمون‌های کلاسی هر یک از ارقام ۲۵ تا ۳۳ درصد به عنوان تعداد برگه‌های بالا و پایین مناسب است.

محاسبه ضریب دشواری سؤال

بنابر تعریف درصد کل آزمودنی‌هایی که به یک سؤال جواب مثبت می‌دهند، ضریب دشواری آن سؤال است که با حرف P نشان داده می‌شود. اگر در تحلیل سؤال کلیه برگه‌ها دخالت داشته باشند. برای محاسبه ضریب دشواری می‌توان تعداد کل افرادی که جواب درست داده‌اند (R) تقسیم بر تعداد کل آزمون‌شوندگان (T) کرد و نتیجه را در ۱۰۰ ضرب کرد. مقدار بدست آمده ضریب دشواری آن سؤال خاص می‌باشد.

مطابق فرمول زیر:

$$P = 100 \left(\frac{R}{T} \right)$$

نکته: هر اندازه ضریب دشواری یک سؤال بزرگتر (به ۱۰۰ نزدیکتر) باشد آن سؤال آسان‌تر است و هر اندازه این ضریب کوچکتر باشد (به صفر نزدیکتر) سؤال دشوارتر است بنابراین به جای ضریب دشواری می‌توان از ضریب آسانی یا سهولت نام برد اما معمول این است که به آن ضریب دشواری بگویند.

تفسیر ضریب دشواری

در آزمون‌های وابسته به هنجار هرچه واریانس نمرات حاصل از یک آزمون وابسته به هنجار بزرگتر باشد آن آزمون بهتری است بنابراین یکی از راه‌های قضاوت درباره مفید بودن سؤالات یک آزمون آن است که ببینیم سؤال تا چه اندازه به پراکندگی یا واریانس نمرات کمک می‌کند.

برای محاسبه واریانس یک سؤال از فرمول زیر استفاده می‌شود:

$$P \times (1 - P) = \text{واریانس سؤال}$$

در این فرمول، P برابر است با درصد کسانی که به سؤال جواب درست دادند یعنی همان ضریب دشواری است. یک سؤال زمانی دارای حداقل واریانس است که ضریب دشواری آن ۰ و یا ۱ باشد. اما زمانی که ضریب دشواری سؤال $P = 0.5$ باشد واریانس سؤال حداکثر خواهد بود. در نتیجه از لحاظ انتخاب برای گنجاندن در فرم نهایی آزمون سؤال‌هایی بهتر هستند که ضریب دشواری آنها از ۱ کمتر و از صفر بیشتر و به ۰/۵ نزدیک باشد.

بطور کلی ضریب دشواری بین ۰/۳ تا ۰/۷ حداکثر اطلاع را درباره تفاوت بین آزمودنی‌ها به دست می‌دهد.

محاسبه ضریب تمیز سؤال

برخلاف ضریب دشواری که میزان آسان بودن یا دشوار بودن یک سؤال را برای گروه آزمون‌شوندگان نشان می‌دهد ضریب تمیز را که با D نشان می‌دهند قدرت سؤال را در تمایزگذاری یا تشخیص بین گروه قوی و ضعیف آزمون‌شوندگان مشخص می‌کند. یعنی معلوم می‌نماید که سؤال تا چه اندازه می‌تواند گروه قوی را از گروه ضعیف جدا سازد. برای محاسبه ضریب تمیز از فرمول زیر استفاده می‌شود:

ضریب تمیز مساوی است با انتخابهای درست گروه پایین - انتخابهای درست گروه بالا تقسیم بر تعداد افراد یک گروه (بالا یا پایین).

تفسیر ضریب تمیز

هرقدر ضریب تمیز بزرگتر باشد قوه تمیز آن سؤال بیشتر است و هرقدر این ضریب کوچکتر باشد، قوه تمیز آن کمتر خواهد بود. مثلاً اگر قوه تمیز سؤالی ۹۰٪ باشد آن سؤال آزمون‌شوندگان قوی و ضعیف را خیلی خوب از هم جدا خواهد کرد؛ اما اگر ضریب تمیز ۱۰٪ باشد آن سؤال از عهده جداسازی دانش‌آموزان قوی و ضعیف به خوبی برنخواهد آمد. ضریب تمیز صفر حاکی از این است که آن سؤال به هیچ وجه نتوانسته بین گروه قوی و ضعیف تمایز قائل شود.

گاه اتفاق می‌افتد که ضریب تمیز منفی می‌شود. این مسأله حاکی از آن است که در آن سؤال گروه قوی بدتر از گروه ضعیف عمل کرده است که این‌گونه سؤال‌ها دارای اشکالات اساسی هستند.

همبستگی سؤال با کل آزمون

برای آنکه پی ببریم که سؤالات مطرح شده دارای قوه تمیز بالا می‌باشد باید بین هر سؤال و کل آزمون همبستگی

بگیریم اگر همبستگی قوی و مثبت باشد، سؤال از قوه تمیز خوبی برخوردار است و اگر همبستگی صفر یا منفی باشد، سؤال مطرح شده قدرت تمیز خوبی ندارد یا قوه تمیز آن صفر است. لذا باید از ضریب همبستگی دو رشته‌ای نقطه‌ای استفاده گردد. هرچه همبستگی بیشتر باشد به این معنی است که با سؤال‌های دیگر همخوانی دارد و لذا بین افراد قوی و ضعیف تمیز قائل می‌شود و از امتیازهای استفاده از ضریب همبستگی بین سؤال و کل آزمون این است که می‌تواند این ضریب را از لحاظ معنی‌داری آماری آزمون کرد. دوم اینکه، از لحاظ عملی می‌توان معنی‌داری ضریب را مورد قضاوت قرار داد. برای مثال، اگر ضریب همبستگی بین یک سؤال و کل آزمون $0.4+$ باشد، می‌توانیم بگوییم که این سؤال ۱۶ درصد ($16 = 4^2$) واریانس نمرات را به خود اختصاص داده است.

همبستگی میان سؤالات آزمون

بهرتر است به هنگام تحلیل آزمون، علاوه بر ضریب همبستگی بین هر سؤال و کل آزمون، ضریب همبستگی تک تک سؤالات را با یکدیگر نیز تعیین کنیم. این کار چند فایده دارد: نخست اینکه، با در دست داشتن تعداد سؤال‌های آزمون و میانگین ضرایب همبستگی بین سؤال‌ها می‌توان ضریب پایایی آزمون را محاسبه کرد. دوم اینکه، با استفاده از این نوع همبستگی می‌توان به مفهوم قوه تمیز سؤال دست یافت. به سخن دیگر، با استفاده از همبستگی بین سؤالات یک آزمون می‌توان فهمید که چرا بعضی از سؤالات توانایی متمایز ساختن بین آزمون‌شوندگان قوی و ضعیف را ندارند.

اگر همبستگی بین یک سؤال و کل آزمون کم باشد، دو دلیل برای آن وجود دارد: یکی اینکه ممکن است سؤال مورد نظر با هیچ کدام از سؤالات آزمون همبسته نباشد. در این حالت یا باید سؤال را بازنویسی کرد یا آن را به کناری نهاد. امکان دوم این است که سؤال با بعضی از سؤالات آزمون همبستگی قابل ملاحظه نشان می‌دهد، اما با دیگر سؤال‌ها همبستگی نزدیک به صفر یا حتی منفی دارد.

رابطه بین ضریب دشواری و ضریب تمیز

برای آنکه بتوانیم بگوییم آزمونی دارای سؤال‌های خوب است باید ضریب دشواری متوسط و ضریب تمیز بالایی داشته باشد. (ضریب دشواری متوسط در حدود 0.5 و ضریب تمیز بالا حدود 1 است).

تحلیل گزینه‌های انحرافی

هدف از قرار دادن گزینه‌های انحرافی در سؤال‌ها منحرف کردن آزمون‌شوندگانی است که جواب درست سؤال را نمی‌دانند. در تحلیل گزینه‌های انحرافی قاعده کلی به شرح زیر است: هر گزینه انحرافی دست کم باید یک نفر از افراد گروه ضعیف را به خود جلب کند و اگر هر دو گروه قوی و ضعیف را به خود جلب می‌کند، تعداد افراد گروه

ضعیف باید بیشتر از تعداد افراد گروه قوی باشد. بعد از تحلیل سؤالات آزمون باید سؤالاتی که ضریب دشواری آنها بسیار بالا یا بسیار پایین است یا ضریب تمیز آنها خیلی کم است تجدید نظر گردد و گزینه‌های نامناسب اصلاح گردد. در صورتی یک سؤال به خوبی عمل می‌کند که افراد ضعیف بیشتر از افراد گروه قوی گزینه‌های انحرافی آن سؤال را انتخاب نمایند.

تحلیل سؤال‌های آزمون‌های تشریحی و آزمون‌های عملکردی

هرچند که تحلیل بیشتر برای آزمون‌های عینی متداول است، با این حال سؤال‌های غیر عینی را نیز پس از اجرا باید مورد تجزیه و تحلیل قرار داد و اشکالات آنها را برطرف کرد. دو روش معروف برای تحلیل آزمون‌های غیر عینی وجود دارد.

یکی از روش‌ها، روش وتینی و سیبرز است که به صورت زیر می‌باشد:

۱- در این روش ابتدا دو گروه ۲۵ درصدی بالا و پایین از کل دانش‌آموزان را به شرحی که قبلاً گذشت (۲۵٪)

دانش‌آموزانی که بیشترین نمره و ۲۵ درصد از دانش‌آموزانی که کمترین نمره را دارند) تعیین می‌کنیم.

۲- برای هر سؤال مجموع نمرات گروه بالا و پایین را حساب می‌کنیم.

۳- برای محاسبه ضریب‌های دشواری و تمیز نتایج را در فرمول‌های زیر قرار می‌دهیم:

در این فرمول‌ها:

$$\sum H = \text{مجموع نمره‌های افراد یا برگه‌های گروه بالا (۲۵ درصد)}$$

$$\sum L = \text{مجموع نمره‌های افراد یا برگه‌های گروه پایین (۲۵ درصد)}$$

Scoremax = بالاترین نمره ممکن در سؤال

Score min = پایین‌ترین نمره ممکن در سؤال

روش نیتکو در تحلیل سؤال آزمون‌های تشریحی و عملکردی

در این روش، ضریب دشواری یک سؤال تشریحی یا عملکردی به صورت نمره میانگین گروه آزمون‌شوندگان برای آن سؤال تعریف شده است. دقت کنید که نمره میانگین سؤالی که به صورت ۰ و ۱ نمره‌گذاری می‌شود برابر است با

ارزش **p** که آن از فرمول $p = \frac{R}{T}$ محاسبه می‌شود. برای آشنا شدن با چگونگی محاسبه ضریب دشواری سؤال‌های

تشریحی و عملکردی با روش نیتکو از فرمول زیر استفاده می‌شود:

$$\text{نمره میانگین سؤال} = \frac{\text{نمره میانگین سؤال}}{\text{ضریب دشواری}}$$

دامنه ممکن نمرات سؤال

پایین ترین نمره ممکن - بالاترین نمره ممکن

و برای محاسبه ضریب تمیز از فرمول زیر استفاده می شود:

$$\text{ضریب تمیز} = \frac{\text{تفاوت بین نمرات میانگین گروه های بالا و پایین برای سؤال}}{\text{دامنه نمرات ممکن برای سؤال}} = \frac{\text{نمره میانگین سؤال برای گروه پایین - نمره میانگین سؤال برای گروه بالا}}{\text{پایین ترین نمره ممکن سؤال - بالاترین نمره کل سؤال}}$$

تحلیل سؤالات آزمون های وابسته به ملاک

در آزمون های وابسته به ملاک، استفاده از ضریب های دشواری و تمیز کاربرد چندانی ندارد، زیرا نتیجه آرمانی آزمون های وابسته به ملاک این است که در آنها اکثر آزمون شونده ها را درست جواب می دهند. در چنین حالتی ضریب دشواری سؤال ها بسیار بزرگ و ضریب تمیز آنها بسیار کوچک خواهد بود، یعنی برای گروه آزمودنی مورد نظر اکثر سؤال ها آسان و قدرت تمیز آنها بسیار کم خواهد بود.

تمیز بین پیش و پس آزمون

بعضی از متخصصان ارزشیابی پیشرفت تحصیلی پیشنهاد می کنند که آزمون های وابسته به ملاک مورد استفاده در سنجش یادگیری در حد تسلط تنها باید شامل سؤال هایی باشند که پس از آزمون مطالب درسی آسانتر از پیش از آموزش آن مطالب جلوه کنند. برای تحقق این امر انجام مراحل زیر پیشنهاد شده است:

۱- اجرای آزمون پیش از آموزش و پس از آن

۲- محاسبه ضریب دشواری هر سؤال برای پیش آزمون و پس آزمون

۳- محاسبه تفاوت ضرایب دشواری سؤال های پیش آزمون و پس آزمون که این ضریب را ضریب تمیز پیش

آزمون و پس آزمون می نامند و آن را با D نشان می دهند. نام دیگر این ضریب، ضریب حساسیت به آثار آموزش است (S). این ضریب با فرمول زیر محاسبه می گردد:

$$S \text{ یا } D = \frac{R_A - R_B}{T}$$

در این فرمول:

$R_A = P_A$ تعداد کسانی که پس از آموزش به سؤال جواب درست داده اند.

تعداد کسانی که پیش از آموزش به سؤال جواب درست داده اند.

$P_B R_B =$

(تعداد کل افراد) تعداد کسانی که هم پیش و هم پس از آموزش به سؤال جواب داده اند $T =$

تمیز بین حد تسلط و غیر حد تسلط

برای محاسبه این ضریب یک شکل چهارخانه‌ای درست می‌شود که در خانه‌های عمودی آن حد تسلط و غیر حد تسلط و در خانه‌های افقی آن درست و غلط قرار داده می‌شوند. در این جدول، حرف a معرف تعداد آزمون‌شوندگانی است که سؤال را درست جواب داده‌اند اما در کل آزمون نمره‌ای کمتر از حد تسلط گرفته‌اند، b تعداد آزمون‌شوندگانی است که سؤال را درست جواب داده‌اند و در نمره کل آزمون نیز به حد تسلط رسیده‌اند، حرف c نشان‌دهنده تعداد آزمون‌شوندگانی است که سؤال را غلط جواب داده‌اند و در کل به حد تسلط نرسیده‌اند و حرف d معرف تعداد کسانی است که سؤال مورد نظر را درست جواب نداده‌اند اما در نمره کل آزمون به حد تسلط رسیده‌اند. تعداد آزمون‌شوندگانی که به حد تسلط رسیده‌اند برابر است با $b + d$ و تعداد کسانی که به حد تسلط نرسیده‌اند مساوی است با $a + c$.

ضریب تمیز یعنی D برابر است با تفاوت بین سطوح دشواری سؤال برای کسانی که به حد تسلط رسیده و آنهایی که به حد تسلط نرسیده‌اند، مطابق با فرمول زیر است:

a	b
c	d

$$D = \frac{b}{b+a} - \frac{a}{a+c}$$

تمیز بین حد تسلط و غیر حد تسلط

تحلیل آزمون به کمک نظریه سؤال - پاسخ

نظریه سؤال - پاسخ بر نظریه یا الگوی صفت مکنون استوار است. نظریه صفت مکنون بر این فرض استوار است که یک صفت یا ویژگی زیربنایی وجود دارد که به شخص امکان می‌دهد تا در یک تکلیف شناختی معین موفقیت کسب نماید. از جمله اینگونه صفات می‌توان «دانش معنی لغات»، «استدلال ریاضی» یا «تجسم فضایی» را نام برد. نظریه سؤال - پاسخ، به جای تأکید بر نمرات کل آزمون، بر پاسخ آزمودنی‌ها به تک تک سؤالات آزمون تکیه دارد. برای تحلیل سؤال‌های این آزمون‌ها بیشتر از کامپیوتر استفاده می‌شود.

منحنی ویژه سؤال و استفاده از آن برای تحلیل سؤال آزمون

منحنی ویژه سؤال که برای آزمون رسم می‌شود. احتمال پاسخ درست دادن به هر سؤال را به توانایی آزمون‌شونده ربط می‌دهد. به سخن دیگر، نظریه منحنی ویژگی سؤال یک بازنمایی نموداری از رابطه بین احتمال پاسخ درست دادن به یک سؤال و موقعیت آزمون‌شونده در صفت مورد اندازه‌گیری توسط آزمون است. در این منحنی، بر روی محور افقی (x) نمره کل آزمون و بر روی محور عمودی (y) نسبت به آزمون‌شوندگانی که به سؤال پاسخ درست داده‌اند مشخص می‌شود.

از روی منحنی ویژگی سؤال می‌توان ضریبهای دشواری و تمیز سؤال را تعیین کرد. ضریب دشواری عبارت است

از «نمره معیاری که در آن ۵۰ درصد آزمون‌شوندگان سؤال را درست پاسخ داده‌اند.» ضریب تمیز برابر است با شیب منحنی ویژگی سؤال. اگر همه آزمون‌شوندگان به یک سؤال جواب درست بدهند، منحنی ویژگی آن سؤال یک خط افقی واقع در بالای محور عمودی خواهد بود که از نسبت (۱) می‌گذرد. در این صورت ضریب دشواری سؤال $p = 1$ است. به عکس، اگر همه آزمون‌شوندگان به یک سؤال جواب غلط بدهند، منحنی ویژگی آن سؤال نیز یک خط افقی خواهد بود که این بار از پایین محور عمودی و از نسبت صفر می‌گذرد. ضریب دشواری این سؤال $P = 0$ است.

نکات مهم و کلیدی

- آزمون‌های که سوالات آنها درهم و برهم و خوانا نیستند موجب دلسردی آزمون‌شونده می‌شوند.
- هرچه سطح زیر منحنی ویژگی سؤال بیشتر باشد، ضریب دشواری سؤال بزرگتر خواهد بود.
- هرچه منحنی ویژگی یک سؤال حالت پلکانی بیشتری داشته باشد همبستگی بین آن سؤال و کل آزمون بیشتر است.
- حد تسلط را باید همیشه نمره ۲۰ در نظر گرفت.
- ترتیب بندی سوالات برای ورقه امتحانی: ۱- صحیح - غلط ۲- جورکردنی ۳- چند گزینه‌ای ۴- کوتاه پاسخ ۵- تشریحی
- مهمترین عامل در اجرای آزمون اضطراب است.
- در تحلیل سوالهای آزمون نقاط قوت و ضعف یک آزمون و کیفیت سوالهای آن تعیین می‌شود.
- در شرایطی که کل آزمون شونده‌گان بیش از ۴۰ نفر است بهترین رقم برای گروه بالا و پایین ۲۷٪ کل برگه‌هاست.
- ضریب دشواری ۳۵ یعنی سوال دشوار بوده است؛ ۳۵ درصد آسان بوده است و ۶۵ درصد سخت بوده است.
- هر اندازه ضریب دشواری یک سوال بزرگتر باشد؛ سوال آسانتر است.
- ضریب دشواری همان ضریب آسانی یا ضریب سهولت است.
- هرچه واریانس نمرات حاصل از یک آزمون وابسته به هنجار بزرگتر باشد آن آزمون، آزمون بهتری است چون دارای پراگندگی بیشتری است.
- سوالی که با سایر سوالات همبستگی ندارد دارای واریانس کوچکتری است و کمک زیادی به واریانس کل نمی‌کند.
- انتخاب سوالهایی با ضریب دشواری نامناسب به نوع سوالهای آزمون مربوط است.
- منظور از سطح موفقیت یا حدس زدن عبارت است از نسبت پاسخهای درستی که با حدس زدن به دست می‌آید.
- ضریب دشواری بین ۰/۳ تا ۰/۷ حداکثر اطلاع را درباره تفاوت بین آزمودنی‌ها به دست می‌دهد.
- برای سوالات کارشناسی ارشد با سوالات دشوار باشد یعنی ضریب دشواری کم باشد.
- ضریب دشواری میزان آسان بودن یا دشوار بودن یک سوال را برای گروه نشان می‌دهد.
- هر قدر ضریب تمیز بزرگتر باشد قوه تمیز آن سوال بیشتر است.
- ضریب تمیز صفر حامی از آن است که سوال به هیچ وجه نتوانسته است بین گروه قوی و ضعیف تمایز قائل شود.

- ضریب تمیز منفی نشان می‌دهد که گروه قوی بدتر از گروه ضعیف عمل کرده است.
- سوالات دارای قوه تمیز خوب با کل آزمون همبسته‌اند.
- ضرایب تمیز سوالات باید مثبت باشد.
- ضریب تمیز سوال نوعی ضریب توافق سوال با کل آزمون است.
- همبستگی بین سوال و آزمون مستقیماً به پایایی آزمون مربوط است.
- اگر ضریب دشواری برابر $0/8$ باشد ضریب تمیز بین $0/4 +$ و $0/4 -$ خواهد بود.
- ضریب تمیز می‌تواند مثبت، منفی و صفر باشد.
- سوال با ضریب تمیز بالا نیاز به اصلاح مجدد دارد.
- ضریب دشواری و ضریب تمیز صرفاً برای آزمونهای وابسته به هنجار مناسبند.
- ضریب تمیز پیش و پس آموزش را ضریب حساسیت نسبت به آثار آموزش نیز گویند.
- ضریب تمیز منفی نشان دهنده این است که سوال دارای مشکل اساسی است.
- نظریه سوال پاسخ بر نظریه صفت مکنون استوار است.
- نظریه سوال پاسخ به جای تاکید بر نمرات کل آزمون بر پاسخهای آزمودنی ها به تک تک سوالات آزمون تاکید دارد.
- منحنی ویژگی سوال احتمال پاسخ درست دادن را نسبت به صفت مکنون تعیین می‌کند.

فصل یازدهم:

استفاده از روش های آماری برای تفسیر نمره های

آزمون

گرایش مرکزی

اندازه های گرایش مرکزی برای طبقه بندی و خلاصه کردن اطلاعات، دقیق تر از جدول توزیع فراوانی هستند. ساده ترین روش برای تعیین جایگاه کلی نمره ها، شاخص معدل است. سه شاخص گرایش مرکزی عبارتند از نما، میانه و میانگین که هر کدام به شیوه متفاوتی به توصیف معدل می پردازند.

نما

نما عدد یا نمره ای است که در توزیع فراوانی یکسری داده دارای بیشترین فراوانی است. نما ساده ترین شاخص گرایش مرکزی است که با ملاحظه توزیع نمره ها تعیین می شود.

مثال: در توزیع اعداد: ۲-۴-۱۶-۲۰-۷-۳-۶-۳-۵-۳-۳-۳ نما عدد ۳ با فراوانی ۵ است.

عددی که دارای بیشترین فراوانی است اغلب در نزدیکی مرکز توزیع اعداد قرار دارد. نما در چنین شرایطی یک شاخص گرایش مرکزی محسوب می‌شود لکن نما همیشه در نزدیک مرکز توزیع فراوانی قرار ندارد به همین دلیل نمی‌توان به عنوان یک شاخص گرایش مرکزی به آن اطمینان کرد. در بین شاخص‌های گرایش مرکزی نما شاخصی بی‌ثبات است که مورد استفاده محدودی دارد و خصوصاً در گروه‌های کوچک دارای اعتبار نیست زیرا میزان آن تنها تابع چند عدد است. نما را برای هر نوع توزیعی می‌توان به کار برد اما تنها زمانی یک شاخص مرکزی است که با مقیاس اسمی بکار برده شود. محققان اغلب زمانی نما را به عنوان یک شاخص مرکزی بکار می‌گیرند که بخواهند بیش کلی درباره شاخص مرکزی کسب کنند. نما را نمی‌توان به صورت ریاضی دستکاری کرد برای همین از نظر محاسبات آماری هم دارای محدودیت است.

توزیع تک‌نمایی و بدون نما

توزیع تک‌نمایی، توزیعی است که در آن یک عدد یا نمره دارای بیشترین فراوانی است. مانند مثال بالا که عدد ۳ با فراوانی ۵ است. در مقابل در برخی از توزیع اعداد هیچ عددی که فراوانی بیشتری از بقیه اعداد داشته باشد وجود ندارد، که چنین توزیعی دارای نما نمی‌باشد.

مثال: در توزیع اعداد: ۲-۴-۶-۳-۷-۲۰-۱۶

توزیع دونمایی و چندنمایی

توزیع دو نمایی، توزیعی است که در آن بیش از یک عدد یا نمره دارای بیشترین فراوانی است مانند مثال ۱ با اعداد ۷ و ۴؛ توزیعی که در آن بیش از دو عدد یا نمره دارای بیشترین فراوانی است، توزیع چند نمایی گفته می‌شود. مانند مثال ۲ با اعداد ۷ و ۱.

مثال ۱: ۲-۳-۸-۷-۷-۷-۶-۵-۶-۴-۴-۴-۱-۹-۸-۹

مثال ۲: ۲-۳-۴-۴-۴-۴-۱-۹-۸-۹-۸-۷-۷-۷-۶-۵-۱-۱-۱

ویژگی‌های نما

- نما از بین شاخص‌های گرایش مرکزی دارای کمترین مفروضه است.
- مقیاس اندازه‌گیری اسمی مناسب‌ترین مورد استفاده نما است.
- اگر بر آورد سریعی از اندازه‌های مرکزی توزیع لازم باشد نما مورد استفاده قرار می‌گیرد.
- اگر بخواهیم معدل یا عدد دارای بیشترین فراوانی را مشخص کنیم نما مورد استفاده قرار می‌گیرد.

• سهولت محاسبه

• نما را می‌توان با توجه به روابط موجود بین شاخص‌های گرایش به مرکز نیز محاسبه کرد:

$$\bar{X} - 2Md = 3Mod$$

در داده‌های طبقه‌بندی شده نما را می‌توان از طریق فرمول زیر محاسبه کرد. استفاده از این فرمول زمانی مناسب است که توزیع مورد نظر بهنجار و تک‌نمایی باشد.

میانه

میانه نقطه وسط در توزیع نمره‌هاست به عبارت دیگر میانه نقطه‌ای است که نیمی از توزیع نمره‌ها در بالای آن و نیمی دیگر در پایین آن قرار دارد. اولین قدم در محاسبه میانه مرتب کردن اعداد می‌باشد و سپس نمره‌ای که توزیع را به دو نیم تقسیم می‌کند به عنوان میانه انتخاب می‌نماییم.

• در صورتی که تعداد نمره‌ها در توزیع فرد باشد، میانه عددی است که در وسط آن قرار دارد.

مثال: ۱-۳-۴-۱۸-۱۰-۱۱-۱۲-۱۰-۱۶

در مثال فوق عدد وسط توزیع ۱۲ است که ۴ عدد در سمت پایین و ۴ عدد در سمت بالای آن قرار دارند.

• در صورتی که تعداد نمره‌ها زوج باشد، میانه عبارت است از معدل دو نمره‌ای که در وسط واقع می‌شوند.

مثال: ۱-۳-۴-۱۸-۱۰-۱۱-۱۲-۱۴-۱۱-۱۶

در مثال فوق عدد وسط توزیع ۱۲ و ۱۴ است چون تعداد نمره‌ها زوج است بنابراین خواهیم داشت:

$$12 + 14 = 26 / 2 = 13$$

پس عدد ۱۳ میانه محسوب می‌شود.

• چنانچه نمره یا عددی که توزیع را به دو قسمت تقسیم می‌کند تکراری باشد، میانه به صورت زیر محاسبه می‌شود. به نمره‌های زیر توجه کنید:

مثال: ۱۰-۹-۹-۸-۷-۷-۷-۶-۵-۴

چون تعداد اعداد ۱۰ می‌باشد، میانه نقطه‌ای است که ۵ عدد در بالای آن و ۵ عدد در پایین آن واقع می‌شود. به منظور تعیین نقطه‌ای که توزیع را به دو قسمت تقسیم کند، فرض بر این است که سه تا عدد ۷ در حدود واقعی عدد ۷، یعنی $7/5 - 6/5$ پراکنده شده‌اند، این مفروضه با این فرض آماری که اعداد به صورت پیوسته توزیع شده‌اند هماهنگی دارد. بدین ترتیب میانه در محدوده $6/5 - 7/5$ قرار دارد.

سه عدد زیر عدد $6/5$ (حد پایین ۷) قرار گرفته است. بنابراین برای یافتن میانه از بین کلیه اعدادی که در محدوده $6/5 - 7/5$ قرار دارند، فقط به دو عدد نیاز داریم چون سه تا هفت وجود دارد و ما تنها به دو تا از آنها نیاز داریم،

فصل یازدهم : استفاده از روش‌های آماری برای تفسیر نمره‌های آزمون / ۹۵

بنابراین دو سوم ($\frac{2}{3}$) فاصله‌ای را که بین $6/5 - 7/5$ قرار دارد، انتخاب می‌کنیم. عدد ۲ که در صورت قرار دارد عبارت است از تعداد اعدادی که لازم داریم و عدد ۳ نشان دهنده تعداد اعداد ۷ است که در حدود واقعی $6/5 - 7/5$ قرار دارند.

$$\text{بنابراین میانه برابر است با: } 6/5 + \frac{2}{3} = 6/5 + 0/67 = 7/17$$

محاسبه میانه در توزیع فراوانی بصورت جدول

برای مثال جدول اعداد زیر را در نظر بگیرید.

x	۳	۴	۵	۶	۷	۸	۹	۱۰
f	۱	۲	۲	۴	۵	۳	۲	۱

N=20

تعداد اعداد توزیع فوق ۲۰ عدد است که یک توزیع زوج محسوب می‌شود. بنابراین باید ۱۰ عدد زیر نقطه یا عددی باشد که وسط توزیع محسوب می‌شود. عدد وسط در این توزیع ۷ است. اما عدد ۷ پنج بار تکرار شده است. بنابراین باید حدپایین عدد ۷ یعنی $6/5$ با یک پنجم یا $.2$ جمع کرد حاصل $6/7$ یا میانه توزیع است.

$$6/5 + .2 = 6/7$$

برای محاسبه میانه از فرمول زیر استفاده می‌شود:

$$Md = L + \frac{\frac{N}{2} - cf}{f}(i)$$

L = حد پایین طبقه‌ای که میانه در آن است.

cf = فراوانی تراکمی تا طبقه‌ای که میانه در آن است.

f = فراوانی مطلق طبقه‌ای که میانه در آن است.

i = فاصله طبقات

N = تعداد نمره‌ها

با استفاده از فرمول مذکور برای توزیع نمرات مثال بالا خواهیم داشت:

$$Md = 6/5 + \frac{\frac{10}{2} - 3}{3}(1) = 6/5 + .67 = 7/16$$

محاسبه میانه با استفاده از فرمول $Md = L + \frac{\frac{N}{2} - cf}{f}(i)$ کارآترین روش محاسبه میانه می‌باشد.

محاسبه میانه برای اعداد طبقه بندی شده با فاصله بزرگتر از یک

تقسیم تعداد کل نمرها یا N به عدد ۲ یعنی وسط توزیع یا دو قسمت مساوی در توزیع نمره‌ها؛

تعیین اولین طبقه دارای فراوانی تراکمی بیشتر یا مساوی $\frac{N}{2}$: برای انجام این کار باید از طبقه پایین شروع کرد.
تعیین L یا حد پایین طبقه مورد نظر
استخراج cf و f از جدول توزیع فراوانی

جایگزینی اعداد استخراج شده در مراحل قبل در فرمول محاسبه میانه یعنی $Md = L + \frac{\frac{N}{2} - cf}{f}(i)$

ویژگی‌های میانه

اندازه یا حجم واحدهای اندازه‌گیری در میانه تأثیری ندارد.

چون میانه نسبت به اعداد بزرگ یا کوچک حساس نیست بهترین شاخص برای نشان دادن تمرکز اعداد در وسط توزیع محسوب می‌شود.

مجموع قدر مطلق انحراف نمره‌ها از میانه کوچکتر یا مساوی قدر مطلق انحراف نمره‌ها از هر عدد دیگری است.

$$\sum |x - Md| < \sum |x - c|$$

چون میانه بر اساس رتبه بندی شده به دست می‌آید، مورد استفاده آن زمانی است که مقیاس اندازه‌گیری رتبه‌ای باشد.

میانه را می‌توان با مقیاس‌های فاصله‌ای و نسبتی هم مورد استفاده قرار داد اما مقیاس‌های فاصله‌ای و نسبتی در محاسبه میانه مورد استفاده قرار نمی‌گیرند.

میانگین: مشهورترین و معتبرترین شاخص گرایش مرکزی

متداولترین شاخص مکان مرکزی که کاربرد گسترده‌ای دارد، میانگین است. میانگین، معدل حسابی گروهی از

نمره‌ها محسوب می‌شود. که حاصل محاسبه زیر است:

جمع کل نمره‌ها تقسیم بر تعداد کل نمره‌ها = میانگین

مثال: محاسبه میانگین اعداد:

$$15 - 12 - 11 - 6 - 5 - 44 - 11 - 16 - 1$$

$$\bar{X} = 166 \div 10 = 16.6 \text{ یا } 16\frac{6}{10} = 16.6 \text{ یا } 16\frac{3}{5}$$

میانگین حسابی، شاخص مناسبی از مکان مرکزی برای متغیرهای فاصله‌ای و نسبی است و روش محاسبه آن نیز از طریق جمع نمره‌ها تقسیم بر تعداد آنها بدست می‌آید:

$$\bar{X} = \frac{\sum X}{N} \rightarrow n\bar{x} = \sum X$$

میانگین جامعه با علامت μ نمایش داده می‌شود و فرمول محاسبه میانگین جامعه عبارت است از:

$$\mu_x = \frac{\sum X}{N}$$

میانگین نمونه با علامت \bar{x} نمایش داده می‌شود و فرمول محاسبه میانگین نمونه عبارت است از:

$$\bar{x} = \frac{\sum x}{n}$$

دقت کنید که نحوه محاسبه میانگین نمونه و جامعه یکسان است و فقط علائم آنها تفاوت دارند.

محاسبه میانگین در جدول توزیع فراوانی با فاصله طبقات یک

در توزیع فراوانی عددی که ارزش x بیش از یک فراوانی باشد برای حساب کردن میانگین باید علاوه بر فراوانی باید fx را هم محاسبه کنیم.

X: اعداد	۱	۲	۳	۴	۵	۶	۷	۸
F: فراوانی اعداد	۱	۱	۱	۱	۲	۳	۲	۱
Fx: فراوانی اعداد ضربدر اعداد	۱	۲	۳	۴	۱۰	۱۸	۱۴	۸

$$N = 20 \dots \dots \dots \sum fx = 280$$

میانگین توزیع فوق عبارت است از:

$$\mu_x = \frac{\sum fx}{N} \text{ یا } 280 \text{ تقسیم بر } 20 \text{ مساوی } 14$$

محاسبه میانگین در جدول توزیع فراوانی با فاصله طبقات بیشتر از یک

برای محاسبه میانگین در جدول توزیع فراوانی با فاصله طبقات بیشتر از یک از نقطه میانی هر طبقه استفاده می‌شود و میانگین مساوی حاصل ضرب نقطه میانی هر طبقه در فراوانی همان طبقه تقسیم بر تعداد کل اعداد است.

۱- محاسبه نقطه میانی طبقه‌ها

- ۲- ضرب فراوانی هر طبقه در نقطه میانی همان طبقه
 ۳- محاسبه مجموع حاصلضرب نقطه میانی در فراوانیها
 ۴- تقسیم مجموع بند ۳ بر تعداد کل نمره‌ها یا N

فرمول محاسبه:

$$\bar{x} = \frac{\sum fx'}{N}$$

محاسبه میانگین اعداد طبقه‌بندی شده از راه کوتاه

در این روش باید فاصله طبقات مساوی باشند که با طی کردن مراحل زیر بدست می‌آید:

- ۱- مشخص کردن طبقه‌ای که فرض می‌شود میانگین در آن طبقه است
 - طبقه‌ای که در وسط قرار دارد.
 - طبقه‌ای که دارای بزرگترین فراوانی است.
- ۲- تنظیم ستون اکس کلاه دار \bar{x} :
 - گذاشتن عدد صفر در جلوی طبقه‌ای که فرض می‌شود میانگین در آن است.
 - مشخص کردن فاصله هریک از طبقات با طبقه فرضی
 - گذاشتن علامت منفی به طبقات پایین و گذاشتن علامت مثبت به طبقات بالا
- ۳- ضرب فراوانی هر طبقه در عدد ستون \bar{x} و تشکیل ستون $f\bar{x}$
- ۴- جمع جبری اعداد ستون $f\bar{x}$
- ۵- جایگزینی اعداد در فرمول:

$$\bar{x} = x' + \frac{\sum f\bar{x}}{N}(i)$$

مثال:

طبقات	۱-۳	۴-۶	۷-۹	۱۰-۱۲	۱۳-۱۵	۱۶-۱۸	۱۹-۲۱	۲۲-۲۴	جمع
F	۱۲	۴	۵	۶	۲۱	۶	۴	۱	N= 59
\bar{x}	-۴	-۳	-۲	-۱	۰	+۱	+۲	+۳	
$f\bar{x}$	-۴۸	-۱۲	-۱۰	-۶	۰	۶	۸	۳	$\sum f\bar{x} = -59$

با توجه به داده‌های فوق و فرمول $\bar{x} = x' + \frac{\sum f\bar{x}}{N}(i)$ خواهیم داشت:

$$\bar{x} = 14 + \frac{-59}{59}(3) = 11$$

میانگین میانگین

محاسبه میانگین میانگین‌ها با محاسبه میانگین مرکب یا موزون امکان‌پذیر است. میانگین میانگین به دو شکل و در دو موقعیت قابل محاسبه است:

۱- اگر حجم نمونه‌ها مساوی باشد: میانگین مرکب عبارت از مجموع میانگین‌ها تقسیم بر تعداد آنها. فرمول محاسبه:

$$\bar{x}_T = \frac{\sum \bar{x}}{N}$$

۲- اگر حجم نمونه‌ها مساوی نباشد: مجموع حاصل ضرب میانگین هر گروه در تعداد آن تقسیم بر مجموع کل نمونه‌ها یا گروه‌ها. فرمول محاسبه:

$$\bar{x}_T = \frac{\sum \bar{x}_i n_i}{N_T}$$

میانگین هندسی

چنانچه n نمره یا اندازه‌گیری در اختیار داشته باشیم، میانگین هندسی از فرمول زیر به دست می‌آید:

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n}$$

به عنوان مثال، میانگین هندسی دو عدد ۲ و ۸ برابر با ۴ است.

$$G = \sqrt[2]{(2)(8)} = \sqrt{16} = 4$$

میانگین همساز یا هارمونیک

میانگین همساز یا هارمونیک یکی دیگر از میانگین‌های مورد استفاده در برخی مسائل پژوهشی است که با فرمول زیر محاسبه می‌شود.

$$HM = \frac{1}{\frac{1}{\mu} \left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_N} \right)}$$

در ارتباط با میانگین‌های مختلف می‌توان گفت میانگین حسابی بزرگتر از میانگین هندسی و میانگین هندسی بزرگتر از میانگین همساز است.

$$\bar{X} > G > HM$$

میانگین حسابی < میانگین هندسی < میانگین همساز یا

ویژگی‌ها و موارد استفاده میانگین

موارد استفاده	ویژگی‌ها
<p>برای تفسیر داده‌های فاصله‌ای</p> <p>دخالت ارزش عددی همه نمره‌ها در محاسبه</p> <p>برای سؤال پژوهشی کمی</p> <p>استفاده از میانگین برای محاسبات دیگر</p>	<p>یک شاخص فاصله‌ای است</p> <p>استفاده از عملیات ریاضی بیشتر</p> <p>استفاده از همه اعداد و نمره‌ها در محاسبه</p> <p>تاثیر نمره‌های خیلی بزرگ و کوچک در آن</p> <p>مجموع انحراف از میانگین همیشه صفر است</p> <p>با ثبات ترین شاخص مرکزی است</p> <p>بهترین برآورد از میانگین جامعه است</p> <p>کاربرد با مقیاس‌های فاصله‌ای و نسبی</p>

مقایسه سه شاخص گرایش مرکزی

- میانگین و میانه دارای ثبات و حساسیت بیشتری نسبت به نما هستند.
- میانه یک شاخص ترتیبی است و تحت تاثیر ارزش عددی نمره‌ها قرار نمی‌گیرد در حالی که میانگین به ارزش عددی تمام نمره‌ها حساس است.
- در یک منحنی طبیعی در صورتی که توزیع طبیعی، متقارن و تک‌نمایی باشد میانگین، میانه و نما بر روی هم قرار می‌گیرند.
- در یک توزیع دونمایی متقارن، میانگین و میانه برابر ولی نما در طرفین قرار دارد.
- در یک توزیع دارای کجی مثبت یا منفی سه شاخص برابر نیستند و در سه نقطه مختلف توزیع قرار خواهند گرفت.

- در توزیع‌های دارای کجی مثبت میانگین همیشه بزرگتر از میانه است و نما کوچکتر از میانه است.
توزیع‌های دارای کجی مثبت = میانگین < میانه < نما یا $M < Md < \bar{x}$
- در توزیع‌های دارای کجی منفی میانگین همیشه کوچکتر از میانه است و نما بزرگتر از میانه است.
توزیع‌های دارای کجی منفی = میانگین > میانه > نما یا $M > Md > \bar{x}$
- میانگین پایدارترین و معتبرترین شاخص مرکزی است.
- برای استنباط جامعه از نمونه میانگین معتبرترین برآورد از میانگین جامعه است.

معیارهای انتخاب یک اندازه گرایش مرکزی (مقیاس اندازه‌گیری)

- برای مقیاس اسمی نما بهترین شاخص است.
- برای مقیاس ترتیبی میانه بهترین شاخص است.
- برای مقیاس فاصله‌ای یا نسبی میانگین بهترین شاخص است.

هدف محاسبه:

- برای شرکت دادن ارزش یا مقدار هریک از نمره‌ها در شاخص مرکزی، میانگین بهترین شاخص است.
- برای جلوگیری از تاثیر نمره‌های خیلی بزرگ یا خیلی کوچک در محاسبه شاخص مرکزی، میانه بهترین شاخص است.
- اگر هدف استنباط یا برآورد باشد، میانگین بهترین شاخص است.

شاخص‌های پراکندگی

پراکندگی یکی از ویژگی‌های عمومی توزیع نمره‌ها است. شاخص‌های پراکندگی، میزان تغییرات و پراکندگی را که در داخل یک توزیع وجود دارد نشان می‌دهد. دامنه تغییرات، دامنه چارکی متوسط، انحراف متوسط، واریانس و انحراف معیار از شاخص‌های پراکندگی محسوب می‌شوند. وجود یا عدم وجود پراکندگی ذاتاً خوب یا بد نیست بلکه با توجه به هدف پژوهش این وضعیت مشخص می‌گردد.

دامنه تغییرات

ساده‌ترین شاخص پراکندگی دامنه تغییرات است که از رابطه زیر حاصل می‌شود:

$$\text{دامنه تغییرات} = \text{کوچکترین عدد} - \text{بزرگترین عدد} \quad R = X_H - X_L \quad \text{یا} \quad R = (X_H - X_L) + 1$$

- عدد به اضافه یک به منظور نشان دادن حدپایین و بالای هر طبقه منظور شده است.

ویژگی‌های دامنه تغییرات

- محاسبه خیلی سریع و سهل.
- عدم توصیف حقیقی توزیع نمره‌ها.
- یک شاخص پایدار پراکندگی نیست.
- تحت تاثیر بزرگترین یا کوچکترین نمره توزیع است.
- مورد استفاده با مقیاس فاصله‌ای.
- برای مقیاس‌های اسمی و ترتیبی شاخص مناسبی نیست.
- بی اعتبارترین و بی ثبات ترین شاخص پراکندگی دامنه تغییرات است.
- پایداری انحراف چارکی از دامنه تغییرات بیشتر است.
- مقدار دامنه تغییرات وابسته به حجم نمونه است.
- انحراف چارکی

انحراف چارکی که از دامنه تغییرات باثبات تر است عبارت است از نصف فاصله بین چارک اول و چارک سوم.

$$Q = \frac{Q_3 - Q_1}{2}$$

چارکها بر روی مقیاس اندازه‌گیری نقاطی هستند که کل نمره‌ها را به چهار قسمت مساوی یعنی چهار تا ۲۵ درصد تقسیم می‌کنند. به گونه‌ای که نقطه ۲۵ درصدی را چارک اول، نقطه ۵۰ درصدی را چارک دوم و نقطه ۷۵ درصدی را چارک سوم می‌نامند. دامنه تغییرات بین چارکها از تفاضل چارک سوم بر چارک اول حاصل می‌شود.

$$Q_3 - Q_1$$

چارک اول (Q_1) نقطه‌ای در روی مقیاس اندازه‌گیری است که ۲۵ درصد نمره‌ها را از پایین جدا می‌کند و در مکانی قرار گرفته که درست $\frac{1}{4}$ نمره‌ها از آن کوچکتر و $\frac{3}{4}$ نمره‌ها از آن بزرگتر است. فرمول محاسبه آن بصورت زیر می‌باشد:

$$Q_1 = L + \frac{\frac{3n}{4} - cf}{f} (i)$$

چارک دوم (Q_2) برابر میانه است، یعنی نقطه‌ای که توزیع را به دو قسمت مساوی تقسیم می‌کند.

چارک سوم (Q_3) نقطه‌ای است که $\frac{3}{4}$ نمره‌ها در زیر آن و بقیه در بالای آن واقع شده‌اند. به چارک سوم نقطه

۷۵ درصدی نیز گفته می‌شود. فرمول محاسبه آن بصورت زیر می‌باشد:

$$Q_3 = L + \frac{\frac{3n}{4} - cf}{f} (i)$$

محاسبه چارک اعداد طبقه‌بندی نشده

اعداد را از کوچک به بزرگ مرتب کرده، میانه اعداد که چارک دوم خواهد بود محاسبه می‌کنیم و بعد از آن میانه سمت چپ را محاسبه می‌کنیم که چارک اول خواهد بود و در نهایت میانه سمت راست را محاسبه می‌کنیم که چارک سوم خواهد بود.

محاسبه چارک اعداد طبقه‌بندی شده

برای محاسبه چارک اعداد طبقه‌بندی شده به توزیع فراوانی تراکمی نیاز هست. در ستون توزیع فراوانی تراکمی عددی که مساوی یا بزرگتر از $\frac{N}{4}$ باشد، طبقه‌ای است که چارک اول در آن جای دارد. در ادامه از فرمول زیر استفاده می‌کنیم:

$$Q_1 = L + \frac{\frac{N}{4} - cf}{f} (i)$$

Q_1 = چارک اول

L = حدپایین طبقه‌ای که چارک اول در آن قرار دارد

N = تعداد نمره‌ها

Cf = فراوانی تراکمی تا طبقه‌ای که چارک اول در آن قرار دارد

f = فراوانی مطلق طبقه‌ای که چارک اول در آن قرار دارد

i = فاصله طبقات

برای محاسبه چارک سوم از ستون فراوانی تراکمی طبقه مساوی یا بزرگتر از $\frac{3N}{4}$ را پیدا می‌کنیم و از فرمول

زیر استفاده می‌کنیم:

$$Q_3 = L + \frac{\frac{3n}{4} - cf}{f} (i)$$

بعد از محاسبه چارک اول و سوم و جایگزینی آنها در فرمول $Q = \frac{Q_3 - Q_1}{4}$ انحراف چارکی حاصل می‌شود.

ویژگی‌های شاخص انحراف چارکی

- محاسبه انحراف چارکی به دقت بیشتری نیاز دارد.
- نتیجه انحراف چارکی دارای ثبات بیشتری است.
- خیلی کم تحت تاثیر نمره‌های انتخاب شده تصادفی است.
- همانند میانه تحت تاثیر نمره‌های خیلی بزرگ یا کوچک قرار ندارد.
- بهترین مورد استفاده شاخص انحراف چارکی زمانی است که نمره‌های خیلی بزرگ یا خیلی کوچک شکل توزیع نمره‌ها را از بین می‌برد.
- شرایط مورد استفاده از میانه و انحراف چارکی مشابه اند و هر دو یک خانواده آماری را تشکیل می‌دهند.
- محاسبه انحراف چارکی با مقیاس فاصله‌ای صورت می‌گیرد.
- در محاسبه انحراف چارکی کلیه ارزشهای مقداری تمام اعداد، مورد بررسی قرار نمی‌گیرند.
- زمانی که نمره‌ها دارای کجی مثبت یا منفی باشند انحراف چارکی شاخص پراکندگی مناسبی است.

انحراف متوسط

انحراف به فاصله بین هر عدد از هریک از شاخص‌های مرکزی اشاره دارد؛ میانگین قدر مطلق انحرافها از میانگین را انحراف متوسط می‌نامند.

فرمول محاسبه انحراف متوسط

$$MD = \frac{\sum |x_i|}{N}$$

ویژگی‌های انحراف متوسط

- در صورتی که پراکندگی نمره‌ها بیشتر شود انحراف نمره‌ها از میانگین هم بیشتر می‌شود.
- نحوه محاسبه انحراف متوسط برای داده‌های طبقه بندی شده و نشده همانند است با این تفاوت که در اعداد طبقه بندی شده نقاط میانی طبقه‌ها از میانگین محاسبه می‌شود.
- با مقیاس فاصله‌ای مورد استفاده قرار می‌گیرد.
- انحراف متوسط میزان پراکندگی موجود در توزیع نمره‌ها را نشان می‌دهد.
- محاسبه انحراف متوسط نسبتاً آسان است.

- میزان انحراف متوسط از توزیعی به توزیع دیگر دستخوش تغییر است. (مشکل انحراف متوسط)
- بکارگیری همراه با احتیاط به دلیل عدم امکان انجام عملیات جبری.
- در محاسبه انحراف متوسط علائم اعداد مورد بررسی قرار نمی‌گیرند.

واریانس

واریانس یک شاخص پراکندگی است که از طریق محاسبه انحراف نمره‌ها از میانگین محاسبه می‌شود به عبارتی میانگین انحراف نمره‌ها از میانگین یا مجموع مجذور انحراف نمره‌ها از میانگین تقسیم بر تعداد نمره‌ها است.

محاسبه واریانس اعداد طبقه‌بندی نشده

۱. محاسبه میانگین \bar{x}
۲. محاسبه انحراف نمره‌ها از میانگین $(x - \bar{x})$
۳. به توان دو رساندن انحرافهای محاسبه شده $(x - \bar{x})^2$
۴. محاسبه مجموع مجذور انحراف از میانگین $\sum(x - \bar{x})^2$
۵. تقسیم مجموع مجذور انحراف از میانگین بر تعداد کل نفرات یا N

فرمول محاسبه واریانس نمونه:

$$S^2 = \frac{\sum(x - \bar{x})^2}{n - 1} \text{ or } S^2 = \frac{\sum x^2}{n - 1}$$

فرمول محاسبه واریانس جامعه:

$$\sigma^2 = \frac{\sum(x - \mu)^2}{N} = \frac{\sum x^2}{N}$$

در صورتی که میانگین اعشار داشته باشد محاسبه واریانس طولانی و دشوار خواهد شد بنابراین بجای میانگین مقدار مساوی آن را در فرمول محاسبه قرار می‌دهیم. محاسبه واریانس از راه اعداد خام عبارت است از:

$$\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}$$

در شرایطی که هدف برآورد واریانس جامعه باشد از فرمول زیر استفاده می‌کنیم:

$$S^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

محاسبه واریانس اعداد طبقه‌بندی شده

روش اول: محاسبه واریانس با استفاده از روش مستقیم

$$S^2 = \frac{\sum fx^2}{n} \text{ محاسبه واریانس با استفاده از نمره‌های انحرافی:}$$

طبقات	۱-۳	۴-۶	۷-۹	۱۰-۱۲	۱۳-۱۵	۱۶-۱۸	۱۹-۲۱	۲۲-۲۴	جمع
F	۱۲	۴	۵	۶	۲۱	۶	۴	۱	N= 59
x'	۲	۵	۸	۱۱	۱۴	۱۷	۲۰	۲۳	
fx'	۲۴	۲۰	۴۰	۶۶	۲۹۴	۱۰۲	۸۰	۲۳	۶۴۹
انحراف از میانگین $x = x' - \bar{x}$	-۹	-۶	-۳	۰	۰	۶	۹	۱۲	
x^2	۸۱	۳۶	۹	۰	۰	۳۶	۸۱	۱۴۴	
fx^2	۹۷۲	۱۴۴	۴۵	۰	۰	۲۱۶	۳۲۴	۱۴۴	۱۸۴۵

با توجه به داده‌های فوق و فرمول $\bar{x} = x' + \frac{\sum fx'}{N}(i)$ میانگین عبارت است از:

$$\bar{x} = 14 + \frac{-59}{59}(3) = 11$$

با توجه به داده‌های فوق و فرمول $S^2 = \frac{\sum fx'^2}{n}$ خواهیم داشت:

$$S^2 = \frac{\sum fx'^2}{n} = \frac{1845}{59} = 31/27$$

روش دوم: محاسبه واریانس با استفاده از اعداد خام

در این روش از فرمول $S^2 = \frac{\sum fx'^2}{n} - \left(\frac{\sum fx'}{n}\right)^2$ استفاده می‌شود. مراحل این محاسبه به ترتیب عبارتند از

۱. محاسبه نقطه میانی x'
۲. ضرب فراوانی در نقطه میانی fx'
۳. جمع حاصلضرب فراوانی در نقطه میانی $\sum fx'$
۴. حاصلضرب فراوانی در نقطه میانی ضربدر نقطه میانی fx'^2

۵. جمع حاصلضرب فراوانی در نقطه میانی ضربدر نقطه میانی $\sum fx'^2$

۶. جایگزینی اعداد محاسبه شده در فرمول $S^2 = \frac{\sum fx'^2}{n} - \left(\frac{\sum fx'}{n}\right)^2$

روش سوم: محاسبه واریانس با استفاده از میانگین فرضی یا روش غیر مستقیم

در این روش با استفاده از فرمول زیر واریانس را محاسبه می‌کنیم

$$S^2 = (i)^2 \left[\frac{\sum f\bar{x}^2}{n} - \left(\frac{\sum f\bar{x}}{n}\right)^2 \right]$$

ویژگی‌های واریانس

- واریانس کاربرد فراوانی در آمار استنباطی دارد.
- واریانس کاربرد محدودی در آمار توصیفی دارد.
- یک شاخص پراکندگی با ثبات است.
- واریانس با مقیاس فاصله‌ای و نسبی بکار می‌رود.

انحراف استاندارد یا معیار

محصول جذر واریانس انحراف استاندارد است انحراف استاندارد را می‌توان جذر ریشه دوم میانگین مجذور انحراف نمره‌ها نامید. انحراف معیار (σ) یا انحراف استاندارد یا جذر واریانس معتبرترین و پایدارترین شاخص پراکندگی است و با مقیاس‌های فاصله‌ای و نسبی بکار می‌رود. هر قدر مقدار انحراف معیار کوچکتر باشد جامعه یا افراد مورد مطالعه متجانس تر هستند. انحراف معیار مانند میانگین حسابی تحت تأثیر همه نمره‌ها قرار می‌گیرد.

ویژگی انحراف معیار بهنجار بودن توزیع آن است. شاید مهمترین فایده انحراف معیار این باشد که با استفاده از آن می‌توان مشخص کرد که چه نسبتی از نمره‌ها در فواصل مختلف نسبت به میانگین قرار گرفته‌اند. همچنین زمانی که بخواهیم ضرایب همبستگی را محاسبه کنیم، انحراف معیار شاخص مناسبی خواهد بود. انحراف معیار هرگز از انحراف متوسط کوچکتر نمی‌شود و نسبت انحراف معیار بر انحراف متوسط تقریباً برابر ۱/۲۵ است.

✚ فرمول انحراف استاندارد زمانی که میانگین نمره‌ها به صورت یک عدد صحیح است. این فرمول به محاسبه انحراف استاندارد با استفاده از میانگین معروف است:

$$\delta = \sqrt{\frac{\sum X^2}{N}}$$

✚ فرمول انحراف استاندارد زمانی که میانگین نمره‌ها به صورت یک عدد صحیح نیست یا عدد اعشاری است. این فرمول به محاسبه انحراف استاندارد با استفاده از اعداد خام معروف است:

$$\delta = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}}$$

ویژگی‌های انحراف استاندارد

- اگر به تمام نمره‌های یک توزیع عدد ثابتی اضافه شود انحراف استاندارد تغییری نمی‌کند.
- اگر به تمام نمره‌های یک توزیع عدد ثابتی ضرب شود انحراف استاندارد توزیع مذکور هم در همان عدد ضرب می‌شود.
- اگر همه نمره‌ها را در عدد ثابتی ضرب و یا تقسیم کنیم واریانس نمره‌ها در مجذور آن عدد ثابت ضرب یا تقسیم می‌شود اما انحراف استاندارد در خود آن عدد ضرب یا تقسیم می‌گردد.
- اگر کلیه اعداد را به یک عدد ثابت تقسیم نماییم مقدار واریانس تغییر نمی‌کند.
- با مقیاس اندازه‌گیری فاصله‌ای مورد استفاده است.
- معتبرترین، پایدارترین و باثبات‌ترین شاخص پراکندگی است.
- امکان انجام عملیات ریاضی را دارد.
- کاربرد گسترده‌ای در آمار استنباطی دارد.
- پایداری انحراف چارکی از انحراف استاندارد کمتر است.
- زمانی که توزیع دارای کجی است باید از انحراف استاندارد با احتیاط استفاده کرد.

تصحیح شیپرد

تصحیح برای اصلاح خطای طبقه‌بندی در انحراف استاندارد مورد استفاده قرار می‌گیرد. خطای طبقه‌بندی هم در محاسبه میانگین و هم در محاسبه انحراف استاندارد مطرح است. خطای مذکور در محاسبه میانگین آن قدر زیاد نیست که از نظر آماری دارای اهمیت باشد، زیرا انحراف‌های مثبت و منفی یکدیگر را خنثی می‌کنند. لکن در محاسبه انحراف معیار، خطای طبقه‌بندی به دلیل اینکه انحراف‌ها مجذور می‌شوند، بیشتر است. بنابراین برای تعدیل این خطا انحراف معیار محاسبه شده باید تصحیح گردد. میزان این خطا تابع فاصله طبقاتی است. هرچه فاصله

طبقاتی بزرگتر باشد، این خطا زیادتر خواهد بود. ضمناً باید توجه داشت که این تصحیح زمانی بکار می‌رود که تعداد طبقات از ۱۲ کمتر باشد زیرا در غیر این صورت مقدار خطا ناچیز خواهد بود. انحراف معیار با استفاده از فرمول زیر که به تصحیح شپرد معروف است، تصحیح می‌گردد.

فرمول تصحیح شپرد:

$$S_c = \sqrt{S^2 - \frac{i^2}{12}}$$

انحراف معیار تصحیح شده، S_c

انحراف استاندارد یا معیار S

فاصله طبقاتی i

انحراف استاندارد مرکب

برای محاسبه انحراف استاندارد مرکب از فرمول مک نیمار استفاده می‌شود. مقصود از انحراف استاندارد مرکب، محاسبه انحراف استاندارد دو چند انحراف استاندارد است. همانند محاسبه میانگین میانگینها.

فرمول محاسبه:

$$S_T = \sqrt{\frac{n_A(\bar{X}_{A^2} + S_{A^2}) + n_B(\bar{X}_{B^2} + S_{B^2})}{n_{A+n_B}} - \bar{X}_T^2}$$

ضریب پراکندگی

اگر پژوهشگر بخواهد پراکندگی یک ویژگی از گروه را با پراکندگی دیگری از همان گروه مقایسه کند. مانند دو ویژگی قد و وزن در یک گروه می‌تواند از ضریب پراکندگی (v) استفاده کند. مثلاً آیا پراکندگی قد یک گروه همانند پراکندگی وزن آنهاست. گاهی اوقات به ضریب پراکندگی، ضریب نسبی واریانس نیز گفته می‌شود. در محاسبه ضریب پراکندگی از میانگین و انحراف استاندارد استفاده می‌شود.

فرمول محاسبه ضریب پراکندگی:

$$v = \left(\frac{S}{\bar{X}} \right) (100)$$

تفسیر انحراف استاندارد

مقدار انحراف استاندارد به تناسب تغییر در پراکندگی نمرات تغییر می‌کند؛ به عبارتی اگر توزیع نمرات نسبت به میانگین پراکندگی زیادی داشته باشند انحراف استاندارد بزرگتر خواهد شد. در مقابل اگر توزیع نمرات نسبت به

میانگین پراکندگی کمی داشته باشند و در اطراف میانگین انباشته شوند واریانس و انحراف استاندارد کوچکتر خواهد شد. باید توجه داشته باشیم که بزرگ یا کوچک بودن انحراف استاندارد نسبی است و زمانی بکار می رود که قصد محاسبه انحراف استاندارد مرکب را داشته باشیم.

گشتاورهای پیرامون میانگین

میانگین و انحراف استاندارد به دسته‌ای شاخص‌های آمار توصیفی که گشتاور نامیده می‌شود، تعلق دارند. میانگین انحراف هریک از نمره‌ها را از میانگین مطابق فرمول زیر گشتاور اول می‌نامند:

$$m_1 = \frac{\sum (X - \bar{X})}{N} = 0$$

گشتاور رتبه اول صفر است، گشتاور رتبه دوم $\frac{n-1}{n}$ برابر واریانس نمونه است، گشتاور رتبه سوم برای محاسبه اندازه کجی و گشتاور رتبه چهارم برای محاسبه اندازه کشیدگی بکار می‌رود. یکی از موارد استفاده شاخص‌های پراکندگی مانند انحراف چارکی تعیین کجی توزیع است. کجی یعنی انحراف منحنی از حالت تقارن چون که در حالت تقارن کجی نداریم یا صفر است.

• در توزیع متقارن فاصله بین چارک اول تا میانه با فاصله چارک سوم تا میانه مساوی است.

• منحنی متقارن است: $Q_3 - Q_2 = Q_2 - Q_1$

• کجی مثبت: $Q_3 - Q_2 > Q_2 - Q_1$

• کجی منفی است: $Q_3 - Q_2 < Q_2 - Q_1$

مقدار کجی را نیز می‌توان از فرمول زیر محاسبه کرد:

$$Sk = \frac{m_3}{m_2 \sqrt{m_2}} = \frac{\frac{\sum x^3}{n}}{\left(\sqrt{\frac{\sum x^2}{n}}\right)^3}$$

یا از طریق فرمول کجی پیرسون

$$g_1 = \frac{\bar{X} - m_0}{S}$$

کشیدگی را نیز می‌توان از فرمول زیر محاسبه کرد

$$k\mu(g_r) = \frac{m_r}{(m_r)^r} - r = \frac{\frac{\sum x^r}{n}}{\left(\frac{\sum x^r}{n}\right)^r} - r$$

• اگر مقدار کشیدگی صفر باشد توزیع نرمال است.

- اگر مقدار کشیدگی مثبت باشد برآمدگی توزیع نرمال در نقطه اوج است.
- اگر مقدار کشیدگی منفی باشد منحنی توزیع نرمال در نقطه اوج خوابیده است.

نمره‌های استاندارد

نمره‌های استاندارد برای تعیین موقعیت نسبی هریک از اعداد در توزیع نمرات مربوط به آنها بکار می‌رود. به عبارتی نمره‌های استاندارد برحسب شاخص مرکزی میانگین و شاخص پراکندگی انحراف استاندارد موقعیت و محل نمره‌ها را در یک توزیع اعداد نشان می‌دهد. با استفاده از نمرات معیار موقعیت نسبی دانش‌آموزان را در یک گروه با نشان دادن فاصله نمره او از میانگین نمرات نشان می‌دهند. می‌توان افراد مختلف را از لحاظ عملکرد با یکدیگر مقایسه کرد. رایجترین نمره‌های استاندارد عبارتند از: نمره سیگمایی (Z)، نمره تراز شده (T)، نمره دانشگاهی (zcb)، نمره هوشبهر (IQ) و نمره‌های نه بخشی (stenine).

ویژگی‌های نمرات استاندارد

- | | | |
|---|---|---|
| <ul style="list-style-type: none"> • نمره‌های استاندارد وضعیت افراد یا نمره‌ها را نسبت به میانگین تعیین می‌کنند. • نمره‌های استاندارد با مقیاس فاصله‌ای بکار می‌روند. • نمره‌های استاندارد مستقیماً از نمرات خام بدست می‌آیند. • انجام محاسبات ریاضی در نمره‌های استاندارد ممکن است به دلیل فاصله‌ای بودن مقیاس مورد استفاده باشند. | } | رتبه‌های درصدی فاقد این دو ویژگی هستند. |
|---|---|---|

- نمره‌های استاندارد مشخص می‌کنند که یک نمره، چند انحراف استاندارد بالا یا پایین میانگین است.

نمره Z

نمره Z یک نمره استاندارد بنیادی است. نمره‌ها در اثر انتقال خطی نمره‌های خام بدست می‌آید. در انتقال خطی شکل توزیع نمره‌های استاندارد دقیقاً همانند شکل نمره‌های اصلی است. این دسته از نمرات دارای توزیع طبیعی با میانگین صفر و انحراف معیار یک خواهند بود. در واقع نمره Z از تقسیم انحراف نمره خام از میانگین بر انحراف استاندارد بدست می‌آید. فرمول محاسبه نمره Z عبارت است از:

$$Z = \frac{X - \mu_x}{\sigma} = \frac{x}{\sigma}$$

μ_x = میانگین نمرات

σ = انحراف استاندارد توزیع

x = انحراف نمره‌ها از میانگین

برای محاسبه نمره Z اول اعداد خام را از میانگین کم کرده و حاصل را بر انحراف استاندارد تقسیم می‌کنیم. برای محاسبه نمره Z میانگین مبداء و انحراف استاندارد واحد تعیین میزان اختلاف از میانگین است. نمره‌های Z نمرات مجردی محسوب می‌شوند که واحد اندازه‌گیری آنها با واحد اصلی اندازه‌گیری متفاوت است. نمره Z به ما نشان می‌دهد که یک نمره خام چقدر بالاتر یا پایین‌تر از میانگین است و نمرات بالاتر از میانگین دارای نمره معیار مثبت و نمرات پایین‌تر از میانگین دارای نمره منفی است. به علت منفی بودن و اعشاری بودن Z معیار دیگری مانند T استفاده می‌شود تا این محدودیت‌های نمره Z رفع شود.

مثال:

در نظر بگیرید در یک توزیع نمره یا اعداد که میانگین مساوی ۴۵ و انحراف استاندارد مساوی ۵ است نمره شما ۶۳ بوده است. نمره Z برابر خواهد بود با:

$$Z = \frac{X - \mu_x}{\sigma} = \frac{x}{\sigma} = \frac{63 - 45}{5} = 3/6$$

عدد یا نمره ۳/۶ بیانگر این است که نمره شما تقریباً سه و نیم انحراف استاندارد بالاتر از میانگین قرار دارد. اگر در مثال فوق محمد دوست شما نمره ۳۵ گرفته باشد خواهیم داشت:

$$Z = \frac{X - \mu_x}{\sigma} = \frac{x}{\sigma} = \frac{35 - 45}{5} = -2$$

نمره -۲ بیانگر این است که نمره محمد دو انحراف استاندارد پایینتر از میانگین قرار دارد.

اشکال دیگر نمره Z

آزمون‌های هوشی، نمره خام را به توزیعی با میانگین ۱۰۰ و انحراف استاندارد ۱۵ تبدیل می‌کنند. مانند فرمول زیر:

$$IQ = 15 \left(\frac{X - \mu}{\sigma} \right) + 100$$

$$IQ = 15Z + 100$$

نمرات تافل مثال دیگری در این مورد با میانگین ۵۰۰ و انحراف استاندارد ۱۰۰ است. در این آزمون نمره دارای ۱/۵ انحراف استاندارد بالاتر از میانگین مساوی:

$$(1/5)(100) + 500 = 650$$

ویژگی‌های نمره Z

- میانگین نمرات Z صفر و انحراف استاندارد آنها یک است.
- انتقال نمرها از خام به Z شکل توزیع نمرات را تغییر نمی‌دهد.

- نسبت اختلاف نمره‌ها در توزیع اصلی مساوی نسبت اختلاف بین نمره‌های Z آنها است.
- با مقیاس فاصله‌ای مورد استفاده قرار می‌گیرد.
- واحد اندازه‌گیری انحراف استاندارد است.
- نمره‌های بالای میانگین +Z و نمره‌های پایین میانگین -Z است.
- تبدیل نمره‌های خام به نمره Z مقایسه نمرات را سهل‌تر می‌کند.
- به سادگی برای کلیه افراد جامعه قابل محاسبه نیست. (عیب)
- مستلزم آشنایی با میانگین و انحراف استاندارد است. (عیب)
- داشتن اعشار و علامت منفی موجب دشواری فهم نمره Z می‌شود. (عیب)

نمره T

بیان شد که گاهی نمره‌های Z اعشاری و یا منفی هستند برای رفع این مشکل از نمره‌های T استفاده می‌شود. نمره T دارای میانگین ۵۰ و انحراف استاندارد ۱۰ است. و بر خلاف نمره Z همیشه مثبت بوده و دامنه آن بین ۲۰ تا ۸۰ می‌باشد. برای محاسبه T از فرمول زیر استفاده می‌شود:

$$T = 10Z + 50$$

ویژگی‌های نمره T

- معمول‌ترین و ساده‌ترین روش تبدیل نمره Z به توزیعی با میانگین ۵۰ و انحراف استاندارد ۱۰ است.
- نمره T همیشه مثبت است.
- نمره T دامنه بین ۲۰ تا ۸۰ است.

مثال: در مثال نمره Z «در نظر بگیرید در یک توزیع نمره یا اعداد که میانگین مساوی ۴۵ و انحراف استاندارد مساوی ۵ است نمره شما ۶۳ بوده است. نمره Z برابر خواهد بود با:» اگر $Z = 3$ باشد. برای تبدیل مثال فوق به نمره T نیاز خواهیم داشت:

$$T = 10Z + 50 = 10(3) + 50 = 80$$

تفسیر رابطه فوق بدین گونه خواهد بود که عدد ۸۰ بیانگر این است نمره شما ۳ انحراف استاندارد بالاتر از میانگین است.

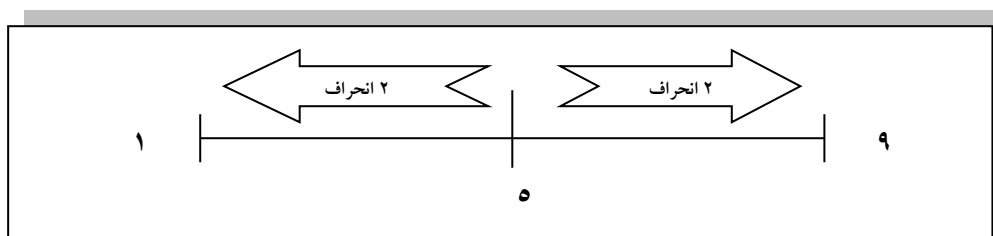
نمرات نه گانه یا نه بخشی

نمرات نه گانه نمرات استandar دی هستند که دامنه آنها بین ۱ تا ۹ است. این نمرات به صورت نمرات معیار یک رقمی نشان داده می‌شود. علت نام‌گذاری این است که توزیع این نمرات به نه نمره محدود می‌شود. نمرات نه گانه

دارای میانگین ۵ و انحراف معیار ۲ است. نمرات نه گانه را برای نمره‌گذاری آزمون‌های کلی و تشریحی که ارزشیابی آنها جنبه کیفی دارد به ویژه آزمون‌های فهمیدن و خواندن به کار می‌برند. همچنین از این آزمون برای تصحیح امتحانات شفاهی استفاده می‌شود. فرمول آن به صورت زیر است:

$$2Z + 5 \text{ یا } 2 \frac{X - \mu}{\sigma} + 5$$

کوچکترین نمره در این شاخص ۱ است که ۲ انحراف استاندارد پایین تر از میانگین و بزرگترین نمره ۹ است که ۲ انحراف استاندارد بالاتر از میانگین است.



رتبه درصدی

رتبه درصدی، رتبه نسبی یک نمره از توزیع را بر مبنای مقیاس ۱۰۰ بیان می‌کند؛ به عبارتی به ما می‌گوید که چند درصد نمره‌ها در توزیع پایین‌تر از نمره خاصی قرار دارد. برای مثال رتبه درصدی ۹۱ بیانگر این است که ۹۱ درصد نمره‌ها پایین‌تر از نمره کسی است که ۹۱ گرفته است. در محاسبه این شاخص باید توجه داشت که عملکرد فرد در گروه به تنهایی مقایسه می‌شود. برای مثال اگر دانش آموزی نمره ۱۷ در درس علوم بگیرد ممکن است احساس شود که نمره بالایی گرفته است اما مطابق شاخص رتبه درصدی نمره ۱۷ زمانی نمره بالایی خواهد بود که: ۱- همه نمرات بالای ۱۷ نباشند، و ۲- ۱۷ در مقایسه با دیگران بالا باشد.

محاسبه رتبه درصدی

۱. مرتب ساختن نمره‌ها از بزرگ به کوچک
۲. مشخص ساختن فراوانی تراکمی هر عدد
۳. مشخص کردن عددی که می‌خواهیم رتبه آن را پیدا کنیم
۴. مشخص کردن فراوانی تراکمی عددی که می‌خواهیم رتبه آن را پیدا کنیم
۵. تقسیم فراوانی تراکمی (عددی که می‌خواهیم رتبه آن را پیدا بکنیم) تقسیم بر مجموع نمرات یا N
۶. عدد بدست آمده در بند ۵ نمایانگر رتبه درصدی عددی است که می‌خواستیم رتبه آن را پیدا کنیم.

مثال در یک توزیع می‌خواهیم رتبه درصدی عدد ۵۹ را پیدا بکنیم که $N=30$ و $cf=18$ است. پس خواهیم داشت:

$$\frac{18}{30} = 0.6 \text{ یا } 60\% \text{ درصد نمرات بالاتر است. یعنی نمره ۵۹ از ۶۰}$$

$$PR = \frac{cf}{n} (100) \text{ فرمول محاسبه رتبه درصدی:}$$

ویژگی‌های شاخص رتبه درصدی

- رتبه درصدی یک شاخص رتبه‌ای است.
- محل نمره را بر مبنای رتبه آن در توزیع نشان می‌دهد.
- با استفاده از مقیاس ۱۰۰ بدون استفاده از تعداد نمره‌های توزیع موقعیت نمره را نشان می‌دهد.
- دارای معنای کلی هستند.
- نمره‌ای که رتبه درصدی آن نزدیک صفر است در گروه نمره پایینی است.
- نمره‌ای که رتبه درصدی آن نزدیک ۵۰ است در گروه نزدیک میانگین است.
- رتبه درصدی نمره یا رتبه فرد را در گروه مشخص می‌کند نه بطور مطلق.
- امکان دست یافتن به رتبه درصدی ۹۹/۹ وجود دارد لکن دسترسی به رتبه درصدی ۱۰۰ غیرممکن است.
- بالاترین نمره در توزیع هم ممکن است رتبه درصدی ۹۹ کسب بکند لکن رتبه ۱۰۰ بدست نخواهد آورد.

نقاط درصدی

نقاط درصدی یک توزیع، نقاطی هستند که معادل رتبه‌های درصدی هستند. برای مثال میانه نقطه ۵۰ درصدی توزیع است، چارک اول نقطه ۲۵ درصدی توزیع است و چارک سوم نقطه ۷۵ درصدی توزیع است. پس می‌توان گفت:

$$P_{75} = \text{چارک سوم} = P_{25} = \text{چارک اول} = P_{50} = \text{میانه}$$

$$P_x = L + \left(\frac{P_n - cf}{f_i} \right) i \text{ فرمول محاسبه:}$$

در شرایطی که نقطه درصدی بین دو یا چند نقطه واقع شود، نقطه درصدی مساوی با میانگین حد بالای پایین‌ترین و بالاترین نقطه است.

گفتیم که نقطه درصدی نقطه‌ای در مقیاس نمره‌گذاری پیوسته است که درصد معینی از افراد زیر آن قرار می‌گیرند. مثلاً نقطه ۸۰ درصدی، نقطه‌ای است که نمره‌های ۸۰ درصد افراد کمتر از آن است و نمره ۲۰ درصد از افراد بالاتر از آن می‌باشد. فرمول کلی محاسبه صدکها و دهکها به صورت زیر است.

در بالا گفته شد که چارک‌های مختلف چه نقاط درصدی را بیان می‌کنند. اما در دهک‌ها نقاط درصدی به ده بخش تقسیم می‌شوند. که با D مشخص می‌شوند؛ یعنی خواهیم داشت.

$$D_1 = 10 - \text{درصد نمرات} = D_2 = 20 - \text{درصد نمرات} = D_3 = 30 - \text{درصد نمرات} \dots\dots\dots$$

$$D_5 = P_{50} = Q_2 \quad D_1 = P_{10}$$

ویژگی‌های رتبه درصدی: معایب و مزیت‌ها

- رتبه درصدی یک شاخص ترتیبی است.
- رتبه درصدی رتبه فرد یا عدد را نسبت به افراد یا اعداد دیگر نشان می‌دهد.
- به کمک رتبه درصدی نمی‌توان مشخص کرد نمره یا رتبه فرد تا چه اندازه از رتبه یا نمره افراد دیگر بهتر است.
- تغییر جزئی در نمره‌های درصدی خیلی بالا یا پایین، موجب اختلاف در نمره‌های خام می‌شود.

همبستگی

ارتباط بین دو یا چند متغیر را همبستگی گویند. در روش همبستگی رابطه میان متغیرها براساس هدف تحلیل می‌گردد. زمانی که افزایش در یک متغیر با افزایش در متغیر دیگر یا کاهش یک متغیر با کاهش متغیر دیگر همراه باشد، همبستگی بین دو متغیر مستقیم و مثبت است. اما اگر افزایش در یک متغیر با کاهش متغیر دیگر همراه باشد همبستگی بین دو متغیر منفی و معکوس است. همبستگی بین دو متغیر به معنی رابطه علت و معلولی بین دو متغیر نیست.

ضریب همبستگی

ضریب همبستگی شاخص آماری است که میزان و حدود رابطه بین متغیرها را نشان می‌دهد. شاخص عددی یک بالاترین ضریبی است که امکان دارد برای همبستگی بدست بیاید و بیانگر همبستگی کامل بین متغیرهای مورد مطالعه می‌باشد. برای ضریب همبستگی داریم:

$$\text{همبستگی کامل مثبت: } r = 1$$

$$\text{همبستگی کامل منفی: } r = -1$$

همبستگی صفر: عدم همبستگی بین متغیرها

تفاوت دو ضریب همبستگی کامل منفی و مثبت در جهت رابطه است و شدت همبستگی به وسیله قدر مطلق ضریب همبستگی مشخص می‌شود نه به وسیله علامت آن. برای همین هم $r = -85$ مساوی با $r = +85$ است.

تحقیقات همبستگی بر حسب هدف به سه دسته تقسیم می‌شوند:

۱- همبستگی دو متغیری

۲- تحلیل رگرسیون

۳- تحلیل ماتریس همبستگی یا کواریانس

تحقیق همبستگی دو متغیری: مقیاس فاصله‌ای

در این نوع مطالعات هدف تعیین میزان هماهنگی تغییرات دو متغیر است. در غالب پژوهش‌های همبستگی دو متغیری از مقیاس فاصله‌ای استفاده می‌شود و ضریب همبستگی محاسبه شده در آنها از نوع ضریب همبستگی گشتاوری پیرسون است.

محاسبه ضریب همبستگی پیرسون: مقیاس فاصله‌ای - نسبی

معروف‌ترین و پرکاربردترین شاخص ضریب همبستگی است. این ضریب زمانی بکار می‌رود که داده‌ها با مقیاس فاصله‌ای یا نسبی اندازه‌گیری شده باشند. فرمول محاسبه ضریب همبستگی پیرسون:

$$Y_{xy} = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$

X = نمره آزمون در متغیر اول

h = تعداد آزمودنیها

Y_{xy} = ضریب همبستگی

\sum = مجموع

y = نمره آزمودنی در متغیر دوم

روش دوم محاسبه ضریب همبستگی که ساده‌تر هم هست و در آن از میانگین‌های انحراف نمره‌های x و y

استفاده می‌شود. فرمول روش:

$$Y = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} \quad Y_{xy} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

انحراف نمره‌ها از میانگین $y =$

انحراف هر نمره از میانگین $x =$

ضریب همبستگی $y =$

مفروضات ضریب همبستگی پیرسون

- رابطه بین دو متغیر خطی باشد.
- توزیعها دارای شکل مشابه باشند.
- نمودار پراکندگی یکسان باشد.

ضریب همبستگی رتبه‌ای اسپیرمن

این ضریب برای توصیف درجه همبستگی دو متغیر رتبه‌ای مناسب است و زمانی بکار می‌رود که نمره‌ها رتبه‌بندی شده باشند و یا بجای اعداد رتبه‌های آنها در دست باشد. فرمول محاسبه آن عبارت است:

$$Y_s = 1 - \frac{6 \sum d^2}{n(n-1)(n+1)}$$

ضریب همبستگی رتبه‌ای اسپیرمن یا

$d =$ اختلاف بین یک جفت رتبه

$\sum =$ مجموع

$n =$ تعداد سوال‌ها

ضریب همبستگی رتبه‌ای کندال

این روش زمانی مورد استفاده قرار می‌گیرد که متغیرهای مورد مطالعه با استفاده از مقیاس رتبه‌ای، یا بالاتر اندازه‌گیری شده باشند. ضریب همبستگی کندال t می‌تواند جانشین ضریب همبستگی اسپیرمن شود.

ضریب تعیین

ضریب تعیین $(Y^2 \times 100)$ مشخص می‌کند چند درصد از کل واریانس X ناشی از واریانس Y است. برای مثال اگر همبستگی بین متغیرهای A و B مساوی ۱ باشد ضریب تعیین برابر خواهد بود با:

$$\text{ارتباط مثبت کامل} \rightarrow \text{برابر با } 100 = (100)^2 (1)$$

نکته: ضریب تعیین هیچ وقت منفی نخواهد بود زیرا برای محاسبه آن، همبستگی بین متغیرها مجذور می‌شود،

ضریب همبستگی (-1) یک رابطه کامل معکوس است که افزایش یکی با کاهش دیگری همراه است.

تفسیر ضریب همبستگی

همبستگی تغییر در یک متغیر هماهنگ با تغییر در متغیر دیگر را بیان می‌کند. و ضریب همبستگی جهت و شدت رابطه بین دو متغیر را تعیین می‌کند. در تفسیر ضریب همبستگی باید به دو مسئله توجه داشت:

۱- تفسیر ضریب همبستگی نباید برحسب درصد باشد و نمی‌توان آن را به صورت نسبت مورد مقایسه و تفسیر قرار داد. برای مثال نمی‌توان گفت ضریب همبستگی ۹۰ درصد، ۹۰ درصد رابطه بین دو متغیر را بیان می‌کند و همچنین نمی‌توان گفت که ضریب فوق دقیقاً مساوی ضریب همبستگی ۴۵ است.

۲- همبستگی بین چند متغیر الزاماً بیانگر رابطه علت و معلولی بین متغیرها نمی‌باشد.

ضریب تعیین

ضریب همبستگی، اندازه همبستگی بین متغیرها را نشان می‌دهد و وجود همبستگی یا بالا و پایین بودن نسبی آن را نشان می‌دهد. لکن ضریب تعیین مشخص می‌کند که چند درصد از کل واریانس x ناشی از واریانس y است.

$$V = (r_{xy})^2 (100) = 0$$

برای مثال اگر ضریب همبستگی بین متغیرهای x و y مساوی صفر باشد ضریب تعیین مساوی صفر خواهد بود.

$$V = (r_{xy})^2 (100) = (0)^2 \times 100 = 0$$

در تعلیم و تربیت ضریب تعیین به عنوان یک شاخص تعمیمی یا تکوینی بین مهارت‌ها مورد استفاده قرار می‌گیرد. توجه داشته باشید که ضریب تعیین هیچگاه منفی نخواهد شد زیرا برای محاسبه آن ضریب همبستگی مجذور می‌شود.

نکات مهم و کلیدی

- برای محاسبه ضریب همبستگی بین نمرات یک سوال چندگزینه‌ای و نمرات کل آزمون روش دو رشته‌ای نقطه تعیین همبستگی بهترین است.
- متغیر نمادی است که اعداد یا ارزشها به آن منتسب می‌شود. بعبارت دیگر متغیر عبارت از ویژگی واحد موردمشاهده است.
- متداول‌ترین روش تعیین همبستگی، روش محاسبه ضریب همبستگی گشتاوری پیرسون است که این ضریب همبستگی با r نشان داده می‌شود.
- واریانس توزیع یک دسته نمره معرف شاخص پراکندگی است.
- هرگاه بخواهیم بین سؤالات آزمون همبستگی را محاسبه کنیم، تراکوریک را استفاده می‌کنیم.
- اکثر ضریب های همبستگی مورد خاصی از ضریب همبستگی پیرسون هستند.
- زمانی از ضریب همبستگی دو رشته‌ای نقطه‌ای استفاده می‌کنیم که یک دسته نمرات فاصله‌ای یا نسبی و دسته دیگر اسمی باشد.
- ضریب همبستگی اسپیرمن را فقط برای نمرات مقیاس ترتیبی می‌توان به کار برد.
- ضریب همبستگی پیرسون متداول‌ترین ضریب همبستگی است.
- همبستگی بین نمرات دو متغیرمقدار $+1$ نشان دهنده رابطه کامل و مثبت دو متغیر و -1 نشان‌دهنده رابطه کامل منفی و صفر نشانه نبود رابطه است.
- وقتی نمره‌های افراد در یک توزیع با هم مساوی باشد پراکندگی به وجود نمی‌آید.
- وقتی نمره‌های افراد در یک توزیع پراکنده باشند، ناهمانگی ایجاد می‌شود.
- وقتی نمره‌های افراد در یک توزیع پراکندگی کمتری داشته باشد یکنواختی افزایش می‌یابد.
- ساده‌ترین شاخص پراکندگی دامنه تغییرات است با فرمول $R = X_H - X_L$
- ساده‌ترین روش برای تعیین جایگاه کلی نمره‌ها، شاخص معدل است.
- نما ساده‌ترین شاخص گرایش مرکزی است.
- اندازه یا حجم واحدهای اندازه‌گیری در میانه تأثیری ندارد.
- میانه را می‌توان با مقیاس‌های فاصله‌ای و نسبی هم مورد استفاده قرار داد اما مقیاس‌های فاصله‌ای و نسبی در محاسبه میانه مورد استفاده قرار نمی‌گیرند.
- میانگین مشهورترین و معتبرترین شاخص گرایش مرکزی است.
- متداولترین شاخص مکان مرکزی که کاربرد گسترده‌ای دارد، میانگین است.

- میانگین و میانه دارای ثبات و حساسیت بیشتری نسبت به نما هستند.
- میانه یک شاخص ترتیبی است و تحت تاثیر ارزش عددی نمره‌ها قرار نمی‌گیرد در حالی که میانگین به ارزش عددی تمام نمره‌ها حساس است.
- در یک منحنی طبیعی در صورتی که توزیع طبیعی، متقارن و تک‌نمایی باشد میانگین، میانه و نما بر روی هم قرار می‌گیرند.
- در یک توزیع دو‌نمایی متقارن، میانگین و میانه برابر ولی نما در طرفین قرار دارد.
- در توزیع‌های دارای کجی مثبت میانگین همیشه بزرگتر از میانه است و نما کوچکتر از میانه است.
- توزیع‌های دارای کجی مثبت = میانگین < میانه < نما یا $M < Md < \bar{x}$
- در توزیع‌های دارای کجی منفی میانگین همیشه کوچکتر از میانه است و نما بزرگتر از میانه است.
- توزیع‌های دارای کجی منفی = میانگین > میانه > نما یا $M > Md > \bar{x}$
- میانگین پایدارترین و معتبرترین شاخص مرکزی است.
- برای استنباط جامعه از نمونه میانگین معتبرترین برآورد از میانگین جامعه است.
- ساده‌ترین شاخص پراکندگی دامنه تغییرات است: دامنه تغییرات = کوچکترین عدد - بزرگترین عدد
- چارک دوم (Q_2) برابر میانه است، یعنی نقطه‌ای که توزیع را به دو قسمت مساوی تقسیم می‌کند.
- چارک سوم (Q_3) نقطه‌ای است که $\frac{3}{4}$ نمره‌ها در زیر آن و بقیه در بالای آن واقع شده‌اند. به چارک سوم نقطه ۷۵ درصدی نیز گفته می‌شود.
- واریانس با مقیاس فاصله‌ای و نسبی بکار می‌رود.
- بزرگترین عیب میانگین این است که به نمرات خیلی کوچک و نمرات خیلی بزرگ حساس است.
- از بزرگی و کوچکی مقدار واریانس و انحراف معیار به زیادی و کمی پراکندگی نمرات پی می‌بریم.
- هر چه گروه نامتجانس‌تر باشد ضریب همبستگی حاصل بزرگتر است.
- مجذور ضریب همبستگی نشان دهنده درصد وجه مشترک دو متغیر است.
- تنها تفاوت بین ضریب همبستگی مثبت و منفی جهت رابطه بین متغیرهایست.
- در همبستگی فی هر دو متغیر باید دو ارزشی واقعی باشند.
- ضریب همبستگی تتراکوریک بین جفت جفت روابط سوالهای یک آزمون را محاسبه می‌کند.
- تتراکوریک یک برآورد سریع از همبستگی گشتاوری پیرسون است.
- مقادیر ضریب همبستگی دو رشته‌ای گاهی از +۱ بزرگتر و گاهی از -۱ کوچکتر است.
- ضریب همبستگی تتراکوریک و ضریب همبستگی و رشته‌ای از خانواده ضریب همبستگی گشتاوری پیرسون نیستند.

فصل دوازدهم: هنجارها و نیمرخها

هنجارها و نیمرخها

پس از تصحیح برگه‌های آزمون برای هریک از دانش‌آموزان نمره‌ای به نام نمره خام بدست می‌آید. نمرات خام به خودی خود قابل تفسیر و استفاده نیستند و باید در رابطه با آنها اقداماتی انجام داد تا قابل استفاده گردند. یکی از روش‌هایی که می‌توان به کار برد استفاده از هنجارها و نیمرخها برای این موضوع می‌باشد. هنجارها و نیمرخها روش‌هایی هستند که به ما کمک می‌کنند تا سطح موفقیت دانش‌آموزان را در درس‌های مختلف با هم مقایسه کنیم و میزان توفیق آنها را نسبت به یکدیگر بسنجیم.

لازم به یادآوری است که مقایسه نمرات دانش‌آموزان با یکدیگر مخصوص آزمون‌های وابسته به هنجارهاست نه آزمون‌های وابسته به ملاک بنابراین مطالبی که در اینجا می‌آوریم عمدتاً ویژه آزمون‌های وابسته به هنجارها هستند نه آزمون‌های وابسته به ملاک.

تعریف هنجار

اگر برای تفسیر نمره یک دانش‌آموز آن نمره را با نمره یک گروه مرجع مقایسه کنیم به آن گروه مرجع گروه هنجارها می‌گویند. گروه مرجع از کسانی تشکیل می‌یابد که به گونه‌ای شبیه دانش‌آموز مورد نظر هستند. بنابراین فایده هنجار این است که برای قضاوت درباره سطح نسبی عملکرد یک فرد به عنوان معیار به کار می‌رود. هنجارها یکی از مراحل اصلی در استاندارد کردن آزمون‌های وابسته به هنجار است. برای تعیین هنجار پس از انتخاب گروه هنجار نمرات گروه را در آزمون مورد نظر تعیین می‌کنیم و یکی از شاخص‌های مرکزی (معمولاً میانه) را برای نمرات آن آزمون مورد استفاده در هنجاریابی هستند و این نمرات صرفاً عملکرد بهنجار یا نوعی را گزارش می‌کنند نه عملکرد بسیار خوب و مطلوب را. گراندولین در رابطه با نمرات مشتق یا هنجارهای آزمون و امکان انجام مقایسه‌های فوق در میان نمرات آزمون چند فایده را ذکر می‌کند:

- (۱) پیش‌بینی موفقیت احتمالی دانش‌آموز در زمینه‌های مختلف
- (۲) تشخیص نقاط قوت و ضعف عملکرد دانش‌آموز
- (۳) اندازه‌گیری رشد آموزش

انواع هنجارها

نمرات هنجار که به آنها نمرات مشتق نیز گفته می‌شود دارای انواع مختلفی هستند که معروف‌ترین آنها عبارتند از:

- ۱- هنجارهای سنی یا معادل‌های سنی
- ۲- هنجارهای کلاسی یا معادل‌های کلاسی
- ۳- هنجارهای درصدی
- ۴- نمرات معیار (استاندارد)

الف) هنجار سنی

هنجار سنی بر میانگین یا میانه نمرات معینی که دانش‌آموزان در سنین مختلف کسب می‌کنند و به صورت معادل‌های سنی نشان داده می‌شود دلالت دارد. برای مثال اگر دانش‌آموزانی که ۱۰ سال و دو ماه سن دارند در یک آزمون نمره ۱۸ بگیرند به این نمره معادل سنی ۲ - ۱۰ تعلق می‌گیرد. از محدودیت‌های هنجار سنی این است که در آن عملکرد آزمون در قالب واحدهای نامساوی معرفی می‌شود. مثلاً تغییرات الگوی رشد در سنین مختلف متفاوت است. به رغم محدودیت بالا استفاده از هنجار سنی در سطح دبستان بهترین مرحله کاربرد این نوع هنجار است. زیرا

رشد در این سنین منظم و یکنواخت است. در سطح متوسطه نیز هنجار سنی کاربرد دارد. اما برای سطح بالاتر استفاده از هنجار سنی خیلی مفید نیست. از هنجار سنی برای آزمون‌های هوش و استعداد استفاده می‌شود. از محدودیت‌های هنجار سنی این است که در آن عملکرد آزمون در قالب واحدهای نامساوی معرفی می‌شود.

ب) هنجار کلاسی

هنجار کلاسی شبیه به هنجار سنی است. تفاوت عمده این دو نوع هنجار این است که در هنجار سنی گروه‌های مرجع گروه‌های سنی هستند اما در هنجار کلاسی از گروه‌های کلاسی به عنوان گروه‌های مرجع استفاده می‌کنند. معادل‌های کلاسی را به صورت دو رقمی نشان می‌دهند. رقم اول نشان‌دهنده سال و رقم دوم نشان‌دهنده ماه است که از صفر تا ۹ را شامل می‌شود. علت آنکه رقم دوم فقط تا ۹ را شامل می‌شود این است که سال تحصیلی فقط نه ماه است.

هنجار سنی نمایی

معمولاً دانش‌آموزان یک کلاس معین که براساس نمرات آنها هنجارهای کلاسی تعیین می‌شود دامنه سنی بزرگی را تشکیل می‌دهند. لذا نمرات بعضی از دانش‌آموزان که از لحاظ سنی خیلی بزرگتر یا خیلی کوچکتر از دانش‌آموزان متوسط هستند. نیز در گروه هنجاریابی شرکت داده می‌شوند. در این صورت نیاز به یک هنجار محدود شده‌ای است که به آن هنجار سنی نمایی گفته می‌شود.

هنجار سنی ذهنی

هنجار سنی ذهن یا سن عقلی که اولین بار توسط بنیه معرفی شد، نوعی هنجار سنی است که در آزمون‌های هوش به کار می‌رود. روش محاسبه سن ذهنی شبیه محاسبه معادل‌های سنی است. بنابراین اگر نمره خام یک دانش‌آموز در یک آزمون هوش برابر با میانه نمرات خام کودکان ۹ ساله باشد، سن ذهنی او ۹ خواهد بود.

هوشبهر

هوشبهر یا بهره هوشی (IQ) از فرمول زیر بدست می‌آید:

$$IQ = \frac{\text{سن ذهنی}}{\text{سن تقویمی}} \times 100$$

هنجار درصدی

در هنجار درصدی فرد یا گروه سنی با کلاس خودش یعنی گروهی که منطقاً او را می‌توان عضوی از آن دانست مقایسه می‌شود. در هنجارهای درصدی از رتبه‌های درصدی یا صدک‌ها استفاده می‌شود. رتبه درصدی یا نمره درصدی وضعیت نسبی فرد در گروه را برحسب کسانی که نمره پایین‌تر از او گرفته‌اند مشخص می‌کند در اینجا نیز هر نمره خام یک معادل درصدی دارد.

می‌توان صدک را نوعی رتبه در یک گروه هزار نفری تلقی کرد مثلاً اگر نمره خام دانش‌آموز ۱۶ و صدک او هفتادم باشد، یعنی نمره دانش‌آموز از ۷۰ درصد دانش‌آموزان بهتر است.

صدک پنجاهم را با نشان P_{50} می‌دهند و برابر با میانه است و روش محاسبه صدک‌ها شبیه به روش محاسبه میانه است. از معایب هنجارهای درصدی این است که واحدهای مربوط به صدک‌ها در طول مقیاس مساوی نیستند چون اکثر نمرات در وسط مقیاس قرار می‌گیرند.

هنجارهای مربوط به نمرات معیار

با استفاده از نمرات معیار که موقعیت نسبی یک دانش‌آموز را در یک گروه با نشان دادن فاصله نمره او از میانگین نمرات نشان می‌دهد. می‌توان افراد مختلف را از لحاظ عملکرد با یکدیگر مقایسه کرد.

نمره Z

این دسته از نمرات دارای توزیع طبیعی با میانگین صفر و انحراف معیار یک خواهند بود. برای تبدیل نمرات خام به نمره Z فرمول زیر استفاده می‌شود:

$$Z = \frac{X - \bar{X}}{sd}$$

x = نمره خام

\bar{X} = میانگین نمرات

SD = انحراف معیار نمرات

نمره Z به ما نشان می‌دهد که یک نمره خام چقدر بالاتر یا پایین‌تر از میانگین است و نمرات بالاتر از میانگین دارای نمره معیار مثبت و نمرات پایین‌تر از میانگین دارای نمره منفی است. به علت منفی بودن و اعشاری بودن Z معیار دیگری مانند T استفاده می‌شود تا این محدودیت‌های نمره رفع Z شود.

نمره T

نمره T برای رفع مشکلات نمرات ابداع شده و از این راه بدست می‌آید که میانگین ۵۰ و انحراف معیار ۱۰

می‌باشد. برای محاسبه T از فرمول زیر استفاده می‌شود:

$$T = 10Z + 50$$

نمرات SAT

نمرات این گروه مرجع به صورت بهنجار با میانگین ۵۰۰ و انحراف معیار ۱۰۰ می‌باشد و فرمول آن به صورت زیر است:

$$\text{نمرات ۹ بخشی} = 100Z + 500 = \text{نمره SAT}$$

این نمرات به صورت نمرات معیار یک رقمی نشان داده می‌شود. علت نام‌گذاری این است که توزیع این نمرات به ۹ بخش محدود می‌شود.

این مقیاس دارای میانگین ۵ و انحراف معیار ۲ است. نمرات ۹ بخشی را برای نمره‌گذاری آزمون‌های کلی و تشریحی که ارزشیابی آنها جنبه کیفی دارد به ویژه آزمون‌های فهمیدن و خواندن به کار می‌برند. همچنین از این آزمون برای تصحیح امتحانات شفاهی استفاده می‌شود. فرمول آن به صورت زیر است:

$$\text{Stanines} = 2z + 5$$

هوشبهر انحرافی

یک هنجار مربوط به نمرات معیار است و دارای میانگین ۱۰۰ و انحراف معیار ۱۵ یا ۱۶ است. فرمول آن به صورت زیر است و تفسیر آن مانند T و Z است.

$$\text{DIQ} = 15Z + 100$$

توزیع بهنجار و منحنی بهنجار

منظور از توزیع بهنجار، توزیعی است که بر منحنی بهنجار منطبق است و دارای ویژگی‌هایی است که متقارن است. اکثر نمرات در وسط توزیع جمع شده است. میانه و میانگین و نما بر روی هم قرار دارند. کل مساحت زیر منحنی برابر با یک واحد فرض می‌گردد.

نیمرخ

با استفاده از نیمرخ عملکرد یک دانش‌آموز را می‌توان در آزمون‌های مختلف مستقیماً با هم مقایسه کرد. به کمک نیمرخ که یک بازنمایی نگاره‌ای یا تصویری از نمرات آزمون هاست. می‌توان نقاط ضعف یک دانش‌آموز را مشخص کرد. (جهت مطالعه بیشتر به صفحات ۴۰۶ الی ۴۰۹ کتاب اندازه‌گیری، سنجش و ارزشیابی آموزش سیف ۱۳۸۵ نشر دوران مراجعه کنید.)

نکات مهم و کلیدی :

- نمرات T برای رفع مشکلات نمرات Z ابداع شدند.
- نمرات T نمراتی با میانگین ۱۰۰ و انحراف استاندارد ۵۰ هستند.
- نمرات SAT نمراتی با میانگین ۵۰۰ و انحراف استاندارد ۱۰۰ هستند.
- نمرات SAS نمراتی با میانگین ۱۰۰ و انحراف استاندارد ۱۶ هستند.
- نمرات NCE نمراتی با میانگین ۵۰ و انحراف استاندارد ۲۱/۰۶ هستند.
- نمرات Z نمراتی با میانگین ۰ و انحراف استاندارد ۱ هستند.
- نمرات نه بخشی نمراتی با میانگین ۵ و انحراف استاندارد ۲ هستند.
- نمرات DIQ نمراتی با میانگین ۱۰۰ و انحراف استاندارد ۱۵ هستند.
- هنجارهای درصدی دارای مقیاس رتبه ای هستند.
- نمرات NEC دارای دامنه ۱ تا ۹۹ هستند.
- در منحنی بهنجار میانگین، نما و میانه روی هم قرار دارند.

فصل سیزدهم:

روایی و پایایی آزمون

تعریف روایی

متخصصان سنجش برای آزمون‌های مختلف ویژگی‌های بسیاری را برمی‌شمارند که مهمترین آنها روایی و پایایی آزمون است. روایی یا اعتبار اصطلاحی است که به هدف آزمون برای تحقق بخشیدن به آن درست شده است. یعنی آزمونی دارای روایی است که برای اندازه‌گیری آنچه مورد نظر است مناسب باشد. مثلاً آزمون هوش فقط هوش را بسنجد. یک آزمون از جهاتی ممکن است روا باشد لکن از جهات دیگر روا نباشد، بنابراین روایی یک امر نسبی است نه یک ویژگی که هست یا نیست.

انواع روایی

با توجه به آزمون‌های مورد استفاده در روانشناسی و آموزش و پرورش ما با سه نوع روایی سروکار داریم:

۱. روایی محتوایی^۱

۲. روایی صوری^۲

۳. روایی ملاکی^۳

الف) روایی محتوایی

از آنجا که آزمون‌های پیشرفت تحصیلی در یک زمان محدود بکار می‌روند و از آنجا که هیچ آزمونی نمی‌تواند تمامی محتوا و هدف‌های آموزش را شامل شود پس یک آزمون از مجموعه‌ای سؤال است که تنها نمونه‌ای از تمامی محتوا و هدف‌های آموزش درس را تشکیل می‌دهد. در روایی محتوایی به این مطلب تأکید می‌شود که نمونه سؤال‌های مورد استفاده تا چه حد معرف جامعه سؤال‌های ممکن است که می‌توان از محتوا یا موضوع مورد نظر تهیه کرد. جدول مشخصات بهترین روش برای چنین چیزی است. برای تعیین روایی محتوایی روش آماری وجود ندارد. لذا از قضاوت متخصصان در این باره استفاده می‌شود.

روش تعیین روایی محتوایی

روش‌های تعیین روایی محتوایی به روش‌های قضاوتی معروف هستند. که در آنها خود ابزار اندازه‌گیری مورد قضاوت قرار می‌گیرد. هیچ‌گونه روش آماری برای تعیین ضریب روایی محتوایی وجود ندارد. در عوض برای تعیین روایی محتوایی یک آزمون از قضاوت متخصصان در این باره که سؤال‌های آزمون تا چه میزانی معرف محتوا و هدف‌های برنامه یا حوزه محتوایی است استفاده می‌شود.

نکته قابل توجه دیگر در مورد روایی محتوایی این است که قضاوت‌های مربوط به روایی محتوایی نه قطعی هستند و نه نهایی بلکه آزمون‌ها از سطوح مختلفی از روایی برخوردارند و متخصصان در قضاوت‌های خود همیشه با هم توافق نظر ندارند.

ب) روایی صوری

روایی صوری از مشتقات روایی محتوایی است و به این مطلب اشاره دارد که سؤالات آزمون تا چه حد در ظاهر شبیه به موضوع است که برای اندازه‌گیری آن تهیه شده است.

مثلاً آزمون ریاضی برای مهندسان هواشناسی باید مسائل هواشناسی را دربر گیرد. برای تعیین روایی صوری نیز از نظر متخصصان استفاده می‌شود.

در تعریف ساده تر روایی صوری به درجه‌ای از منطقی بودن یک آزمون به نظر کسانی که به آن جواب می‌دهند

^۱Content Validity

^۲face Validity

^۳Criterion Validity

گفته می‌شود.

در واقع روایی صوری نمی‌تواند نوعی روایی باشد بلکه تنها یک ویژگی آزمون است که در پاره‌ای مواقع وجود آن مفید به نظر می‌رسد.

هرچند که روایی صوری یک ویژگی مطلوب آزمون است با این حال در بعضی آزمون‌ها این ویژگی نه تنها ضروری نیست بلکه بهتر است که وجود نداشته باشد.

روش تعیین روایی صوری

روش‌های تعیین روایی صوری به روش‌های قضاوتی معروف هستند. از آنجا که روایی صوری نوع خاصی از روایی محتوایی است، همان روشی که برای تعیین روایی محتوایی به کار می‌رود، یعنی استفاده از نظر متخصصان روایی صوری هم مورد استفاده قرار می‌گیرند.

ج) روایی ملاکی

منظور میزان ارتباط بین نمرات حاصل از یک آزمون با نمرات حاصل از آزمون دیگر است. لذا هر زمان از نمرات یک آزمون برای پیش‌بینی عملکرد افراد در آزمونی دیگر استفاده کنیم، با روایی ملاکی سر و کار داریم. آزمون دوم که عملکرد فرد در آن پیش‌بینی می‌شود، آزمون پیش‌بینی می‌گویند و اگر مربوط به زمان حال و بدون فاصله زمانی باشد، به آن روایی همزمان می‌گویند. از روایی پیش‌بینی برای پیش‌بینی موفقیت افراد مورد تحصیلی یا شغلی استفاده می‌شود. مثال: استفاده از کنکور برای پیش‌بینی موفقیت دانش‌آموزان در دانشگاه.

روش تعیین روایی ملاکی

روش‌های تعیین روایی ملاکی به روش‌های تجربی معروف هستند. زیرا در آنها ماهیت پاسخ‌هایی که به سئوالهای آزمون داده می‌شود مورد بررسی قرار می‌گیرد.

روایی پیش‌بینی

برای تعیین روایی پیش‌بینی باید نمرات حاصل از دو آزمون ضریب همبستگی محاسبه گردد. هرچه ضریب بزرگتر باشد، روایی بیشتر خواهد بود.

روایی پیش‌بینی ویژه آزمون‌هایی است که برای پیش‌بینی موفقیت افراد در امور تحصیلی و شغلی به کار می‌روند.

روش تعیین روایی پیش‌بینی

از آنجا که در روایی پیش‌بینی همیشه نوعی ملاک مورد نظر است که نمرات آزمون پیش‌بینی با آن مقایسه شود به

آن روایی ملاکی نیز گفته می‌شود. روایی ملاکی به فن تجربی مطالعه رابطه بین نمرات آزمون پیش بینی و نوع اندازه بیرونی مستقل یعنی ملاک مربوط می‌شود.

در تعیین روایی پیش بینی از چندین ملاک استفاده می‌شود استفاده از چندین آزمون پیش بین نیز این نوع روایی را افزایش می‌دهد و بر قدرت پیش بینی می‌افزاید.

روایی همزمان

روایی همزمان نوعی روایی پیش بینی است که در آن بجای تعیین رابطه بین دو آزمون پس از گذشت یک فاصله زمانی، رابطه بین دو آزمون به طور همزمان تعیین می‌شود. هدف روایی همزمان بین دو آزمون این است که معلوم سازیم آیا می‌توان یکی از دو آزمون را به جای آزمون دیگر به کار برد یا نه؟ در روایی پیش‌بینی ملاک در آینده اندازه‌گیری می‌شود. هدف روایی پیش‌بینی، پیش‌بینی آینده است اما هدف روایی همزمان جانسین کردن آزمون پیش‌بینی با آزمون ملاک است. یک ملاک مطلوب دارای چهار ویژگی ربط داشتن بی‌طرفی پایایی و در دسترس بودن است.

روش تعیین روایی همزمان

روش‌های تعیین روایی همزمان به روش‌های تجربی معروف هستند.

روایی سازه

روایی سازه بیشتر از روایی محتوایی و روایی پیش‌بینی جنبه نظری دارد روایی سازه نوعی از روایی است که نمرات حاصل از اجرای آزمون به مفاهیم یا سازه‌های نظریه مورد نظر مربوط می‌شود. مانند سازه اضطراب.

روش تعیین روایی سازه

برای تعیین روایی سازه باید نخست سازه یا متغیر مورد نظر را تعریف کرد.

راههای تعیین روایی سازه

۱. تعیین همبستگی آزمون با سایر آزمون‌های روا
۲. تمایز سنی
۳. تحلیل عاملی
۴. همسانی درونی

تعیین همبستگی آزمون با سایر آزمون‌های روا

یکی از راه‌های تعیین روایی سازه تعیین همبستگی آزمون با سایر آزمون‌های موجود روا که همان سازه را اندازه می‌گیرد، است.

روایی همگرا و واگرا

هنگامی که از راه تعیین همبستگی آزمون با سایر آزمون‌های روایی موجود به مشخص کردن روایی سازه می‌پردازیم دو نوع روایی خواهیم داشت: ۱- اگر در تعیین روایی سازه به روش تعیین همبستگی آزمون با سایر آزمون‌های روایی موجود آزمون جدید با آزمون موجود همبستگی زیادی داشته باشد فرض می‌شود که دو آزمون سازه واحدی را اندازه می‌گیرند و در نتیجه آزمون جدید نیز برای سازه مورد نظر روا می‌باشد که به این نوع روایی، روایی همگرا گفته می‌شود و در بعد دیگر آزمون مورد نظر باید با آزمون‌های دیگری که سازه‌های متفاوتی را اندازه‌گیری می‌کنند همبستگی نداشته باشد، این گونه روایی را روایی واگرا یا روایی تمیزی می‌نامند.

تمایز سنی

تمایز سنی یکی دیگر از راه‌های تعیین روایی سازه است. برای مثال عملکرد کودکان خردسال در آزمون‌های هوشی مخصوص کودکان ضعیف است اما با بالا رفتن سن به سئوال‌های بیشتری جواب می‌دهند

تحلیل عاملی

روش تحلیل عاملی روابط درونی میان داده‌های مورد تحلیل را تعیین می‌کند و برای ایجاد سهولت متغیرها را به خوشه‌ها یا عوامل کاهش می‌دهد.

همسانی درونی

همسانی درونی ساخت درونی آزمون بررسی می‌شود ملاک مورد استفاده برای بررسی همسانی درونی نمره کل آزمون است. روش معمول برای این کار محاسبه همبستگی بین نمرات آزمون فرعی با نمره کل آزمون است.

روایی آزمون‌های وابسته به ملاک

اصطلاحات روایی محتوایی، پیش بین و سازه برای آزمون‌های وابسته به هنجار به کار می‌روند اما کاربرد آنها در مورد آزمون‌های وابسته به ملاک هم کاملاً امکان پذیر است. با توجه به مقاصد دو نوع آزمون شواهد مربوط به آنها متفاوت هستند.

روایی محتوایی آزمون‌های وابسته به ملاک

روایی محتوایی آزمون‌های وابسته به هنجار بسیار شبیه آزمون‌های وابسته به ملاک است اما در تعیین روایی برای آزمون‌های وابسته به ملاک وابستگی سئوال‌های آزمون به محتوا و هدف‌های مشخص یک موضوع بیشتر است. در

تعیین روایی آزمون‌های وابسته به هنجار چون هدف‌های دقیق موردنظر نیستند لذا کوشش آزمون ساز به نمونه‌گیری از محتوای مورد سنجش منحصر می‌شود.

تعیین روایی پیش‌بین آزمون‌های وابسته به ملاک

در آزمون‌های وابسته به ملاک روایی پیش‌بین به دو منظور بکار می‌رود:

۱. پیش‌بینی عملکرد یادگیرنده در یک هدف آموزشی براساس هدف آموزشی دیگر
۲. پیش‌بینی یادگیری در حد تسلط

روش تعیین روایی پیش‌بین عملکرد یادگیرنده در یک هدف آموزشی بر اساس هدف آموزشی دیگر

۱. مشخص ساختن ضریب همبستگی عملکرد دانش آموزان در سئوال‌های مربوط به اهداف آموزشی

روش تعیین روایی پیش‌بین یادگیری در حد تسلط

۱. مشخص ساختن ضریب همبستگی عملکرد دانش آموزان با توجه به دو حد تسلط و عدم تسلط

روایی سازه آزمون‌های وابسته به ملاک

برای تعیین روایی سازه آزمون‌های وابسته به ملاک دست کم به دو نوع تحلیل نیاز داریم:

۱. تحلیل منطقی
۲. تحلیل آماری

عوامل مؤثر بر روایی آزمون را می‌توان به دو دسته مؤثر تقسیم کرد:

۱. عوامل درونی
۲. عوامل بیرونی

عوامل درونی

یکی از عوامل درونی مؤثر کیفیت سئوال‌ات آزمون است. سئوال‌اتی که طبق قواعد درست سئوال‌نویسی تهیه شده‌اند. به روایی آزمون می‌افزاید. یکی دیگر از عوامل ترتیب قرار گرفتن سئوال‌ات از ساده به دشوار است که روایی را افزایش می‌دهد و عامل سوم افزایش تعداد سئوال‌ات آزمون است. هرچه تعداد سئوال‌ات بیشتر باشد، در روایی محتوایی نمونه بهتری از محتوا را اندازه‌گیری می‌کنند.

عوامل بیرونی

یکی از این عوامل بیرونی اجرای نامناسب آزمون مانند کمبود زمان است. دومین عامل اشکال در تصحیح پاسخ‌ها و

نمره‌گذاری به ویژه در سؤالات تشریحی است. چنانچه آزمون در گروه‌های نامتجانس پیاده گردد، روایی بیشتری دارد. ویژگی روانی پاسخ‌دهندگان مانند انگیزش نیز در روایی مؤثر است.

روش تعیین روایی سازه

(۱) تعیین همبستگی آزمون با سایر آزمون‌های روایی

(۲) تمایز سنی

(۳) تحلیل عاملی

(۴) همسانی درونی

روایی محتوایی آزمون‌های وابسته به ملاک

روایی محتوایی آزمون‌های وابسته به هنجار و آزمون‌های وابسته به ملاک شبیه هم هستند. با وجود این در تعیین روایی برای آزمون‌های وابسته به ملاک وابستگی سؤال‌های آزمون به محتوا و هدف‌های مشخص یک موضوع بیشتر است. در تعیین روایی آزمون‌های وابسته به هنجار، چون هدف‌های دقیق آموزشی موردنظر نیستند، لذا کوشش آزمون‌ساز صرفاً به نمونه‌گیری از محتوای مورد سنجش منحصر می‌شود. در حالیکه چون در آزمون‌های وابسته به ملاک کاربرد هدف‌های دقیق آموزشی ضرورت دارد، این آزمون‌ها در صورتی روا هستند که هدف‌های آموزشی درس یا موضوع مورد نظر را به دقت بسنجند.

روایی پیش‌بینی آزمون‌های وابسته به ملاک

روایی پیش‌بینی آزمون‌های وابسته به ملاک به دو منظور به کار می‌رود:

۱- پیش‌بینی عملکرد یادگیرنده در یک هدف آموزشی براساس هدف آموزشی دیگر

۲- پیش‌بینی یادگیری در حد تسلط

روایی سازه آزمون‌های وابسته به ملاک

آزمون‌هایی که برای اندازه‌گیری سازه‌های روانی درست شده‌اند، بطور مستقیم سازه‌ها را نمی‌سنجند بلکه عملکرد یا رفتارهایی را که گمان می‌رود منعکس‌کننده سازه‌ها هستند می‌سنجند. برای تعیین روایی سازه آزمون‌های وابسته به ملاک دست کم در تحلیل یکی منطقی و دیگری آماری مورد نیاز است.

پایایی آزمون

پایایی به دقت آزمون اشاره می‌کند و در صورتی آزمون پایاست که اگر در یک فاصله زمانی کوتاه چندین بار به

گروه واحدی از افراد آن را بدهیم نتایج معکوس یکسانی را بدست بیاوریم. یک آزمون باید پایا باشد تا بتواند روا باشد اما روایی برای پایایی ضروری نیست.

روش‌های تعیین پایایی

۱. روش پایایی مصححان
۲. باز آزمایی
۳. روش‌های هم ارز
۴. روش همسانی درونی

روش پایایی مصححان

این روش در آزمون‌های انشایی یا تشریحی که نمرات تحت تأثیر قضاوت مصححان قرار می‌گیرد به کار می‌رود و باید از دو یا چند مصحح استفاده کرد. همبستگی بین نمرات مصححان شاخص پایایی مصححان است.

باز آزمایی

اگر آزمون را در نوبت به گره واحدی از افراد بدهیم و ضریب همبستگی بین نمرات را بدست آوریم به آن پایایی باز آزمایی گویند.

روش‌های هم ارز

در این روش دو آزمون هم ارز را برای یک مطلب تهیه می‌کنند و در فاصله زمانی کوتاه اجرا می‌کنند. همبستگی حاصل از نمرات ضریب پایایی آن می‌باشد. مشکل عمده این روش این است که تهیه فرم‌های هم ارز به سادگی میسر نیست. لذا روش دو نیمه کردن توصیه می‌شود.

روش همسانی درونی

در همسانی درونی بر یکنواختی ماده‌ها یا اجزای تشکیل دهنده یک آزمون تأکید می‌شود و شامل روش زیر است

۱. دو نیمه کردن
۲. کودر ریچاردسون
۳. روش آلفای کرونباخ

الف- روش دو نیمه کردن

در روش دو نیمه کردن، آزمون را یکبار اجرا می‌کنیم و سپس دو نیمه می‌کنیم. (براساس سؤالات زوج و فرد) و بین

دو نیمه همبستگی می‌گیریم. همبستگی بدست آمده را در فرمول اسپیرمن براون قرار می‌دهیم تا همبستگی نهایی بدست آید.

برای محاسبه ضریب پایایی کل آزمون، ضریب همبستگی بین نیمه‌ها را در فرمول زیر قرار می‌دهیم:

$$r_{tt} = r_{tt} = \frac{2r \frac{1}{2}}{1 + r \frac{1}{2}}$$

$$r_{tt} = r \frac{1}{2}$$

روش کودریچاردسون

در این روش نیز آزمون یکبار اجرا می‌شود و فرض اساسی آن این است که سؤالات آزمون را می‌توان به صورت یک برای درست و صفر برای غلط تصحیح کرد که دارای دو فرمول است. کودریچاردسون برای بررسی همسانی درونی آزمون و تعیین پایایی آن کاربرد دارد که به KR_{20} و KR_{21} شهرت دارد. این دو فرمول به صورت زیر می‌باشد:

$$KR_{20} = r_{tt} = \left[\frac{n}{n-1} \right] \left[1 - \frac{\sum px}{8^2} \right]$$

$$1 - \frac{\sum dx}{8^2}$$

در این فرمول، r_{tt} ضریب پایایی آزمون، n تعداد سؤالات آزمون، 8^2 واریانس کل سؤالات آزمون، p نسبت دانش‌آموزانی که به سؤال پاسخ درست داده‌اند، x نسبت دانش‌آموزانی که به سؤال پاسخ غلط داده‌اند و $\sum px$ مجموع px ها برای تمام سؤالات است.

$$KR_{21} = r_{tt} = \left(\frac{n}{n-1} \right) \left(1 - \frac{n\bar{px}}{8^2} \right)$$

روش ضریب آلفا

اگر به جای سؤالات آزمون از بخش‌ها یا قسمت‌هایی تشکیل شده باشد از کرانباخ آلفا استفاده می‌شود. روش دو نیمه کردن آزمون نوعی روش آلفای کرانباخ است با افزایش تعداد سؤالات آزمون، پایایی نیز افزایش می‌یابد. افزودن تعدادی سؤال با ضریب تمیز زیاد و ضریب دشواری متوسط پایایی را افزایش می‌دهد پایایی شرط روایی است و هر اندازه از میزان پایایی کاسته شود، به همان نسبت از روایی کم می‌شود.

فرمول محاسبه آن به صورت زیر است:

$$r_{\alpha} = \frac{j}{j-1} \left(1 - \frac{\sum 8j^2}{8^2} \right)$$

در این فرمول r_{∞} ضریب پایایی کل آزمون، j تعداد بخشهای آزمون، $8j^2$ واریانس نمرات هر بخش آزمون و 8^2 واریانس نمرات کل آزمون است.

روش تعیین پایایی آزمونهای سرعت

در رابطه با استفاده از ضریب پایایی، آزمونها را می توان به دو گروه آزمونهای قدرت و آزمونهای سرعت دسته بندی کرد. آزمون قدرت به آزمونی گفته می شود که در اجرای آن به آزمون شوندگان وقت کافی داده می شود تا به همه سؤالهای آن جواب دهند. آزمونهای پیشرفت تحصیلی عموماً آزمونهای قدرت هستند. در مقابل آزمونهای سرعت به آزمونهایی گفته می شود که برای جواب دادن به سؤالهای آن محدودیت زمانی وجود دارد و اکثر آزمون شوندگان در محدوده زمانی تعیین شده نمی توانند آن را به پایان برسانند. رقم ۷۵ درصد آزمونهای سرعت و آزمونهای قدرت را از هم جدا می سازد.

روش بهتر محاسبه ضریب پایایی آزمونهای سرعت، روش بازآزمایی است. از روش دو فرم هم ارز یا موازی هم می توان استفاده کرد.

پایایی نمرات اختلاف

مهم ترین نمره اختلاف مورد استفاده در آموزش اختلاف بین نمره های دانش آموزان در یک آزمون واحد یا در دو فرم موازی یک آزمون است که در دو زمان مختلف اجرا شده اند. نام دیگر این نمره اختلاف نمره افزایش یا نمره تغییر است. برای بدست آوردن نمره اختلاف (d) از فرمول زیر استفاده می شود:

$$r_d = \frac{r_{11} + r_{22} - r_{12}}{1 - r_{12}}$$

عوامل مؤثر در افزایش پایایی آزمون (وابسته به هنجار)

- ۱- با افزودن بر تعداد سؤالات یک آزمون، پایایی آن نیز افزایش می یابد.
- ۲- با متجانس تر و همگون تر کردن سؤالات یک آزمون، پایایی آن بیشتر می شود.
- ۳- با افزودن تعدادی سؤال با ضریب تمیز زیاد به یک آزمون، پایایی آن بالا می رود.
- ۴- با افزودن تعدادی سؤال دارای ضریب دشواری متوسط به یک آزمون، پایایی آن افزایش می یابد.
- ۵- آزمونهای سرعت ضریب پایایی بالاتری نسبت به آزمونهای قدرت دارند.
- ۶- استفاده از ضریب پایایی آزمون بالاتری نسبت به آزمونهای قدرت دارند.

استفاده از ضریب پایایی آزمون برای برآورد نمره حقیقی:

مفهوم پایایی آزمون بر این باور استوار است که هیچ آزمونی کاملاً پایا نیست، لذا عموماً با نمره‌های هر آزمون قدری خطا همراه است. این خطا که بخشی از نمره بدست آمده یا نمره مشاهده شده (نمره خام) است، جزء خطای نمره گفته می‌شود. عوامل مختلفی در پدید آمدن این جزء خطا دخیل‌اند و بنابراین واریانس خطای یک واریانس تصادفی یا غیر منظم است.

اگر X_0 معرف نمره مشاهده شده، X_t نشان‌دهنده نمره حقیقی و x_e علامت نمره خطا باشد، تساوی زیر را خواهیم داشت:

نمره خطا + نمره حقیقی = نمره مشاهده شده

$$X_0 = x_t + x_e$$

اگر واریانس نمرات مشاهده شده آزمونی را با 8_0^2 ، واریانس نمرات حقیقی را با 8_t^2 ، واریانس نمرات خطای آن را با 8_e^2 نشان دهیم، رابطه این واریانس‌ها به شرح زیر خواهد بود:

$$8_0^2 = 8_t^2 + 8_e^2$$

حال می‌توانیم پایایی آزمون را به صورت نسبت واریانس نمره حقیقی به واریانس نمره مشاهده تعریف کنیم،

$$\text{یعنی به صورت: } r_{tt} = \frac{8_t^2}{8_0^2} \text{ (پایایی آزمون)}$$

خطای معیار اندازه‌گیری

خطای معیار اندازه‌گیری برابر است با انحراف معیار توزیع نمرات خطا: فرمول برآورد خطای معیار اندازه‌گیری به شرح زیر است:

$$SEM = SD_t \sqrt{1 - r_{tt}}$$

در این فرمول: خطای معیار اندازه‌گیری برحسب نمرات خام = SEM

انحراف معیار آزمون که همان انحراف معیار نمرات مشاهده شده = SD_t

ضریب پایایی آزمون = r_{tt}

ضریب پایایی آزمون‌های وابسته به ملاک (روش درصد توافق)

تعداد افرادی که در هر دو آزمون به حد تسلط رسیده‌اند + تعداد افرادی که در هر دو آزمون به حد تسلط نرسیده‌اند

تعداد کل گروه (N) × ۱۰۰ = ضریب پایایی

عوامل مؤثر در افزایش پایایی آزمون (وابسته به ملاک)

- ۱- در صورت امکان، آزمونی بسازید که بین میانگین نمرات آن و نمره حد تسلط تفاوت محسوسی وجود داشته باشد.
- ۲- برای هر هدف آموزشی یا هدف یادگیری تا حد امکان سؤال‌های زیادتری بنویسید.
- ۳- هدف‌های خود را تا آنجا که ممکن است دقیق و مشخص بنویسید.
- ۴- از روش عینی نمره‌گذاری استفاده کنید.
- ۵- سؤال‌های نمونه تمرینی، شبیه به سؤال‌های آزمون اصلی در اختیار آزمون‌شوندگان قرار دهید.
- ۶- سؤال‌های آزمون را سازمان‌یافته و روشن در اختیار آزمون‌شوندگان قرار دهید.

پایایی روش‌های سنجش مستقیم

- پایایی سنجش محصولات ماندنی رفتار یا سنجش فرآورده رفتار:

تعداد توافقه‌ها

$$\text{ضریب پایایی} = 100 \times \frac{\text{تعداد توافقه‌ها}}{\text{تعداد عدم توافقه‌ها} + \text{تعداد توافقه‌ها}}$$

تعداد عدم توافقه‌ها × تعداد توافقه‌ها

- پایایی روش ثبت فراوانی رفتار:

ضریب پایایی = $100 \times \frac{\text{کوچکترین فراوانی ثبت شده}}{\text{بزرگترین فراوانی ثبت شده}}$

بزرگترین فراوانی ثبت شده

- پایایی روش‌های سنجش طول مدت و دوره نهفتگی رفتار:

کوچکترین طول مدت ثبت شده

$$\text{ضریب پایایی} = 100 \times \frac{\text{کوچکترین طول مدت ثبت شده}}{\text{بزرگترین طول مدت ثبت شده}}$$

بزرگترین طول مدت ثبت شده

پایایی روش های ثبت فاصله ای و ثبت نمونه گیری زمان:

تعداد توافقهها

$$\text{ضرب پایایی} = 100 \times \frac{\text{تعداد توافقهها}}{\text{تعداد عدم توافقهها} + \text{تعداد توافقهها}}$$

تعداد عدم توافقهها + تعداد توافقهها

نکات کلیدی و مهم

- روش‌های تعیین روایی محتوایی و روایی صوری که مشتقی از روایی محتوایی به روش‌های قضاوتی معروف هستند.
- قضاوت‌های مربوط به روایی محتوایی نه قطعی هستند و نه نهایی.
- هرچند که روایی صوری یک ویژگی مطلوب آزمون است با این حال در بعضی آزمون‌ها این ویژگی نه تنها ضروری نیست بلکه بهتر است که وجود نداشته باشد.
- روایی ملاکی به فن تجربی مطالعه رابطه بین نمرات آزمون پیش‌بینی و نوع اندازه بیرونی مستقل یعنی ملاک مربوط می‌شود.
- روایی پیش‌بینی به موقعیتی اشاره دارد که در آن داده‌های مربوط به ملاک در آینده به دست می‌آیند و به همین دلیل نمرات آزمون واقعاً برای انجام نوعی پیش‌بینی به کار می‌رود.
- در روایی همزمان هدف آزمون پیش بین پیش‌بینی ملاک نیست.
- معنی سازه به معنی مفهوم نزدیک است.
- بنا به گفته مرفی و دیوید شوfer تاکنون هیچ گونه شاخص آماری که نشان دهنده روایی باشد بدست نیامده است.
- برای تعیین روایی پیش‌بینی عملکرد یادگیرنده در آزمون‌های وابسته به ملاک و در حد تسلط از ضریب همبستگی استفاده می‌شود.
- تفاوت بین روش ضریب آلفا و روش دو نیمه کردن آزمون تفاوت در واحد تحلیل است. در روش دو نیمه کردن آزمون، یک نیمه آزمون با نیمه دیگرش مقایسه می‌شود، در حالی که در روش ضریب آلفا هر سؤال آزمون با تک تک سؤال‌های دیگر مقایسه می‌شود.
- ضریب پایایی صفر نشان‌دهنده این است که تمامی واریانس مشاهده شده واریانس خطاست. در مقابل، ضریب پایایی ۱ معرف این است که صد در صد واریانس مشاهده شده واریانس حقیقی است.
- ضریب پایایی ارقامی از صفر تا ۱ را شامل است و هرچه ضریب پایایی کوچکتر باشد، حاکی از پایایی زیادت‌ر آزمون است. خطای معیار اندازه‌گیری عکس پایایی است. مقادیر نزدیک به صفر معرف این است که وسیله اندازه‌گیری دارای پایایی زیاد و خطای کم است و در مقابل، مقادیر بزرگ برای خطای معیار اندازه‌گیری معرف پایایی کم آزمون و خطای زیاد آن است.
- روایی به هدفی که آزمون برای تحقق بخشیدن به آن درست شده است اشاره می‌کند.
- سه هدف عمده آزمونها در آموزش و پرورش: ۱- روایی سازه ۲- روایی محتوایی ۳- روایی ملاکی

- روایی صوری مشتق شده از روایی محتوایی است.
- روایی ملاک شامل دو نوع روایی پیش بین و روایی همزمان است.
- روایی محتوایی بر این اشاره دارد که نمونه سوالی مورد نظر در یک آزمون تا چه حد نعرف کل جامعه سوالهای ممکن است که می‌توان از محتوا با دو نوع مورد نظر تهیه کرد.
- جدول مشخصات بهترین راه برای افزایش روایی محتوا است.
- جدول مشخصات وسیله ای است که در آن محتوای درس و هدفهای آموزشی به خوبی ارائه می‌شوند.
- هیچگونه روش آماری برای تعیین ضریب روایی محتوایی وجود ندارد.
- روایی صوری یک ویژگی مطلوب آزمون است و در بعضی از آزمونها بهتر است وجود نداشته باشد.
- روایی ملاک میزان ارتباط بین نمرات حاصل از یک آزمون با نمرات حاصل از یک آزمون یا وسیله اندازه‌گیری دیگر است.
- برای تعیین روایی پیش بین یک آزمون، اجرای دو آزمون ضروری است.
- روایی همزمان از لحاظ منطقی و عملی با روایی پیش بین تفاوت دارد.
- ویژگی های مطلوب ملاک: ۱- ربط داشتن ۲- بی طرفی ۳- پایایی ۴- در دسترس بودن
- روایی سازه نظری تر از روایی محتوا و روایی پیش بینی است.
- یک آزمون در صورتی دارای روایی سازه است که که نمرت حاصل از اجرای آن مفاهیم یا سازه‌های نظری مورد نظر مربوط باشد.
- راههای تعیین روایی سازه: ۱- تعیین همبستگی آزمون با سایر آزمونهای روا ۲- تمایز سنی ۳- تحلیل عاملی ۴- همسانی درونی
- روایی همگرا: یعنی اینکه این آزمونها یا آزمونهایی که در گذشته در این زمینه اجرا شده‌اند همبستگی داشته باشد.
- روایی واگرا: یعنی اینکه این آزمون با آزمونهایی که در گذشته غیر از این موضوع انجام گرفته متفاوت باشد.
- ملاک مورد استفاده برای بررسی همسانی درونی نمره کل آزمون است.
- بین تجانس آزمون شوندگان و روایی رابطه معکوس وجود دارد.
- سازه از نظریه های روانشناختی استنباط می‌شود.
- آزمونهای وابسته به ملاک و هنجار هر دو باید دارای روایی محتوایی باشند.
- روش تعیین روایی محتوایی آزمون های وابسته به ملاک استفاده از نظر متخصصان است.

- تعیین روایی سازه آزمونهای وابسته به ملاک مستلزم دو تحلیل یکی منطقی و دیگری آماری است.
- نخستین عاملی که می‌تواند بر روایی آزمون تاثیر بگذارد دستورالعمل یا راهنمای آزمون است.
- عوامل درونی روایی: ۱- دستورالعمل آزمون ۲- کیفیت سوالها ۳- ترتیب قرار گرفتن سوالها ۴- طول آزمون با تعداد سوالهای آن
- عوامل بیرونی روایی: ۱- اجرای نامناسب آزمون ۲- اشکالات موجود در تهیه پاسخها ۳- ترکیب آزمون‌شوندگان ۴- ویژگی روانی آزمون‌شوندگان
- هرچه آزمون‌شوندگان نامتجانس تر باشند روایی بیشتر خواهد شد.
- پایایی شرط روایی است اما روایی برای پایایی ضروری نیست.
- روش دو نیمه کردن بیانگر هماهنگی درونی سوالهاست.
- مطلوبترین روش برآورد پایایی آزمون، اجرای فرمهای موازی در زمانهای مختلف است.
- روش دو نیمه کردن آزمون نوعی روش ضریب آلفا است.
- ضریب پایایی می‌تواند ارقام بین ۰ تا +۱ را شامل شود.
- ضریب پایایی صفر نشان دهنده این است که تمام واریانس مشاهده شده واریانس خطا است.
- ضریب پایایی ۱ نشان دهنده این است که صد درصد واریانس مشاهده شده واریانس حقیقی است.
- خطای معیار اندازه‌گیری برابر با انحراف معیار توزیع نمرات خطا است.
- همبستگی میان سوالهای آزمون مستقیماً به پایایی آزمون مربوط است.
- تفاوت بین روش ضریب آلفا و روش دو نیمه کردن آزمون تفاوت در واحد تحلیل است.
- پایایی به دقت نتایج اشاره دارد.
- با کاهش میزان ضریب همبستگی بین دو آزمون ضریب پایایی نمرات اختلاف افزایش می‌یابد.
- واریانس حقیقی منعکس‌کننده توانایی صورت‌سنجش است در حالیکه واریانس خطا ناشی از یک عامل تصادفی است.
- هر اندازه از میزان پایایی یک آزمون کاسته شود به همان نسبت از روایی آن کم می‌شود.

فصل چهاردهم: ارزشیابی معلم

ارزشیابی معلم

همانطور که قبلاً گفته شد یکی از روش‌های ارزشیابی آموزشی ارزشیابی از کارکنان آموزشگاه است که مهم‌ترین آن ارزشیابی معلم است. برای ارزشیابی معلم هدف‌های زیادی را برشمرده‌اند که مهم‌ترین آنها یکی کمک به معلمان در جهت اصلاح و بهبود روش‌ها و فعالیت‌های آموزشی خودشان و دیگری کمک به مدیران و مسئولان سازمان‌های آموزشی در جهت تصمیم‌گیری‌های معقول‌تر و درست‌تر درباره استخدام ترفیع ارتقا و مانند اینهاست.

ارزیابی معلم به عنوان دومین درون داد اصلی نظام آموزشی یک فرایند چندوجهی می‌باشد که دارای دو هدف است: الف) ارزیابی تکوینی عملکرد معلمان از طریق ارائه داده‌های لازم و قضاوت درباره توانایی‌ها و مهارت‌های آنان و پیشنهاد برای بهبودی.

ب) ارزیابی پایانی عملکرد معلمان برای تصمیم‌گیری‌های استخدامی از قبیل ورود به خدمت، نگهداری و ارتقا. برای ارزیابی معلمان، شیوه‌های گوناگونی به کار برده می‌شود. از جمله می‌توان:

- ۱- مصاحبه‌های استخدامی به منظور قضاوت درباره آمادگی معلمان
- ۲- آزمون دانش، نگرش و مهارت‌های آنان
- ۳- نظر همگان (همکاران)
- ۴- مشاهده کلاس درس
- ۵- نظرخواهی از دانشجویان درباره فرایند تدریس معلم
- ۶- خودسنجی معلم
- ۷- ملاحظه رشد صلاحیت‌های حرفه‌ای و تخصصی معلم است.

ارزشیابی معلم عبارت است از تعیین میزان موفقیت معلمان در رسیدن به هدف‌های آموزشی خود.

در ارزشیابی معلم دو دسته اطلاع ضرورینند:

- ۱- ملاک‌ها یا معیارهای موفقیت آموزشی
 - ۲- شاخصی از میزان توفیق معلمان در رسیدن به این ملاک‌ها
- متخصصان ارزشیابی معلم برای مقاصد مختلف ارزشیابی، روش‌های گوناگونی را ابداع کرده‌اند که عمده‌ترین آنها با عناوین زیر معرفی می‌شوند:

- ۱- مصاحبه برای گزینش و نگهداری معلمان
- ۲- مشاهده از کلاس درس

انواع روش‌های ارزشیابی معلم

- ۱- مصاحبه به عنوان وسیله‌ای برای گزینش و نگهداری معلم
- ۲- مشاهده از کلاس درس
- ۳- استفاده از نظر دانش‌آموزان و دانشجویان
- ۵- استفاده از آزمون‌های عملکرد آموزشی
- ۶- خودسنجی یا ارزشیابی از خود
- ۷- ارزشیابی بوسیله همکاران
- ۸- بررسی آثار علمی معلم.

مصاحبه به عنوان وسیله‌ای برای گزینش و نگهداری معلم

روش مصاحبه با دو مقصود می‌تواند مورد استفاده قرار بگیرد:

۱- گزینش معلمان

۲- نگهداری معلمان و دادن امتیازهایی برای آنها

الف) گزینش معلمان

در گزینش معلمان توجه به سه عامل ضروری است:

- علاقه مندی به شغل معلمی

- داشتن ویژگی‌های شخصیتی ویژه معلمی

- داشتن اطلاعات کافی از دانش موضوعی

روش‌های گزینش معلم باید متمرکز بر سه عمل علاقه مندی به شغل معلمی، داشتن ویژگی‌های شخصیتی ویژه معلمی و داشتن اطلاعات کافی از دانش موضوعی باشند که عامل سوم را می‌توان از طریق آزمون‌های مختلفی مورد سنجش قرار داد. اما دو ویژگی اول قابل سنجش نیستند.

روش مصاحبه را نمی‌توان یک روش کامل ارزشیابی از معلم دانست و به هیچ وجه نباید از آن به تنهایی استفاده کرد. روش مصاحبه در انتخاب معلمان از میان داوطلبان واجد شرایط یک امر اجتناب‌ناپذیر است. همین‌طور روش مصاحبه روش خوبی برای بهره‌برداری از نتایج سایر روش‌های ارزشیابی است.

مشاهده از کلاس درس

در روش مشاهده از کلاس درس اطلاعات مورد نیاز به طور مستقیم از اعمال و رفتار معلم در کلاس گردآوری می‌شوند. عمده‌ترین شاخص‌های اثربخشی معلم عبارتند از:

- انتخاب روش‌های آموزشی متناسب با هدفهای آموزش درسی؛

- استفاده از روش‌های متنوع آموزشی؛

- رعایت ویژگی‌های فردی یادگیرندگان؛

- کاربرد طرح درس و سازماندهی درست مطالب؛

- اداره درست کلاس و حداکثر استفاده از وقت کلاس.

مهمترین روش‌های گردآوری اطلاعات با روش مشاهده

۱- روش تعیین فراوانی یا دفعات رفتار

۲- روش درجه بندی

۳- روش نقلی

محاسن و معایب روش مشاهده

الف) معایب روش مشاهده

- اسکریون: استفاده از دیدارهای کلاسی توسط همکاران معلم برای ارزشیابی از آموزش نه تنها نادرست بلکه توهین آمیز است.
- رفتن به کلاس درس معلم برای مشاهده و ارزشیابی جریان طبیعی آموزش را برهم می زند.
- تعداد مشاهدات کلاسی هر چه قدر هم زیاد باشند تعمیم پذیر نخواهند بود.
- مشاهده گران کلاس معلم غالباً بدون تعصب عمل نمی کنند.
- هیچ چیز در مشاهدات کلاس درس نمی تواند معیاری برای نتیجه گیری باشد.

محاسن روش مشاهده:

- روش مشاهده اطلاعاتی به ما می دهد که از هیچ روش دیگری نمی توان بدست آورد.
- درک بهتر روش کار معلم و یافتن راههای درست آن تنها از راه مشاهده مستقیم کلاس درس میسر می شود.

استفاده از نظر دانش آموزان و دانشجویان

در روش استفاده از نظر دانش آموزان و دانشجویان به عنوان ارزشیابی از معلم مسئولان نظر فراگیران را در مورد معلم جویا می شوند. نتایج حاصل از نظر سنجی ها در مورد روش استفاده از نظر دانش آموزان و دانشجویان نشان داده است که بازخورد حاصل از ارزشیابی دانشجویان از اثربخشی آموزش اگر با مشورت و راهنمایی همراه گردد در بالابردن سطح کیفی آموزش موثر خواهد بود. این روش بحث انگیز ترین روش ارزشیابی از کار معلم است.

انواع پرسشنامه های نظرخواهی

پرسشنامه نظرخواهی در یک درس خاص

پرسشنامه نظرخواهی در یک درس خاص ساده ترین شکل پرسشنامه نظرخواهی است که صرفاً برای مقاصد بازخورد و اصلاح و بهسازی آموزش مفید است.

انواع سؤال های پرسشنامه های نظرخواهی

سؤال های پرسشنامه نظرخواهی باید از انواع مختلفی باشد و جوانب مختلف درس را مورد ارزشیابی قرار دهد؛ و مهارتهای مختلف معلم را مورد توجه قرار دهد.

محاسن و معایب روش پرسشنامه های نظرخواهی

محاسن روش پرسشنامه های نظرخواهی

- شاگردان منبع اصلی توفیق معلم در نیل به اهداف آموزشی هستند.
- سهولت انجام ارزشیابی نسبت به دیگر روش‌های ارزشیابی
- نظر شاگردان نسبت به معلم وسیله مناسبی جهت ایجاد ارتباط بین شاگرد و معلم است.

معایب روش پرسشنامه‌های نظرخواهی

- نظر فراگیر به طور جامع نمایانگر شایستگی‌های معلم نیست.
- جنسیت در ارزشیابی از کار معلم تاثیر می‌گذارد.
- تعداد شاگردان کلاس، سطح دشواری درس و سایر عوامل بر ارزشیابی تاثیر می‌گذارد.
- نظر فراگیران در مورد شایستگی‌های معلم تحت تاثیر دوستان و رفتارهای دوستانه معلم با آنها است.

استفاده از یادگیری دانش آموزان

روش استفاده از یادگیری دانش آموزان را می‌توان بهترین روش ارزشیابی معلم دانست. حتی به عقیده برخی این روش عادلانه‌ترین و عینی‌ترین روش اثربخشی معلم است. این روش مستلزم دقت فراوان و رعایت برخی شرایط است.

استفاده از آزمون‌های عملکرد آموزشی

یکی از روش‌های معروف استفاده از یادگیری دانش آموزان برای ارزشیابی معلم آزمون‌های عملکرد آموزشی است.

محاسن و معایب روش استفاده از یادگیری دانش آموزان

محاسن روش استفاده از یادگیری دانش آموزان

- مراجعه به یادگیری بهترین ملاک ارزشیابی از معلم است چرا که یادگیری مهمترین هدف آموزشی است.

معایب روش استفاده از یادگیری دانش آموزان

- بازده یادگیری دانش آموز ملاک مناسبی برای اثربخشی آموزش معلم نیست. زیرا معلم مسئول یادگیری و عدم یادگیری شاگردان نیست و دانش آموز خودش اگر نخواهد یاد نمی‌گیرد.
- یادگیری مطالب درسی تنها دستاورد معلم نیست.
- آزمون‌های عملکرد آموزشی پایا نیستند.
- استفاده از مقدار یادگیری برای ارزشیابی در مورد معلمان که دانش آموزان دیرآموز و غیره دارند عادلانه نیست.

خودسنجی یا ارزشیابی از خود

روش خودسنجی یا ارزشیابی از خود مناسبترین روش برای مطالعه و بهسازی آموزش معلمان است حتی نتایج دیگر

ورشهای ارزشیابی از معلم زمانی بهترین فایده را برای اصلاح کار معلمان فراهم می‌کنند که نتایج آنها در اختیار معلمان قرار بگیرد و آنها با اطلاع از این نتایج به سنجش خود اقدام کنند.

کارول با تعریف خود سنجی به "فضاوت درباره آموزش خود" پنج وسیله عمده برای آن معرفی می‌کند:

- ۱- فرمهای درجه‌بندی شخصی
- ۲- گزارشهای شخصی
- ۳- مواد مطالعه شخصی
- ۴- مشاهده آموزش همکاران
- ۵- ضبط شنیداری و دیداری از تدریس خود

محاسن روش خودسنجی یا ارزشیابی از خود

- روش خود سنجی برای کشف نواقص معلم و کمک به رفع آن از بهترین روش‌های ارزشیابی معلم است.

معایب روش خودسنجی یا ارزشیابی از خود

- استفاده از واژه خود سنجی نقض غرض است چرا که ارزشیابی برای قضاوت در مورد خود معلم انجام می‌گیرد.

ارزشیابی بوسیله همکاران

هرچند که ارزشیابی بوسیله همکاران برای قضاوت در مورد فعالیت‌های معلمان به ویژه استاد دانشگاه فراوان کاربرد دارد. با این حال از این روش خیلی کم استفاده می‌شود.

محاسن روش خودسنجی یا ارزشیابی از خود

- این روش به عنوان مکمل سایر روش‌ها به ویژه در آموزش عالی مفید است.

معایب روش خودسنجی یا ارزشیابی از خود

- سطح پایایی آن پایین است.
- اساتید هر چند نمرات بالایی به هم می‌دهند با این حال در مورد تدریس توافق ندارند.

بررسی آثار علمی معلم

ملاک ارزشیابی در بررسی آثار علمی هم کیفیت و هم کمیت آثار علمی مانند کتاب، مقاله و... است.

نکات کلیدی و مهم

- فرمهای درجه‌بندی شخصی بهترین وسیله برای مقایسه برای مقایسه بین نظر معلم نسبت به خود و سایر منابع اطلاعاتی است.
- بهترین فرصت برای خودسنجی مشاهده فعالیت‌های خود در نوارهای ویدئوی و شنیدن صدای خود از نوارهای ضبط صوت و قضاوت درباره آن است.
- روش خودسنجی یا ارزشیابی از خود مناسبترین روش برای مطالعه و بهسازی آموزش معلمان است.
- به عقیده برخی روش استفاده از یادگیری دانش‌آموزان عادلانه‌ترین و عینی‌ترین روش اثربخشی معلم است.
- پرسشنامه نظرخواهی در یک درس خاص ساده‌ترین شکل پرسشنامه نظرخواهی است.
- روش استفاده از نظر دانش‌آموزان و دانشجویان بحث‌انگیزترین روش ارزشیابی از کار معلم است.
- یکی از روش‌های ارزشیابی آموزشی ارزشیابی از کارکنان آموزشگاه است که مهم‌ترین آن ارزشیابی معلم است.
- ارزشیابی از معلم عبارت است از تعیین میزان موفقیت معلم در رسیدن به اهداف آموزشی خود.
- بهترین روش ارزشیابی معلم استفاده از پیشرفت تحصیلی یادگیرندگان است.

منابع و مأخذ

- سیف، علی اکبر (۱۳۸۶). سنجش و اندازه‌گیری آموزشی. تهران: نشر دوران.
- نادری، عزت‌الله؛ سیف نراقی، مریم. (۱۳۷۵) روش‌های تحقیق در علوم انسانی. تهران : بور
- دلاور، علی (۱۳۸۰). روش‌های تحقیق و اندازه‌گیری. تهران: ویرایش
- هومن، حیدر علی، (۱۳۸۶). اندازه‌گیری های رون و تربیتی. تهران. انتشارات پیک فرهنگ.
- لطف‌آبادی، حسین (۱۳۷۴). سنجش و اندازه‌گیری در علوم تربیتی و روانشناسی. انتشارات حکیم فردوسی.