

۱۱.....	روان‌سنجی
۱۱.....	اندازه‌گیری
۱۲.....	آزمون
۱۲.....	سنجش
۱۳.....	ارزشیابی و مقیاس‌ها
۱۴.....	طبقه‌بندی آزمون‌ها
۱۵.....	فرآیند ارزشیابی
۱۵.....	گام اول: در فرآیند ارزشیابی (توصیف از موضوع مورد ارزشیابی)
۱۵.....	گام دوم: در فرآیند ارزشیابی (بررسی هدف)
۱۶.....	هدف کلی
۱۶.....	تحلیل تکلیف
۱۶.....	اهداف جزئی
۱۶.....	گام سوم: در فرآیند ارزشیابی (تبدیل اهداف جزئی به اهداف رفتاری)
۱۶.....	ویژگی‌های اهداف رفتاری
۱۷.....	طبقه‌بندی اهداف پرورشی
۱۷.....	طبقه‌بندی حیطه شناختی بلوم
۱۹.....	طبقه‌بندی حیطه شناختی کراتول و اندرسون
۲۲.....	طبقه‌بندی حیطه عاطفی کراتول
۲۳.....	طبقه‌بندی حیطه روانی - حرکتی دوه

- گام چهارم: در فرآیند ارزشیابی (جدول هدف - محتوا) ..... ۲۳
- گام پنجم: در فرآیند ارزشیابی (تهیه و تنظیم و کنترل سوالات) ..... ۲۴
- طبقه‌بندی روش‌های تدوین سوال ..... ۲۵
- قواعد تهیه سوالات انشایی یا تشریحی ..... ۲۶
- قواعد تهیه سوال‌های صحیح - غلط ..... ۲۷
- قواعد تهیه سوال‌های چندگزینه‌ای ..... ۲۷
- گام ششم: در فرآیند ارزشیابی (اجرا و نمره‌گذاری) ..... ۲۸
- روش تصحیح آزمون‌های تشریحی ..... ۲۸
- روش‌های مورد استفاده برای کاهش خطای تصحیح ..... ۲۹
- روش تصحیح و نمره‌گذاری آزمون‌های عینی ..... ۳۰
- روش‌های تصحیح ..... ۳۱
- اجرا و نمره‌گذاری آزمون‌های عملکردی ..... ۳۱
- روش‌های سنجش مشاهده‌ای ..... ۳۲
- فنون سنجش هدف‌ها و رفتارهای عاطفی ..... ۳۳
- گام هفتم: در فرآیند ارزشیابی (تحلیل سوال) ..... ۴۲
- محاسبه درجه دشواری ..... ۴۳
- محاسبه ضریب تمیز ..... ۴۴
- رابطه بین ضریب تمیز و درجه دشواری ..... ۴۵
- تحلیل گزینه‌های انحرافی ..... ۴۶
- تحلیل سوال‌های آزمون‌های عملکردی و عاطفی ..... ۴۶
- روش لوپ ..... ۴۷
- روش تحلیل عامل ..... ۴۷
- روش IRT ..... ۴۷
- سوالات ۵۰ درصد اول ..... ۴۹
- گام هشتم: در فرآیند ارزشیابی (شواهد پایایی) ..... ۵۴
- روش‌های تعیین پایایی ..... ۵۴
- روش پایایی مصححان ..... ۵۵
- روش بازآزمایی ..... ۵۵
- معایب روش بازآزمایی ..... ۵۵
- روش فرم‌های هم‌ارز ..... ۵۶
- معایب روش‌های هم‌ارز ..... ۵۶
- روش‌های همسانی درونی ..... ۵۶
- روش دو نیمه کردن ..... ۵۶
- روش ضریب آلفای کرونباخ ..... ۵۷
- روش کودر - ریچاردسون ۲۰ و ۲۱ ..... ۵۸

- ۵۹..... روش تعیین پایایی در آزمون‌های ملاک - مرجع
- ۶۲..... روش درصد توافق
- ۶۲..... روش کاپا
- ۶۳..... عوامل موثر در افزایش پایایی آزمون‌های ملاک مرجع
- ۶۳..... عوامل موثر در افزایش پایایی آزمون‌های هنجار - مرجع
- ۶۳..... گام نهم: در فرآیند ارزشیابی (شواهد روایی)
- ۶۳..... تعریف روایی
- ۶۳..... رابطه بین روایی و پایایی
- ۶۴..... طبقه‌بندی روش‌های تعیین روایی
- ۶۴..... روایی محتوایی
- ۶۵..... روایی صوری
- ۶۵..... روایی منطقی
- ۶۵..... روایی ملاکی
- ۶۵..... روایی سازه
- ۶۷..... عوامل موثر بر روایی
- ۶۷..... گام دهم: در فرآیند ارزشیابی (هنجارها و نیم‌رخ‌ها)
- ۶۸..... تبدیل‌های خطی
- ۷۰..... تبدیل‌های غیرخطی
- ۷۲..... سوالات ۵۰ درصد دوم
- ۷۷..... پاسخنامه سوالات ۵۰ درصد اول
- ۷۷..... پاسخنامه سوالات ۵۰ درصد دوم

## روان سنجی<sup>۱</sup>

ساختار و کاربرد آزمون‌های روانی می‌باشد، همین‌طور به اندازه‌گیری‌های روانی نیز اطلاق می‌گردد. [۱]

برای کمی کردن داده‌های پژوهش‌های روان شناختی و علوم تربیتی باید آنها را اندازه‌گیری کرد. و بطور کلی علم وابسته به بحث اندازه‌گیری می‌باشد. هرچه اندازه‌گیری دقیق‌تر باشد، اطلاعات علمی دقیق‌تر است.

## اندازه‌گیری<sup>۲</sup>

اندازه‌گیری عبارت است از نسبت دادن عدد به یک پدیده. نکته حائز توجه آن است که، به پدیده نمی‌توان عدد نسبت داد، بلکه به صفات هر پدیده می‌توان عدد نسبت داد. بعنوان مثال؛ میز را نمی‌توان اندازه‌گیری کرد، بلکه صفات آن میز از جمله ارتفاع، طول، عرض، حجم و جزآن را می‌توان اندازه‌گیری کرد.

از سوی دیگر خود عدد را نیز نمی‌توان نسبت داد بلکه ویژگی‌های اعداد را نسبت می‌دهیم. بنابراین می‌توان اندازه‌گیری را بدین صورت تعریف نمود؛ نسبت دادن ویژگی‌های عدد بر اساس قواعد به صفت هر پدیده، در این صورت، چون بر اساس قواعدی این فرآیند صورت گرفته است، می‌توان آن را تعمیم داد و نتایج یکسانی را با آزمون‌های استاندارد دیگر بدست آورد. [۲]

اعداد چهار ویژگی دارند:

(۱) به جای اسم به کار می‌روند.

(۲) بزرگی و کوچکی دارند (ترتیب دارند).

(۳) فواصل مساوی دارند.

(۴) قابل نسبت بستن می‌باشند.

بر اساس ویژگی‌های اعداد، مقیاس<sup>۳</sup> شکل می‌گیرد.

- اگر فقط از ویژگی اول اعداد استفاده کنیم: مقیاس اندازه‌گیری اسمی می‌باشد. این مقیاس برای طبقه بندی کردن به کار می‌رود. مثل شماره پلاک خودرو

- اگر از ویژگی‌های اول و دوم اعداد استفاده کنیم: مقیاس اندازه‌گیری رتبه‌ای (ترتیبی) می‌باشد. این مقیاس علاوه بر طبقه بندی کردن می‌تواند رتبه را نیز تعیین کند. مثل رتبه فرد در کنکور

- اگر از ویژگی‌های اول تا سوم اعداد استفاده کنیم: مقیاس اندازه‌گیری فاصله‌ای می‌باشد. این مقیاس علاوه بر طبقه بندی و تعیین رتبه، فاصله اعداد را نیز نشان می‌دهد. مثل نمرات تمام آزمون‌های استاندارد، نگرش سنج‌ها، امتحانات پیشرفت تحصیلی و جزآن

- اگر از تمام ویژگی‌های اعداد استفاده کنیم: مقیاس اندازه‌گیری نسبی می‌باشد. در این مقیاس صفر واقعی است؛ به عبارت دیگر، صفر معنی هیچ می‌دهد.

<sup>۱</sup> Psychometry

<sup>۲</sup> Measurment

<sup>۳</sup> Scale

نکته: نکته: در مقیاس فاصله‌ای صفر قراردادی است. مثل: صفر در آزمون زبان کنکور ( در این حالت صفر به معنای هیچ نمی‌باشد).

نکته: در مقیاس‌های اسمی و ترتیبی اعداد را نمی‌توان باهم جمع، تفریق، ضرب و تقسیم کرد.

نکته: در مقیاس‌های فاصله‌ای و نسبی اعداد را می‌توان باهم جمع، تفریق، ضرب و تقسیم کرد.

## آزمودن<sup>۴</sup>

اندازه‌گیری با استفاده از یک ابزار استاندارد (آزمون<sup>۵</sup>) [۳]

فرق آزمودن و اندازه‌گیری: هر آزمودنی اندازه‌گیری است ولی هر اندازه‌گیری الزاماً آزمودن نیست، زیرا ممکن است ابزار اندازه‌گیری استاندارد نباشد. بنابراین اندازه‌گیری اعم از آزمودن است.

مثال: تست هوش آزمودن است و اندازه‌گیری نیز می‌باشد، اما امتحان ریاضی اندازه‌گیری است ( چون ابزار غیر استاندارد است) ولی آزمودن نیست.

## ارزیابی<sup>۶</sup> (سنجش)

عبارت است از اندازه‌گیری به همراه سنجش کیفی صفت مورد اندازه‌گیری [۳] بنابراین فراتر از اندازه‌گیری است، زیرا اندازه‌گیری‌های کیفی را نیز به همراه دارد.

نکته: نکته: آزمودن وابسته به مجری اندازه‌گیری نیست.

نکته: نکته: ارزیابی از مجری ارزیابی اثر می‌پذیرد.

---

<sup>4</sup> Testing

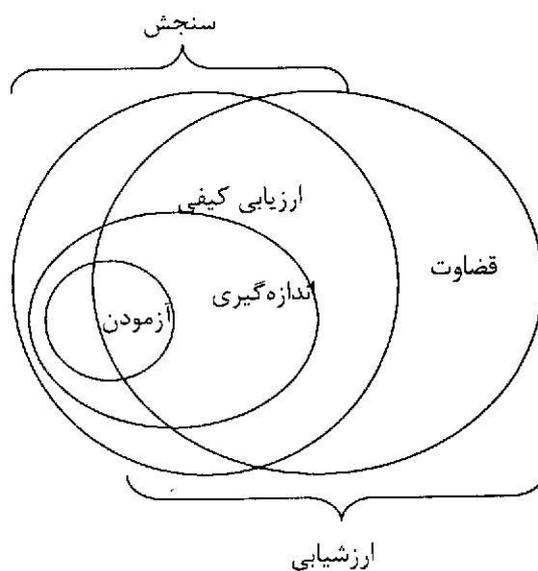
<sup>5</sup> Test

<sup>6</sup> Assessment

## ارزشیابی<sup>۷</sup>

اندازه‌گیری به همراه قضاوت ارزشیاب است. قضاوت یعنی مقایسه با یک ملاک. بنابراین می‌توان روابط اندازه‌گیری، آزمودن، ارزیابی و ارزشیابی را به صورت شکل زیر نمایش داد.

شکل یک: روابط اندازه‌گیری، آزمودن، ارزیابی و ارزشیابی



## طبقه بندی آزمون ها<sup>۸</sup>



نکته: نکته: شباهت آزمون استعداد اختصاصی و پیشرفت تحصیلی: محتوا و هدف است.

نکته: نکته: تفاوت آزمون استعداد اختصاصی و پیشرفت تحصیلی: آزمون استعداد معطوف به آینده است و آزمون پیشرفت تحصیلی معطوف به یادگیری گذشته است.

آزمون‌ها را به دو حالت « هنجار مرجع » و « ملاک مرجع » نیز می‌توان تقسیم کرد: [۳]

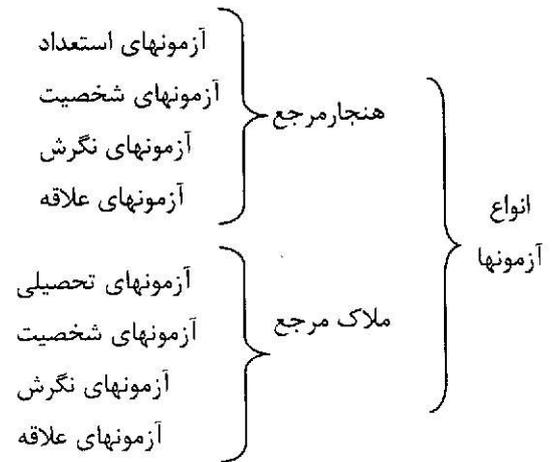
- آزمون‌های ملاک مرجع: بر اساس یک نمره معین، ملاک مطلق، یا معیار که توسط آزمون‌گر تعیین می‌شود، بگوییم قبول یا رد. [۳]

- آزمون‌های هنجار مرجع: در مقایسه با یک جامعه و هنجار قضاوت شده در آن جامعه تصمیم بگیریم. [۳]

نکته: در ملاک مرجع مقایسه با یک ملاک است و در هنجار مرجع مقایسه با یک هنجار است.

<sup>8</sup> این بخش کاملاً برگرفته از کتاب اندازه‌گیری، سنجش، و ارزشیابی آموزشی، دکتر علی اکبر سیف ۱۳۸۲ صفحه ۶۳ می‌باشد.

بنابراین می‌توان آزمون‌ها را بر اساس ملاک مرجع یا هنجار مرجع نیز طبقه بندی نمود:



## فرآیند ارزشیابی

ارزشیابی در طی ده گام صورت می‌گیرد، که به قرار زیر است:

**گام اول:** توصیفی از موضوع مورد ارزشیابی

**گام دوم:** بررسی هدف

**گام سوم:** تبدیل اهداف جزئی به اهداف رفتاری برای آزمون پذیر شدن

**گام چهارم:** تهیه جدول مشخصات

**گام پنجم:** تهیه، ویرایش، و کنترل سؤالات مربوطه

**گام ششم:** اجرا و نمره گذاری

**گام هفتم:** تحلیل سؤال برای انتخاب سؤالات مناسب و حذف سؤالات نامربوط

**گام هشتم:** بررسی شواهد پایایی و بدست آوردن مقدار خطای اندازه‌گیری

**گام نهم:** به دست آوردن شواهد اعتبار

**گام دهم:** هنجاریابی

در ادامه به توضیح تفصیلی ده گام فوق می‌پردازیم:

## گام اول

از موضوع مورد ارزشیابی توصیفی به عمل می‌آوریم شامل اینکه چه اهدافی باید مورد ارزشیابی واقع شوند، پیش نیازها کدام هستند، اهمیت اهداف چگونه است، نحوه اجرا چگونه باید باشد و جز آن.

## گام دوم

بررسی هدف، در این مرحله ارزشیاب باید به بررسی هدف بپردازد. هدفی که به ارزشیاب ارائه می‌گردد هدف کلی یا غایی است.

### هدف کلی یا غایی

بیانی کلی است که قصد و منظور از ایجاد و اداره دوره آموزشی یک جامعه را نشان می‌دهد. [۳] ولی با یک هدف غایی نمی‌توان کل ماجرا را تحلیل کرد، پس باید به اهداف جزئی‌تر تقسیم کرد تا بدین طریق آزمون پذیر شود و این کار را توسط روش تحلیل تکلیف انجام می‌دهند.

### تحلیل تکلیف

برای آزمون سازی، منابع و پژوهش‌های مربوط به آن موضوع را مطالعه کرده و موانع و نکات مهم آن را مشخص می‌سازیم و بر این اساس هدف کلی را به هدف‌های جزئی‌تر تقسیم می‌کنیم و بدین صورت آزمون را طراحی می‌کنیم. نکته حائز توجه آن است اهداف جزئی را باید سلسله مراتب کنیم و اهداف پیش نیاز، پس نیاز و هم نیاز را مشخص نمائیم. اهداف جزئی: اهدافی می‌باشند که مورد قضاوت و ارزشیابی ارزشیابان قرار می‌گیرد [۳]

## گام سوم

تبدیل اهداف جزئی به اهداف رفتاری

در تبدیل اهداف جزئی به اهداف رفتاری به دو دسته ویژگی توجه می‌شود:

- (۱) باید به حیطه و سطح ارزشیابی مورد نظر در آن هدف جزئی توجه کنیم (به ویژه در آزمون‌های توانایی)
- (۲) باید ویژگی‌های هدف رفتاری را در هدف نویسی رعایت کنیم.

### ویژگی‌های اهداف رفتاری به قرار زیر می‌باشند:

شرط (مجموعه امکانات و وسایلی که در اختیار پاسخگو می‌تواند قرار بگیرد یا نباید قرار بگیرد) ملاک (حداکثر درصد خطایی که آزمودنی مجاز است انجام دهد یا حداقل میزان صحت)

نکته: در آزمون‌های سرعت، ملاک زمان می‌باشد.

رفتاری بودن (مشخصاً یک عملکرد/ فعل قابل مشاهده و اندازه‌گیری را در هدف بگنجانند) [۴]

مثال: دانش آموز باید بتواند مفاهیم کلیدی را بدون استفاده از کتاب به درستی تعریف کند.

الف: شرط است.

ب: ملاک است.

ج: رفتاری بودن است. (عملکرد است)

برای تعیین حیطه و سطح ارزشیابی هدف باید به طبقه بندی اهداف توجه کرد:

## طبقه بندی بلوم از اهداف پرورشی

بلوم اهداف پرورشی را به سه حیطه کلی تقسیم کرده است. حیطه شناختی، حیطه عاطفی، حیطه روانی - حرکتی.

## طبقه بندی بلوم از حیطه شناختی

- دانش

- فهمیدن

- کار بستن

- تجزیه و تحلیل

- ترکیب

- ارزشیابی

**نکته:** طبقه بندی بلوم به صورت سلسله مراتب است. یعنی فهمیدن پیش نیاز کار بستن و پس نیاز دانش است.

**نکته:** در سلسله مراتب اهداف جزئی علاوه بر سلسله مراتب اهداف مبتنی بر سازمان در س باید به سلسله مراتب آنها مبتنی بر طبقه بندی بلوم نیز توجه کرد.

سطح دانش به سه زیر سطح دانش، امور جزئی، دانش امور کلی و دانش روندی تقسیم می گردد. سطح دانش به معنی یادآوری است مثل آنکه دانش آموز بتواند معنی لغات را به درستی بگوید یا تشخیص دهد.

دانش امور جزئی به معنی یادآوری امور جزئی است مثل آنکه دانش آموز بدون کمک جدول ضرب را از حفظ بگوید.

دانش امور کلی یادآوری مطالب کلی تر است مثل آن که فراگیر بدون کمک از کتاب یا جزوه تئوری اسکینر را از حفظ بگوید.

دانش امور روندی به معنی یادآوری فرآیند است مثل آن که فراگیر فرآیند تولید یک قطعه را بگوید یا بیان نماید که رودخانه دانوب از کجا سرچشمه می گیرد و به ترتیب از کدام کشورها یا شهرها گذشته، به کدام دریا می ریزد.

**نکته:** دانش همان حفظ کردن است.

سطح فهمیدن به سه زیر سطح ترجمه، تفسیر و برون یابی (استخراج) تقسیم می گردد.

سطح فهمیدن ابتدای یادگیری تلقی می گردد و به معنی هرگونه تغییر نسبت به یادآوری می باشد و از یادآوری و حفظ طوطی وار مطالب عمیق تر است.

ترجمه، اولین زیر سطح فهمیدن است و به معنی کمترین تغییر نسبت به یادآوری می باشد و همانگونه که از اسم آن پیداست، تغییر بیان است. مثل آن که فراگیر معنی شعر را توضیح دهد (به شرط آن که معلم قبلاً آن را نگفته باشد) یا یک فرمول آماری را با استفاده از اطلاعات یک مسئله جایگذاری نماید، یا به زبان خودش یک تئوری را خلاصه نماید، یا یک قطعه را از فارسی به انگلیسی ترجمه نماید.

تفسیر، دومین سطح فهمیدن است. به معنی ربط دادن به قانون علمی است و از طریق برقراری این ارتباط می توان یک پدیده را توضیح داد. به عنوان مثال؛ وقتی از فراگیر می خواهیم قانون مویبگی را به طور خلاصه بیان نماید، در واقع، از او خواسته ایم در

سطح ترجمه توضیح دهد. وقتی از فراگیر بخواهیم دلیل قانون مویبگی را بیان نماید، در واقع، از او خواسته ایم در سطح تفسیر توضیح دهد.

برون یابی (استخراج)، عمیق‌ترین سطح فهمیدن است. به معنی برون یابی یا درون یابی یک قانون یا اطلاع علمی است. برای مثال؛ وقتی از فراگیر بخواهیم برای قانون مویبگی یک مثال بزنند، در واقع، از او خواسته ایم در سطح برون یابی توضیح دهد. یا مثال دیگر: وقتی از فراگیر بخواهیم یک مسئله را با استفاده از قانون تقسیم حل کند، این یک مورد برون یابی است. بنابراین، در کلیه مواردی که به صورت تمرین پس از آموزش یک قانون علمی استفاده می‌شود، برون یابی است. اما در مورد درون یابی یا استخراج می‌توان مثال زیر را بکار برد. در جای ؟ عدد مناسب قرار دهید: ۲۲ ؟ ۷ ۵ ۳ ۱

مثال فوق، مثالی از درون یابی و یا استخراج است. یعنی فراگیر باید روابط بین اعداد را به دست آورده و بر اساس آن جای خالی را پر نماید. در مثال فوق پاسخ صحیح ۱۱ است. زیرا روابط بین اعداد تاریخ پیروزی انقلاب اسلامی را نشان می‌دهد.

اگر توجه کرده باشید در برون یابی یا استخراج قانون به فراگیر ارائه گردیده و مثال یا کاربرد از وی خواسته می‌شود مطابق شکل روبرو قانون ← مثال

سطح کاربرستن، از فهمیدن عمیق‌تر است به معنی فرآیندی معکوس برون یابی است. همانگونه که اشاره گردید؛ برون یابی از قانون به مثال حرکت می‌کرد اما در برون یابی؛ مثالی را به فراگیر ارائه می‌کنند و او باید قانونی را برای حل آن یا توضیح آن تشخیص دهد. به عنوان مثال: وقتی از فراگیر بخواهیم توضیح دهد که به کدام دلیل آب در برخی جاه‌ها بالاتر می‌ایستد و وی برای حل این مسئله از قانون مویبگی استفاده نماید، این مثال در سطح کاربرستن است.

بنابراین، در کلیه مواردی که به صورت امتحان قانون علمی توسط فراگیر به درستی تشخیص داده می‌شود و برای حل مسئله به طور مناسب به کار برده می‌شود، کاربرستن است.



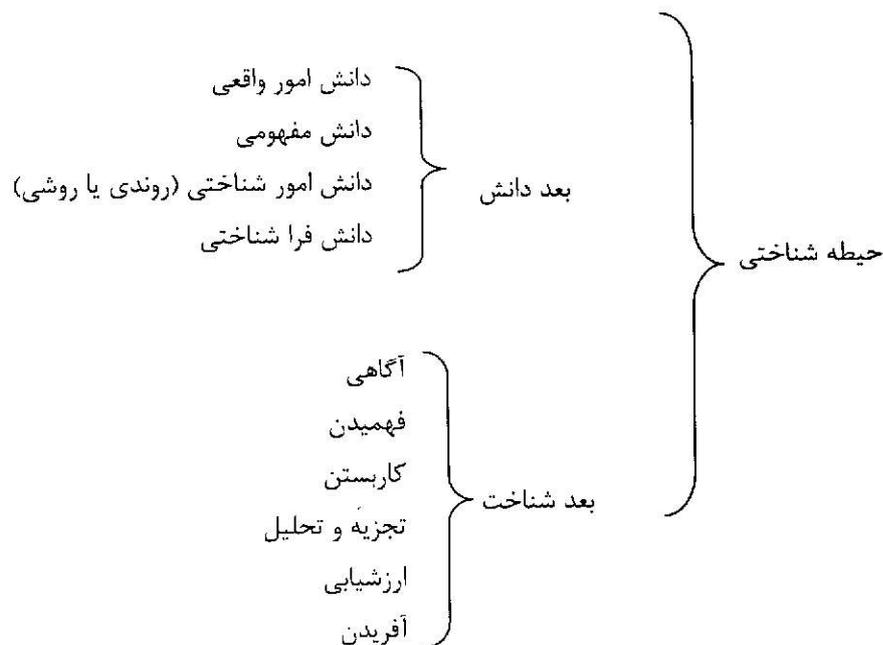
تجزیه و تحلیل: سطحی عمیق‌تر از کاربرستن است. به معنی باز کردن موضوع علمی است در تجزیه و تحلیل با یک مقوله انتزاعی سروکار داریم (مثل مقایسه دو تئوری) باید توجه داشت که وقتی فراگیر یک موتور بنزینی را باز و بسته می‌نماید. (بدون توجه به مهارت‌های حرکتی) این فعالیت در سطح فهمیدن تفسیر و یا حتی کمتر در سطح دانش روندی است و نباید به آن تجزیه و تحلیل اطلاق نمود. زیرا تجزیه و تحلیل مربوط به مقولات انتزاعی است. برای مثال: از فراگیر بخواهیم قانون مویبگی را با قانون ظروف مرتبطه مقایسه نماید و موارد اشتراک و افتراق این دو قانون را بیان نماید.

نکته: در تجزیه و تحلیل با شکستن اصول و نظریه‌ها سروکار داریم.

ترکیب: سطحی عمیق‌تر از تجزیه و تحلیل است. به معنی ترکیب اطلاعات به دست آمده قبلی و خلق دانش جدید است. برای مثال؛ فراگیر بتواند از امکاناتی که در اختیار وی قرار می‌دهیم، دانش جدیدی را خلق کند. یا توانایی ارائه یک طرح پژوهشی، مربوط به سطح ترکیب است. تولید سؤال در سطح کاربرستن و یا عمیق‌تر نیز در سطح ترکیب است.

ارزشیابی : سطحی عمیق تر از ترکیب است . به معنی همان تجزیه و تحلیل به اضافه قضاوت است. به عنوان مثال؛ مقایسه دو تئوری و کشف شباهت‌ها و تفاوت‌ها و قضاوت در مورد مفید بودن یکی از آن‌ها. [۵]

نکته: اگرچه بلوم معتقد به سلسله مراتب بودن این سطوح را داشت ولی به نظر می‌رسد ترکیب پیش نیاز ارزشیابی نباشد بلکه تجزیه و تحلیل پیش نیاز ارزشیابی قرار گیرد. پس از مرگ بلوم کراتول و اندرسون طبقه بندی دیگری از حیثه شناختی ارائه کردند، که دارای دو بعد دانش و شناخت است. و به قرار زیر می‌باشد.



این دو بعد ۲۴ سطح را به وجود می‌آورند که به قرار جدول زیر است:

آفریدن	ارزشیابی	تجزیه و تحلیل	کار بستن	فهمیدن	آگاهی	بعد شناخت / بعد دانش
۲۱	۱۷	۱۳	۹	۵	۱	امور واقعی
۲۲	۱۸	۱۴	۱۰	۶	۲	مفهومی
۲۳	۱۹	۱۵	۱۱	۷	۳	امور شناختی (روندی)
۲۴	۲۰	۱۶	۱۲	۸	۴	فراشناختی

(۱) آگاهی امور واقعی: بازیابی اطلاعات واقعیت‌های اساسی، اصطلاحات و جزآن از حافظه .

برای مثال: فراگیر بتواند تعداد زوایای شش ضلعی را تشخیص دهد.

(۲) آگاهی مفهومی: بازیابی دانش مربوطه به اجزا و روابط متقابل میان آن‌ها از حافظه.

برای مثال: فراگیر بتواند روابط میان عناصر را بنویسد.

(۱) آگاهی روندی: بازیابی دانش مربوط به مراحل اجرای یک فعالیت.

**برای مثال:** فراگیر بتواند سرچشمه رود کارون را بنویسد و شهرهایی را که این رود از آن‌ها عبور می‌کند را به ترتیب نام ببرد و تعیین نماید که به کجا می‌ریزد.

۲) **آگاهی فراشناختی:** بازیابی دانش مربوط به شناخت بر موضوع مورد مطالعه و شناخت روش‌ها و تکنیک‌های یادگیری مناسب از حافظه.

**برای مثال:** فراگیر بتواند برنامه ریزی مورد نیاز برای یادگیری رگرسیون چند متغیر ی را که قبلاً به وی آموزش داده شده است، بیان نماید.

۳) **فهمیدن امور واقعی:** ساختن معنی از ارتباط بین مطالب تازه و قبلی درباره واقعیت‌های اساسی و جزآن .

**برای مثال:** ترجمه یک پارگراف از زبان انگلیسی به فارسی به شرط آن که نکات دستوری و لغات به کار رفته را از قبل بداند.

۴) **فهمیدن امور مفهومی:** ساختن معنی از ارتباط بین مطالب تازه و قبلی درباره طبقات و روابط متقابل میان آن‌ها .

**برای مثال:** طبقه بندی از آزمون‌های آماری که قبلاً به وی تدریس شده است، را انجام دهد.

۵) **فهمیدن روندی:** ساختن معنی از ارتباط بین مطالب تازه و قبلی درباره مراحل انجام یک فعالیت .

**برای مثال:** بیان مراحل وقایع تاریخی از راه مقایسه آن‌ها با یکدیگر.

۶) **فهمیدن فراشناختی:** ساختن معنی از ارتباط بین مطالب تازه و قبلی درباره شناخت بر موضوع مورد مطالعه و شناخت بر روش‌ها و تکنیک‌های یادگیری مناسب.

**برای مثال:** فراگیر الگوهای مختلف موجود در درون مثال‌های متعدد را استنباط نماید.

۷) **کار بستن امور واقعی:** انتخاب روش برای حل مسائل درباره واقعیت‌های اساسی و جزآن

**برای مثال:** حل تمرینات کتاب درباره ضرب دو رقم در دو رقم.

۸) **کار بستن امور مفهومی:** انتخاب روش برای حل مسائل درباره طبقات و روابط متقابل میان آن‌ها.

**برای مثال:** توانایی حل کردن مسائل مالی شرکت‌های خصوصی گوناگون.

۹) **کار بستن روندی:** انتخاب روش برای حل مسائل درباره مراحل انجام یک فعالیت.

**برای مثال:** توانایی عیب یابی و تنظیم موتور انژکتوری.

۱۰) **کار بستن فراشناختی:** انتخاب روشی برای حل مسائل درباره شناخت بر موضوع مورد مطالعه و شناخت بر سبک‌ها و تکنیک‌های یادگیری مناسب.

**برای مثال:** توانایی راهنمایی دانشجویان سال پایین‌تر یا مقطع پایین‌تر برای انجام یک پروژه تحقیقاتی.

۱۱) **تحلیل امور واقعی:** شناخت عناصر و توصیف روابط میان آن‌ها درباره واقعیت‌های اساسی.

**برای مثال:** توانایی بیان نکات اشتراک بین دو رویکرد رفتارگرایی و شناخت‌گرایی در یادگیری ماز توسط یک موش.

۱۲) **تحلیل امور مفهومی:** شناخت عناصر و توصیف روابط میان آن‌ها درباره طبقات و روابط متقابل میان آن‌ها .

**برای مثال:** ارائه یک مقاله در جمع بندی پژوهش‌های انجام شده درباره تکلیف شب.

(۱۳) *تحلیل روندی:* شناخت عناصر و توصیف روابط میان آن‌ها درباره مراحل اجرای یک فعالیت.

**برای مثال:** تحلیل عوامل پیروزی تیم والیبال ایران بر تیم والیبال قطر و تعیین نقاط آسیب پذیر در تیم والیبال ایران.

(۱۴) *تحلیل فراشناختی:* شناخت عناصر و توصیف روابط میان آن‌ها درباره شناخت بر موضوع مورد مطالعه و شناخت بر سبک‌ها و تکنیک‌های یادگیری مناسب.

**برای مثال:** تحلیل فعالیت‌های آموزشی مؤسسه پارسه و تعیین عوامل موفقیت دانشجویان در کنکور سراسری در رشته روانشناسی تربیتی و ارائه راهکارهای آموزشی مناسب.

(۱۵) *ارزشیابی امور واقعی:* داوری کردن با استفاده از یک معیار یا ملاک درباره واقعیت‌های اساسی.

**برای مثال:** قضاوت کردن در این باره که کدام یک از دو روش آماری برای تحلیل یک مجموعه داده مناسب‌تر است.

(۱۶) *ارزشیابی مفهومی:* داوری کردن با استفاده از یک معیار درباره طبقات و روابط متقابل میان آن‌ها.

**برای مثال:** تعیین این که پیشنهادات ارائه شده در یک پژوهش به طور منطقی حاصل نتایج پژوهش می‌باشند یا خیر.

(۱۷) *ارزشیابی روندی:* داوری کردن با استفاده از یک معیار درباره مراحل انجام یک فعالیت.

**برای مثال:** ارزیابی اثر بخشی یک روش درمانی جدید برای بیماری افسردگی دو قطبی.

(۱۸) *ارزشیابی فراشناختی:* داوری کردن با استفاده از یک معیار درباره شناخت بر موضوع مورد مطالعه و شناخت بر سبک‌ها و تکنیک‌های یادگیری مناسب.

**برای مثال:** ارزیابی عوامل موفقیت دانشجویان در کنکور سراسری در رشته روانشناسی تربیتی در مؤسسه پارسه و دیگر مؤسسات. یا ارزیابی اثر بخشی آموزشی در تربیت دبیر فنی در دانشگاه شهید رجایی و دیگر دانشگاه‌ها می‌شود.

(۱۹) *آفریدن امور واقعی:* قرار دادن عناصر در یک ساختار جدید درباره واقعیت‌های اساسی.

**برای مثال:** تولید فرضیه جدید برای توجیه داده‌های مشاهده شده.

(۲۰) *آفریدن مفهومی:* قرار دادن عناصر در یک ساختار جدید درباره طبقات و روابط متقابل میان آن‌ها.

**برای مثال:** طراحی یک برنامه ارزشیابی برای تعیین اثربخشی برنامه‌های آموزش تربیت دبیر فنی.

(۲۱) *آفریدن روندی:* قرار دادن عناصر در یک ساختار جدید درباره مراحل انجام یک فعالیت.

**برای مثال:** طراحی یک برنامه ارزشیابی برای تعیین کارآمدی آموزش جوشکاری.

(۲۲) *آفریدن فراشناختی:* قرار دادن عناصر در یک ساختار جدید درباره شناخت موضوع مورد مطالعه و شناخت سبک‌ها و تکنیک‌های یادگیری مناسب.

**برای مثال:** طراحی یک برنامه پژوهشی جهت ارزیابی عوامل موفقیت دانشجویان در کنکور سراسری در رشته روانشناسی تربیتی

در مؤسسه پارسه. [۳]

**نکته:** بعد شناختی طبقه بندی کراتول و اندرسون و طبقه بندی بلوم همانند هم است. با این فرق که در طبقه بندی جدید دو طبقه آخر جابجا شده اند.

**نکته:** شاید فردی توانایی قضاوت و ارزشیابی را به دلیل داشتن تفکر نقادانه داشته باشد ولی برای آفریدن باید تفکر واگرا باشد.

**نکته:** در این طبقه جدید، طبقه فهمیدن گسترده‌تر از فهمیدن بلوم در نظر گرفته شده است.

## طبقه بندی حیطة عاطفی

توسط کراتول صورت گرفته است که به قرار زیر می‌باشد:

- توجه کردن
- پاسخ دادن
- ارزش گذاری
- سازمان دادن ارزش ها
- تشخیص طلبی

- **توجه کردن:** حساسیت نسبت به وجود بعضی پدیده‌ها و توجه کردن به آن‌ها.

**برای مثال:** فراگیر به فعالیت‌های آموزشی معلم توجه دارد.

- **پاسخ دادن:** توجه فعالانه به پدیده‌ها و انجام فعالیتی درباره آن‌ها.

**برای مثال:** فراگیر به طور داوطلبانه درباره یک موضوع خاص مطالعه کند.

- **ارزش گذاری:** باور پایدار حاکی از ارزشمند بودن یک اندیشه.

**برای مثال:** فراگیر در بحث گروهی از یک اندیشه خاص درباره یک موضوع اجتماعی دفاع می‌کند.

- **سازمان دادن ارزش‌ها:** دسته بندی ارزش‌ها به گونه‌ای که بعضی از آن‌ها مهم‌تر از دیگر ارزش‌ها قلمداد می‌شوند.

**برای مثال:** فراگیر نیاز به یافتن نوعی تعادل بین رشد اقتصادی و حفظ منابع طبیعی را تشخیص می‌دهد.

- **تشخیص طلبی:** تبدیل نظام ارزشی به صورت یک سبک زندگی.

**برای مثال:** فراگیر در موقعیت‌های مختلف متناسب با نظام ارزشی خودش رفتار می‌کند. [۳]

**نکته:** در طبقه بندی کراتول در حیطة عاطفی سلسله مراتب وجود دارد.

**نکته:** اگر آموزش در حیطة شناختی در سطح دانش باشد. حداکثر سطح عاطفی که جذب می‌کند «توجه کردن» است.

**نکته:** اگر آموزش در حیطة شناختی در سطح فهمیدن باشد. حداکثر سطح عاطفی که جذب می‌کند «پاسخ دادن» است.

**نکته:** اگر آموزش در حیطة شناختی در سطح کاربرستن باشد. حداکثر سطح عاطفی که جذب می‌کند «ارزش گذاری» است.

**نکته:** اگر آموزش در حیطة شناختی در سطح تجزیه و تحلیل باشد. حداکثر سطح عاطفی که جذب می‌کند

« سازمان دادن ارزش ها » است.

نکته: اگر آموزش در حیطه شناختی در سطح ترکیب باشد. حداکثر سطح عاطفی که جذب می‌کند « تشخیص طلبی » است.

### طبقه بندی حیطه روانی حرکتی

توسط دوه صورت گرفته است که به قرار زیر می‌باشد:

- مشاهده و تقلید
- اجرای مستقل
- دقت
- هماهنگی حرکات
- عادی شدن
- **مشاهده و تقلید:** در این مرحله فراگیر پس از مشاهده هر مهارت به تقلید آن می‌پردازد. تمرین در این مرحله تقلید کردن است.
- **اجرای مستقل:** تشخیص بدهد پس از هر مرحله در یک مهارت، چه کاری باید انجام دهد.
- **دقت:** در این مرحله فراگیر مهارت را آموخته است و محصول تولید شده باید دقیق باشد و خطا نداشته باشد.
- **هماهنگی حرکات:** فراگیر حرکاتش در حین انجام عمل سریع و هماهنگ باشد.
- **عادی شدن:** سطحی از مهارت که اجرای یک فعالیت با صرف حداقل انرژی و توجه باشد. [۶]

نکته: در مرحله اجرای مستقل ارزشیابی از فرآیند انجام عمل<sup>9</sup> و اجرا صورت می‌گیرد.

نکته: در مرحله دقت ارزشیابی از محصول<sup>10</sup> تولید شده صورت می‌گیرد.

نکته: در طبقه بندی حیطه روانی- حرکتی سلسله مراتب وجود دارد.

### گام چهارم

گام چهارم فرآیند ارزشیابی تهیه جدول محتوا- هدف یا جدول مشخصات است. این جدول دارای دو بعد می‌باشد. یکی محتوا و دیگری هدف. در گام قبلی اهداف را نوشته و سلسله بندی می‌کردیم. در این گام باید اهداف سلسله مراتب شده را در داخل این جدول قرار دهیم و تعیین کنیم که نسبت هر هدف با توجه محتوایش در تولید سؤالات آزمون چقدر باشد. از دو طریق می‌توان نسبت هر هدف را با توجه به محتوایش تعیین کرد. یکی از روی اهمیت موضوع آموزشی و دیگری مدت زمانی که برای آموزش هر هدف در نظر می‌گیریم.

<sup>9</sup> Process

<sup>10</sup> Product

جدول محتوا- هدف را می‌توان به دو صورت نگارش کرد:

- (۱) در جدول محتوا- هدف، نسبت یا درصد یا ضریب هر هدف را با توجه به محتوایش در خانه‌های جدول نگارش کنیم.
- (۲) در جدول محتوا- هدف شماره سؤال‌ها یا تعداد سؤال‌ها را در خانه‌های جدول نگارش کنیم. [۳]

شکل شماره دو: جدول محتوا - هدف

				محتوا هدف
D	C	B	A	1
				2
				3
				4
				5

نکته: جدول هدف - محتوا (مشخصات) برای هر موضوع درسی تهیه می‌گردد.

### گام پنجم

گام پنجم فرآیند ارزشیابی تهیه و تنظیم و کنترل سؤالات است.

هدف طراح سؤال، از طرح سؤال سه نوع ارزشیابی می‌باشد:

(۱) ارزشیابی ورودی

(۲) ارزشیابی تکوینی (مرحله ای)

(۳) ارزشیابی نهایی (پایانی)

- **ارزشیابی ورودی:** برای تعیین اولین نیازهای آموزش برای شروع درس می‌باشد. آزمون جایابی نوعی ارزشیابی ورودی است.

**ارزشیابی تکوینی:** برای تعیین اثربخش آموزش می‌باشد و نشان می‌دهد که آیا معلم به قسمت بعدی برود یا خیر. معمولاً ارزشیابی

تکوینی به صورت شفاهی اجرا می‌شود. همینطور تعیین می‌کند که آیا برنامه آموزشی دارای نواقصی می‌باشد یا خیر؟

- **ارزشیابی نهایی:** با دو هدف صورت می‌گیرد؛ اول اینکه، برای تصمیم‌گیری درباره قبولی یا ردی دانش آموز است. دیگر آن که،

تصمیم در مورد مفید بودن یا نبودن برنامه تدریس می‌باشد. [۳]

نکته: امتحانات کوئیز نوعی ارزشیابی تکوینی می‌باشد.

نوع دیگری ارزشیابی وجود دارد به نام ارزشیابی تشخیصی. این ارزشیابی زمانی به کار می‌رود که معلم با مشکلات متعددی در یک یا

چند دانش آموز مواجهه می‌شود که با روش‌های اصلاحی ارزشیابی تکوینی قابل رفع نیستند. این نوع ارزشیابی در پی علت‌های زیر

بنایی مشکلات دانش آموزان است. [۳]

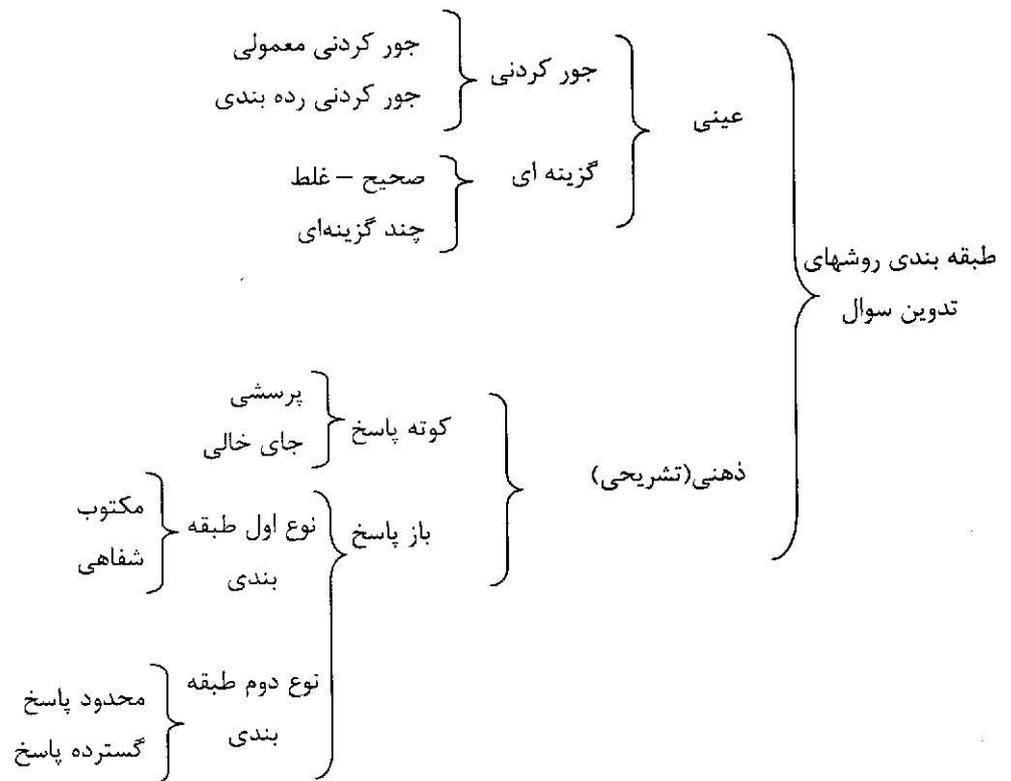
نکته: تفاوت آزمون‌های تشخیصی با آزمون‌های پیشرفت تحصیلی در آن است که آزمون‌های پیشرفت تحصیلی، توفیق دانش آموزان را در درس‌های مختلف می‌سنجد، اما آزمون‌های تشخیصی توصیف کاملی از نقاط ضعف و قوت یادگیرندگان در زمینه‌های مهارت‌های خاص مثل خواندن، نوشتن، محاسبه کردن و جزآن را به دست می‌دهند. [۳]

نکته: ارزشیابی تشخیصی، علاوه بر تشخیص مشکلات دانش آموزان، باید بتواند روش‌های مناسب رفع مشکلات را نیز به معلمان معرفی نماید.

نکته: سؤال‌های ارزشیابی تکوینی هم می‌تواند شبیه ارزشیابی نهایی باشد و یا اختصاصاً تکوینی باشد.

## روش‌های تدوین سؤال

روش‌های تدوین سؤال را می‌توان به صورت زیر طبقه بندی نمود:



در سؤالات عینی؛ طراح سؤال هم سؤال را طرح می‌کند و هم پاسخ را تنظیم می‌کند و پاسخگو انتخاب می‌کند.

در سؤالات تشریحی؛ طراح سؤال، سؤالات را تنظیم می‌کند و پاسخگو پاسخ را طراحی می‌کند. [۴]

نکته: طراحی سؤالات تشریحی آسان است ولی طراحی سؤالات عینی سخت است.

نکته: در آزمون‌های کوتاه پاسخ از این جهت که پاسخ به وسیله پاسخگو تدوین می‌شود، تشریحی تلقی می‌شود و چون امکان تصحیح کمی به مصحح می‌دهد، عینی تلقی می‌شود.

نکته: در سؤالات تشریحی پاسخگو به سه نکته باید توجه کند: (۱) سازمان دادن به اطلاعات، (۲) صحت اطلاعات، (۳) خلاقیت.

نکته: در آزمون‌های کوتاه پاسخ باید توجه داشت که تعداد جای خالی نباید زیاد باشد.

نکته: آزمون‌های کوتاه پاسخ بیشتر در سطح سنجش دانش و فهمیدن قابل استفاده است.

نکته: آزمون‌های تشریحی بیشتر در سطح سنجش کاربستن و بالاتر مناسب است.

نکته: آزمون‌های محدود پاسخ، پاسخگو برای پاسخگویی محدودیت دارد.

نکته: آزمون‌های گسترده پاسخ، پاسخگو برای پاسخگویی هیچگونه محدودیتی ندارد.

## قواعد تهیه سؤالات انشایی یا تشریحی

(۱) سؤال‌ها به طور مستقیم به هدف‌های آموزش مربوط شوند.

(۲) سؤال‌های تشریحی مربوط به اندازه‌گیری هدف‌هایی باید باشد که با روش‌های دیگر قابل اندازه‌گیری نباشند.

(۳) از کلی‌گویی و ابهام در بیان سؤالات تشریحی 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) به آزمون‌شوندگان حق انتخاب چند سؤال از میان تعدادی سؤال را ندهید.

(۷) برای پاسخ دادن به سؤال‌ها زمان کافی را مشخص کنید. [۳]

نکته: در صورتی که به آزمون‌شوندگان حق انتخاب چند سؤال از میان تعدادی سؤال داده شود، بر تفسیر ملاکی ارزشیابی تأثیر می‌گذارد.

در آزمون‌های عینی جور کردنی معمولی: تعدادی سؤال و تعدادی پاسخ وجود دارند که پاسخگو باید سؤال را به پاسخ با یک خط وصل نماید.

در آزمون‌های عینی جور کردنی رده بندی: تعداد زیادی سؤال و تعداد کمی پاسخ وجود دارند که پاسخگو باید برای هر پاسخ چند سؤال را تعیین کند. [۳]

در آزمون‌های عینی گزینه‌ای صحیح - غلط سه روش استفاده می‌شود: (۱) صحیح- غلط، (۲) بله- خیر، (۳) خوشه‌ای .

در حالت اول صحیح- غلط: جملات خبری نوشته می‌شود.

در حالت دوم بله- خیر: سؤالات به صورت پرسش نگارش می‌شوند.

در حالت سوم خوشه‌ای: تنه اصلی سؤال می‌آید و چندین پاسخ دارد که باید برای هر کدام صحیح- غلط را مشخص کنیم. [۳]

نکته: سؤالات صحیح - غلط حداکثر در سطح دانش و فهمیدن قابل استفاده است.

نکته: مهم‌ترین ایراد سؤالات صحیح - غلط وجود احتمال ۵۰ درصد شانس پاسخ صحیح است.

## قواعد تهیه سؤال‌های جور کردنی

- (۱) پرسش‌ها و پاسخ‌های متجانس یا همگون انتخاب کنید.
- (۲) طول فهرست‌های پرسش و پاسخ را کوتاه انتخاب کنید.
- (۳) در صورت امکان فهرست پاسخ‌ها را به طور منطقی مرتب کنید تا در وقت پاسخگویی صرفه جویی شود.
- (۴) در فهرست پرسش‌های سؤالات جور کردنی از کاربرد جملات نیمه تمام خودداری کنید.
- (۵) هر یک از پاسخ‌ها باید برای تمام پرسش‌های یک سؤال درست جلوه کند.

## قواعد تهیه سؤال‌های صحیح - غلط

- (۱) سؤال‌ها را تا حد امکان مختصر، ساده و روشن بنویسید.
  - (۲) از کاربرد کلماتی چون همه، بعضی وقت‌ها، غالباً، هرگز و از این قبیل بپرهیزید.
  - (۳) سعی کنید طول سؤال‌های صحیح و سؤال‌های غلط هم اندازه باشد.
  - (۴) سعی کنید تعداد سؤال‌های صحیح با تعداد سؤال‌های غلط تقریباً مساوی باشد.
  - (۵) سؤال‌های صحیح - غلط را از روی جملات کتاب بنویسید.
  - (۶) سعی کنید هر سؤال مربوط به یک هدف آموزشی باشد.
- معمولاً برای جبران مشکل سؤال صحیح - غلط، از سؤالات چند گزینه‌ای استفاده می‌کنند که مشکل‌ترین بخش آن طراحی گزینه‌ی انحرافی و غلط است که شبیه پاسخ درست باشد. [۳]
- در سؤالات چند گزینه‌ای بین ۳ تا ۵ گزینه می‌آید.

نکته: هر چه قدر تعداد گزینه بیشتر باشد احتمال پاسخ شانسی کمتر می‌شود.

سؤالات چند گزینه‌ای تا سطح کار بستن قابل استفاده است. اگرچه، با افزایش توانایی طراح سؤال تا سطح ترکیب هم می‌تواند گسترش یابد.

نکته: از طریق پاسخ‌های تشریحی گروهی دیگر از پاسخ دهندگان به سؤال می‌توان گزینه‌های انحرافی را تولید کرد.

## قواعد تهیه سؤال‌های چند گزینه‌ای

- (۱) هر سؤال باید به یک هدف آموزشی مربوط باشد.
- (۲) از تکرار مطالب در گزینه‌ها خودداری کنید.
- (۳) مطالب اصلی سؤال را به طور کامل در تنه سؤال بنویسید.
- (۴) همه گزینه‌های یک سؤال باید متجانس و به موضوع واحدی مربوط باشند.
- (۵) سؤال را طوری بنویسید که پاسخ درست قطعاً درست‌ترین پاسخ باشد.
- (۶) گزینه‌های انحرافی را طوری بنویسید که توجه آزمون شوندگان بی اطلاع از موضوع را به خود جلب کند.
- (۷) در سؤال‌های منفی کلمات منفی را برجسته کنید.
- (۸) از نوشتن سؤال‌هایی که در آن‌ها متن سؤال منفی و گزینه‌ها هم منفی هستند، احتراز کنید.

- ۹) از کاربرد عبارت هایی مانند، تمامی موارد فوق، همه موارد، هیچ کدام و جز آن بپرهیزید.  
 ۱۰) محل گزینه درست را در میان گزینه های انحرافی به طور تصادفی انتخاب کنید.  
 ۱۱) در سؤال هایی که تنه آن ها جمله ناتمام است، جای خالی را در قسمت آخر جمله قرار دهید. [۳]

نکته: امتیاز سؤال تشریحی، ارزیابی عمیق ترین اهداف می باشد.

نکته: انتقاد سؤال تشریحی، اطلاعات کمی را می تواند مورد ارزیابی قرار دهد.

نکته: امتیاز سؤال عینی، بیشترین اهداف آموزشی را مورد ارزیابی قرار می دهد.

نکته: انتقاد سؤال عینی، اهداف عمیق را نمی توان با آن بررسی کرد و شانس در آن دخالت دارد.

### گام ششم

گام ششم فرآیند ارزشیابی تصحیح و نمره گذاری می باشد.

### روش تصحیح آزمون های تشریحی (انشایی)

۱) تصحیح تحلیلی (بارم بندی)، ۲) تصحیح کلی.

تصحیح تحلیلی یا بarm بندی بدین صورت است که هر یک از اجزاء پاسخ را مشخص نموده و میزان معینی از نمره را برای آن در نظر می گیریم. این روش تصحیح برای آزمون های محدود پاسخ مناسب است.

مقایسه سؤالات عینی (غیرذهنی) و ذهنی (تشریحی یا انشایی)

سؤالات تشریحی	سؤالات عینی
۱. برای اندازه گیری فرایندهای یادگیری در سطح ترکیب و ارزشیابی مناسب است. (برای سایر زمینه ها به جز دانش هم استفاده می شود).	۱. برای اندازه گیری فرایندهای یادگیری در سطح دانش، فهمیدن، کار بستن و تحلیل مناسب است.
۲. دانش آموز باید پاسخ سؤال را تهیه و در تهیه ی آن تا حد زیادی آزادی عمل دارد.	۲. پاسخ ها کنترل شده، دانش آموز پاسخ درست را فقط علامت می زند. (بازشناسی)
۳. نمونه ی ناقصی از موارد آموزشی را ارزیابی می کند.	۳. نمونه گسترده ای از موارد آموزش داده شده را ارزیابی می کند.
۴. عواملی مانند تسلط بر بیان مطلب توانایی نویسندگی و خلاقیت در ارزشیابی آن مؤثر است.	۴. عواملی مانند تسلط بر بیان مطلب، توانایی نویسندگی و زیبا نویسی در ارزشیابی آن بی اثر است.
۵. عامل حدس و گمان در آن دخالتی ندارد.	۵. عامل حدس و گمان در آن مؤثر است.
۶. نمره گذاری و ارزشیابی ذهنی است. (تأثیر مصحح در تصحیح)	۶. نمره گذاری و ارزشیابی پاسخ ها عینی است.
۷. ثبات و دقت کمتر نمرات امتحانی	۷. ثبات و دقت بیشتر نمرات امتحانی
۸. تهیه ی سؤالات آسان تر و صرف وقت کمتر	۸. تهیه ی سؤالات وقت گیر و دشوار
۹. تصحیح و نمره گذاری مستلزم وقت زیاد	۹. تصحیح و نمره گذاری با سرعت و سهولت
۱۰. تقویت عادات صحیح مطالعه در همه زمینه ها به خصوص ترکیب و ارزشیابی	۱۰. تقویت عادات صحیح مطالعه مانند: یادآوری، تجزیه، تحلیل، تعبیر و تفسیر

تصحیح کلی بدین شکل انجام می‌شود که مصحح تمام پاسخ را خوانده و یک درجه کلی برای آن در نظر می‌گیرد. این روش برای امتحانات شفاهی تشریحی مناسب است. [۳]

**نکته:** برای آزمون‌های کتبی تشریحی، هر دو روش تصحیح قابل استفاده است. روش کلی برای سازمان بندی و خلاقیت و روش تحلیلی برای مفاهیم پاسخ.

**نکته:** تأثیر نظر مصحح در میزان نمره آزمون شوندگان به همین دلیل آزمون‌های ذهنی نامیده می‌شوند.

**نکته:** روش نمره گذاری آزمون‌های کوتاه پاسخ، بر اساس صفر و یک است.

**نکته:** روش نمره گذاری آزمون‌های تشریحی، بر اساس پیوستار است.

دو دلیل اصلی ذهنی بودن نمره گذاری آزمون‌های تشریحی

(۱) تصحیح کاملاً در اختیار مصحح است. (خطای تصحیح)

(۲) سؤال کاملاً برای آزمونگر است و پاسخ کاملاً برای پاسخگو.

**نکته:** مناسب‌ترین از لحاظ آسانی و دقت سیستم نمره گذاری صفر تا بیست می‌باشد.

**نکته:** اصطلاح لنگرگاه در نمره گذاری اوراق انشایی عبارت از ورقه‌ای است که به دانش آموز خیلی قوی تعلق دارد.

**نکته:** در تصحیح اوراق انشایی، تعویض دائمی معیار در طول تصحیح اوراق، سرایت معیار نام دارد.

**نکته:** اثر هاله‌ای عبارت است از: یک ویژگی برجسته، سایر خصوصیات فرد آزمون شونده را تحت الشعاع قرار می‌دهد.

**نکته:** اثر پیگمالیون عبارت است از: اثر انتظارات معلم بر موفقیت دانش آموزان خود.

### روش‌های مورد استفاده برای کاهش خطای تصحیح:

(۱) مصحح همه ورقه‌ها را سؤال به سؤال تصحیح کند.

(۲) از مصحح دیگری برای تصحیح برگه‌ها برای بار دوم استفاده کند. به عبارت دیگر استفاده از بیش از یک مصحح. [۳]

**نکته:** هرچه میزان خطای تصحیح کاهش یابد اعتبار تصحیح افزایش می‌یابد.

اعتبار تصحیح میزانی است که نشان دهنده کاهش مقدار خطای تصحیح می‌باشد.

## روش محاسبه اعتبار تصحیح

این روش میزان اشتراک بین مصححین است.

$$\text{اعتبار تصحیح} = 100 - \frac{\text{اختلاف مصححین}}{\text{حداکثر نمره اختلاف}} \times 100$$

مثال: مصحح اول ۸ نمره خطا گرفته و مصحح دوم ۱۲ نمره سیستم نمره گذاری صفر تا بیست است، اعتبار تصحیح را به دست آورید.

$$\text{اعتبار تصحیح} = 100 - \left[ \left( \frac{12-8}{20} \right) \times 100 \right] = 80\%$$

نکته: حداکثر نمره اختلاف در سیستم نمره گذاری صفر تا بیست، بیست می‌باشد.

نکته: اگر سه نفر مصحح بود، میانگین اختلاف‌ها را به دست می‌آوریم.

$$x = \text{مصحح دوم} - \text{مصحح اول}$$

$$\text{اختلاف مصححین} = \frac{x+y+z}{3}$$

$y = \text{مصحح سوم} - \text{مصحح اول}$

$$z = \text{مصحح سوم} - \text{مصحح دوم}$$

## روش تصحیح و نمره گذاری آزمون‌های عینی

در این آزمون‌ها نظر مصحح در تصحیح دخالت ندارد. نوع نمرات بدست آمده در این آزمون‌ها صفر و یک است.

مهم‌ترین ایراد آزمون‌های عینی، عامل حدس و گمان است.

فرمول تصحیح با حذف عامل حدس برای سؤالات صحیح - غلط

$$S = R - W \quad \text{نمره} = {}^{11}S$$

$$R = \text{تعداد سؤالات صحیح} = {}^{12}R$$

$$W = \text{تعداد سؤالات غلط} = {}^{13}W$$

مثال: در یک آزمون صحیح- غلط فردی به ۳۰ سؤال پاسخ صحیح داده است و ۱۲ سؤال را غلط پاسخ داده است. نمره تصحیح شده او چقدر است؟

$$S = 30 - 12 = 18$$

فرمول تصحیح با حذف عامل حدس برای سؤال‌های چهار گزینه‌ای

$$K = \text{تعداد گزینه‌ها}$$

<sup>11</sup> Score

<sup>12</sup> Right

<sup>13</sup> Wrong

مثال: در یک آزمون پنج گزینه‌ای، فردی به ۵۵ سؤال پاسخ داده که ۳۵ سؤال صحیح و ۱۸ سؤال غلط است. نمره تصحیح شده او چقدر است؟

$$S = 35 - \frac{18}{4} = 30.5$$

نکته: اجرای آزمون‌های عینی به متخصص نیاز ندارد.

نکته: اجرای آزمون‌های تشریحی شفاهی به متخصص نیاز دارد ولی در آزمون‌های تشریحی کتبی به متخصص نیاز ندارد.

## روش‌های تصحیح، اجرا و نمره گذاری آزمون‌های عملکردی (مربوط به حیطه‌های غیر شناختی)

### انواع آزمون‌های عملکردی و سطح مورد سنجش آن‌ها:

- (۱) آزمون‌های شناسایی  
حداکثر در سطح مشاهده و تقلید را می‌سنجند.
- (۲) آزمون‌های کتبی
- (۳) آزمون انجام عملکرد در موقعیت‌های شبیه سازی شده ← حداکثر تا سطح اجرای مستقل را می‌سنجد.
- (۴) آزمون‌های نمونه ی کار ← حداکثر تا سطح دقت را می‌سنجد.

در آزمون‌های فرآیندی هدف تا سطح اجرای مستقل می‌باشد، مانند: آزمون شناسایی، آزمون کتبی، آزمون انجام عملکرد در آزمون‌های فرآورده‌ای هدف تا سطح دقت می‌باشد. مانند: آزمون اجرای نمونه ی کار در آزمون‌های فرآیندی تشخیص اینکه چه موقع باید به مرحله ی بعدی رفت، اهمیت دارد ولی در آزمون‌های فرآورده‌ای نتیجه ی عمل و دقت در اجرای نمونه ی کار اهمیت دارد.

در آزمون‌های کتبی از پاسخگو می‌خواهیم یک مهارت را بنویسید و مهم‌ترین عمل این آزمون‌ها سرند کردن است. [۳]

## روش پوشه کار (پورت فولیو)

نوع جدید آزمون‌های عملکردی است، مجموعه‌ای از کارهای دانش آموزان که نشان دهنده کوشش، پیشرفت و موفقیت تحصیلی او می‌باشد و کارهای منتخب او در یک مدت معین می‌باشد. حداکثر تا سطح هماهنگی را می‌تواند بسنجد. [۳]

نکته: در روش پورت فولیو معلم ضمن ارزشیابی، آموزش نیز می‌دهد.

استفاده از پوشه کار دو هدف عمده دارد: (۱) ارائه بهترین کارها، (۲) نشان دادن رشد تحصیلی یادگیرنده.

نکته: روش پورت فولیو را در حیطه ی شناختی هم می‌توان استفاده کرد.

## روش‌های سنجش مشاهده‌ای

- (۱) فهرست واریسی
- (۲) واقعه نگاری (ثبت رویداد)
- (۳) مقیاس درجه بندی

### (۱) فهرست واریسی

در این روش قضاوت‌های مشاهده گر بودن یا نبودن ویژگی مورد نظر است و حداکثر تا سطح اجرای مستقل مورد استفاده می‌باشد. استفاده از روش فهرست واریسی در سنجش فرآیندی عملکرد و فرآورده عملکرد.  
در سنجش فرایند عملکرد ← تأکید بر رفتارهای فرد (مثلاً ترتیب انجام مراحل)  
در سنجش فرآورده‌ای ← ویژگی‌های محصول به صورت فهرستی از ابعاد محصول خوب (اندازه، رنگ، شکل ...)

### روش تهیه فهرست واریسی

فهرستی از اعمال وصف شده در عملکرد را مشخص کنید و غلط‌های احتمالی را به ترتیبی که انتظار وقوعشان می‌رود را هم به فهرست اضافه کنید. (به منظور تشخیص اشکال‌های عملکرد یا محصول)

### (۲) واقعه نگاری (ثبت رویداد)

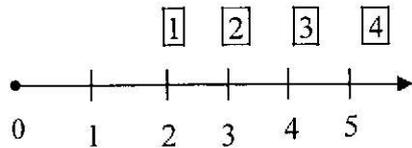
ثبت وقایع زندگی بلافاصله بعد از وقوع، بیشتر جنبه تحلیلی دارد و میزان آن در سطح دقت حداکثر تا سطح هماهنگی می‌باشد.  
**تعریف:** توصیف‌های واقعی از رویدادها و اتفاقات معنی داری که معلم (مشاهده گر) در نتیجه مشاهده از زندگی دانش آموزان به دست می‌آورد.  
← بیشتر برای هدف‌های آموزشی حوزه‌های عاطفی و روانی حرکتی استفاده می‌شود.  
**مثال:** ثبت چگونگی برخورد دانش آموز با مسائل یا پشت کار دانش آموز در ادامه کار.  
← بهترین کاربرد این روش: ثبت و ارزشیابی رفتارها و ویژگی‌هایی از دانش آموز که با روش‌های متداول قابل اندازه‌گیری نیست.

### (۳) روش مقیاس درجه بندی

تعیین درجه یا مقدار رفتار یا ویژگی مورد نظر.  
قابل استفاده برای ارزشیابی از فرایند عملکرد (شیوه اجرا)  
فرآورده عمل (محصول)  
حداکثر تا سطح اجرای مستقل را مورد سنجش قرار می‌دهد.  
← تفاوت روش مقیاس درجه بندی و فهرست واریسی:

در روش فهرست واریسی «بود یا نبود» ویژگی اهمیت داشت ولی در این مقیاس کارهای انجام شده را درجه بندی می‌کنیم.

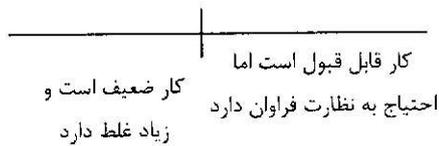
## انواع مقیاس درجه بندی:



(۱) مقیاس درجه بندی عددی  
 تفکیکی  
 پیوستاری



(۲) نگاره‌ای (درجه بندی یکنواخت)  
 مقیاس درجه بندی یکنواخت  
 مقیاس درجه بندی متغیر



(۳) نگاره‌ای توصیفی (نقاط روی مقیاس توصیف می‌شوند).

## سنجش هدف‌های حوزه عاطفی

### طبقه بندی کراتول از حوزه عاطفی

- توجه / دریافت کردن
- پاسخ دادن
- ارزشگذاری
- سازمان دادن ارزش‌ها
- تشخیص طلبی

### فنون سنجش هدف‌ها و رفتارهای عاطفی

(۱) پرسشنامه بسته پاسخ (پاسخ بایست)، متن سؤال ← خبری  
 ← سؤال

شبهه آزمون‌های عینی در حوزه ی شناختی

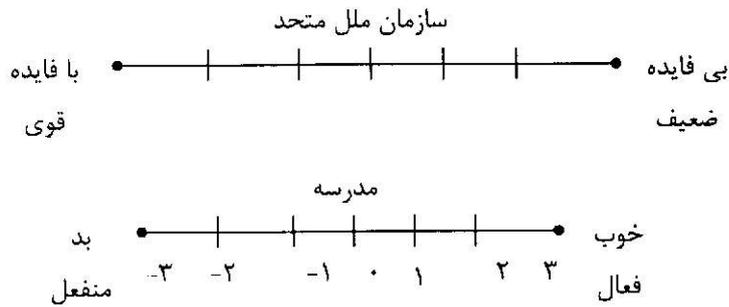
(۲) آزمون‌های باز پاسخ، سؤال‌ها به صورت جملات ناتمام

شبهه آزمون‌های تشریحی در حوزه شناختی ← تا سطح پاسخ دادن را می‌سنجند.

(۳) روش تفکیک معنایی (برای اندازه‌گیری معنای ضمنی مفاهیم).

شامل تعدادی صفت دو قطبی (بد- خوب) روی یک پیوستار هفت بخشی و موضوع نگرش در بالای صفحه نوشته می‌شود.

← تا سطح ارزشگذاری را می‌تواند بسنجد.



(۴) روش یادداشت‌های روزانه

برای سازمان دادن ارزش‌ها، مفید است. [۴]

## ضمیمه آمار

### نمرات استاندارد (نمرات Z، نمرات تراز)

اصل نمرات استاندارد برای مقایسه کردن است. مقایسه بین دو نمره. در صورتی که توزیع نمرات ما نرمال باشد، از فرمول زیر استفاده می‌کنیم.

$$Z = \frac{x - \bar{x}}{S}$$

به این استاندارد نمره تراز هم می‌گوییم.

مثال: اگر علی نمره فیزیکش ۱۲ باشد، معدلش ۱۴ و انحراف معیار ۲ باشد، نمره تراز آن چقدر است؟

$$Z = \frac{12 - 14}{2} = -1$$

نمره تراز در اصل برای مقایسه کردن است.

مثال: فردی دارای نمره همتراز در دو درس می‌باشد، به شرط آن که میانگین درس A ۱۵، و انحراف معیار ۳ باشد. و میانگین درس B، ۵ و انحراف معیار ۲ باشد، نمره ۱۸ از درس A با کدام نمره از درس B همتراز است؟

در چنین شرایطی دو فرمول Z داریم که همتراز هم هستند.

$$\frac{x_A - \bar{x}_A}{S_A} = \frac{x_B - \bar{x}_B}{S_B}$$

$$\frac{18 - 15}{3} = \frac{x - 5}{2} \Rightarrow 1 = \frac{x - 5}{2} \Rightarrow 2 = x - 5 \Rightarrow x = 7$$

اگر نمره خام ما  $\bar{x}$  باشد، مقدار Z چقدر می‌شود؟

- اگر  $x = \bar{x}$  باشد، در این صورت  $Z = 0$  می‌شود.

- اگر  $x = \bar{x} + 1S$  باشد، در این صورت  $Z = 1$  می‌شود.

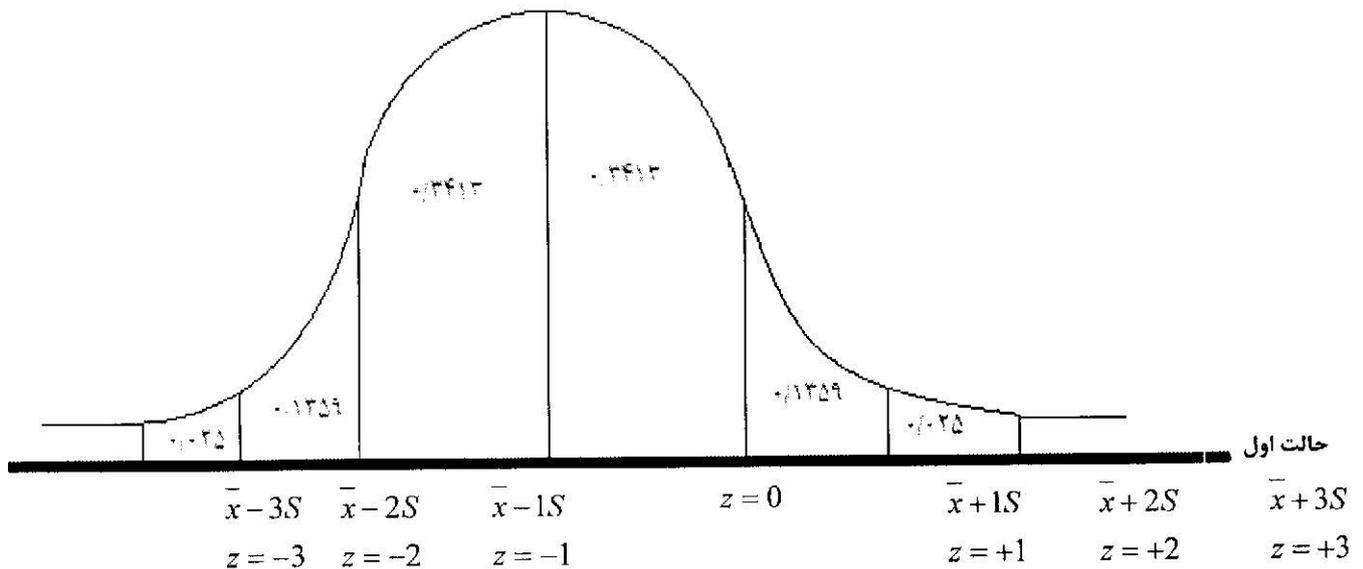
- اگر  $x = \bar{x} - 1S$  باشد، در این صورت  $Z = -1$  می‌شود.

$$z = \frac{(\bar{x} + 1S) - \bar{x}}{S} = 1 \Rightarrow$$

- اگر  $x = \bar{x} + 2S$  باشد، در این صورت  $Z = 2$  می‌شود.

- اگر  $x = \bar{x} - 2S$  باشد، در این صورت  $Z = -2$  می‌شود.

۱۲ حالت وجود دارد.



$Z$  را بدهند، سطح را بخواهند. (ممکن است بگویند احتمال و نسبت که همان معنی را می‌دهد).

**حالت اول:** دو تا  $Z$  بدهند، سطح بین دو تا  $Z$  را بخواهند.

**حالت دوم:** یک  $Z$  بدهند، سطح این طرف و آن طرفش را بخواهند.

مثال: چه نسبتی از نمره‌ها بیشتر از  $Z=1$  است؟

$$0.1359 + 0.025 = 0.16$$

مثال: چه نسبتی (سطحی، احتمالی) از نمره‌ها کمتر  $Z=1$  است؟

$$0.3413 + 0.3413 + 0.1359 + 0.025 = 0.84$$

مثال: چه نسبتی بین  $Z = \pm 1$  قرار دارد؟

$$0.3413 + 0.3413 = 0.68$$

### حالت دوم

ممکن است  $Z$  را بدهند، درصد را بخواهند.

که در این حالت، از روی  $Z$  سطح را به دست می‌آوریم، بعد ضرب در ۱۰۰ می‌کنیم و درصدش را به دست می‌آوریم.

مثال: چند درصد از نمره‌ها بیشتر از  $Z=1$  است؟

$$0.1359 + 0.025 = 0.16 \times 100 = 16\%$$

مثال: چند درصد از نمره‌ها کمتر از  $Z=1$  است؟

$$0.3413 + 0.3413 + 0.1359 + 0.025 = 0.84 \times 100 = 84\%$$

### حالت سوم

$Z$  را می‌دهند و از ما تعداد می‌خواهند، البته در چنین شرایطی حتماً  $n$  را به ما می‌دهند. که در این حالت اول از  $Z$  سطح را به دست می‌آوریم، بعد سطح را ضرب در  $n$  می‌کنیم، تعداد به دست می‌آید.

مثال: در نمونه‌ای به حجم ۵۰۰ نفر چه تعداد  $Z$  بیشتر از ۱ گرفتند؟

$$0.1359 + 0.025 = 0.16 \times 500 = 80$$

مثال: در نمونه‌ای به حجم ۵۰۰ نفر چه تعداد  $Z$  کمتر از ۱ گرفتند؟

$$0.3413 + 0.3413 + 0.1359 + 0.025 = 0.84 \times 500 = 420$$

مثال: در نمونه‌ای به حجم ۵۰۰ نفر چه تعداد بین  $Z = \pm 1$  گرفتند؟

$$0.3413 + 0.3413 = 0.68 \times 500 = 340$$

### حالت چهارم

برعکس حالت اول است، سطح را بدهند،  $Z$  را بخواهند.

مثال: ۰.۱۶ بالایی چه نمره  $Z$  دارد؟

$$0.025 + 0.1359 + 0.3413 + 0.3413 = 0.84 = Z = -1$$

مثال: ۰.۸۴ بالایی چه نمره  $Z$  دارد؟

$$0.025 + 0.1359 + 0.3413 + 0.3413 = 0.84 = Z = -1$$

مثال: ۰.۶۸ وسطی چه نمره  $Z$  دارد؟

یعنی تقسیم بر ۲ می‌کنیم، که می‌شود ۰.۳۴، که همان  $Z = \pm 1$  است، می‌شود.

### حالت پنجم

برعکس حالت دوم است، درصد می‌دهند  $Z$  را می‌خواهند. که در صد را تقسیم بر ۱۰۰ می‌کنیم، بعد سطح را به دست می‌آوریم، از روی سطح،  $Z$  را به دست می‌آوریم.

مثال: ۱۶٪ بالای  $Z$  چقدر است؟

$$0.16 \div 100 = 0.16 = (0.1359 + 0.025) Z = 1$$

مثال: ۸۴٪ بالای  $Z$  چقدر است؟

$$0.84 \div 100 = 0.84 = (0.3413 + 0.3413 + 0.1359 + 0.025) = Z = -1$$

مثال: ۶۸٪ وسطی Z مساوی با چند است؟

$$.68 \div 100 = .068 \div 2 = .034 = Z = \pm 1$$

### حالت ششم

برعکس حالت سوم است، تعداد را می‌دهند، Z را می‌خواهند. تعداد را تقسیم بر n می‌کنیم، سطح به دست می‌آید. از روی سطح Z به دست می‌آید.

مثال: در نمونه‌ای به حجم ۵۰۰ نفر، نمره Z ۸۰ نفر بالایی چقدر است؟

$$\frac{80}{500} = 0.16 \Rightarrow z = 1$$

مثال: در جامعه‌ای به حجم ۵۰۰ نفر، ۴۷۵ نفر وسطی دارای چه Z یا Z هایی می‌باشند؟

$$\frac{475}{500} = 0.95 \div 2 = 0.475 \Rightarrow z = \pm 2$$

مثال:

Z می‌تواند منفی باشد.

شاخص‌های مرکزی می‌توانند منفی باشند.

سطح نمی‌تواند منفی باشد.

شاخص‌های پراکندگی نمی‌توانند منفی باشند.

### حالت هفتم

به ما نمره خام می‌دهند، از ما سطح می‌خواهند. در چنین شرایطی، البته میانگین و انحراف معیار هم می‌دهند، نمره خام را با استفاده از فرمول به نمره Z تبدیل می‌کنیم. بعد از Z سطح را به دست می‌آوریم.

مثال: اگر  $\bar{x} = 15$  و  $S = 3$  باشد، چه نسبتی (سطحی) از افراد بالای ۱۸ گرفته اند؟

$$z = \frac{x - \bar{x}}{S} = \frac{18 - 15}{3} = 1 \Rightarrow z = 1$$
$$0.025 + 0.1359 = 0.16$$

مثال: اگر  $\bar{x} = 15$  و  $S = 3$  باشد، چه نسبتی از نمره‌ها بین نمره ۱۸ و ۱۲ قرار گرفتند؟

$$z_1 = \frac{18 - 15}{3} = 1$$

$$z_2 = \frac{12 - 15}{3} = -1$$

$$z = \pm 1$$

$$0.3413 + 0.3413 = 0.6826$$

### حالت هشتم

نمره خام را می‌دهند، از ما درصد می‌خواهند، البته میانگین و انحراف معیار هم حتماً به ما می‌دهند. در چنین حالتی از روی  $Z$  سطح را به دست می‌آوریم، سطح را ضرب در ۱۰۰ می‌کنیم، درصد به دست می‌آید.

مثال، اگر  $\bar{x} = 15$  و  $S = 3$  باشد، چه درصدی از افراد بالای ۱۸ گرفته اند؟

$$z = \frac{x - \bar{x}}{S} = \frac{18 - 15}{3} = 1 \Rightarrow z = 0.16$$

$$0.16 \times 100 = 16\%$$

### حالت نهم

ممکن است به ما نمره خام را بدهند از ما تعداد بخواهند. حتماً  $\bar{x}$  و  $S$  و  $n$  را هم به ما می‌دهند. از نمره خام، نمره  $Z$  را پیدا می‌کنیم، با استفاده از فرمول، از نمره  $Z$  سطح را به دست می‌آوریم. بعد سطح را ضرب در  $n$  می‌کنیم، تعداد به دست می‌آید.

مثال: در نمونه‌ای به حجم ۵۰۰ در صورتی که  $\bar{x} = 15$  و  $S = 3$  باشد، چه تعداد نمره‌ای بیشتر از ۱۸ گرفته اند؟

$$z = \frac{x - \bar{x}}{S} = \frac{18 - 15}{3} = 1 \Rightarrow z = 1$$

$$0.025 + 0.1359 = 0.16 \times 500 = 80$$

مثال: با توجه به اطلاعات قبل چه تعداد از افراد بین نمره ۱۸ و ۱۲ گرفته اند؟

$$z = \frac{18 - 15}{3} = 1$$

$$z = \frac{12 - 15}{3} = -1$$

$$z = \pm 1$$

$$0.3413 + 0.3413 = 0.68 \times 500 = 340$$

### حالت دهم

برعکس حالت هفتم است، به ما سطح را می‌دهند و از ما نمره خام می‌خواهند. که ما سطح را به  $Z$  تبدیل کرده، سپس در فرمول گذاشته و نمره خام را به دست می‌آوریم.

مثال: اگر  $\bar{x} = 15$  و  $S = 3$  باشد، حداقل نمره ۰،۱۶ بالایی چقدر است؟

$$z = 1$$

$$z = \frac{x - \bar{x}}{S}$$

$$1 = \frac{x - 15}{3} \Rightarrow 3 = x - 15 \Rightarrow x = 18$$

مثال: اگر  $\bar{x} = 15$  و  $S = 3$  باشد، ۰،۶۸ وسطی بین چه نمراتی قرار دارد؟

$$0.68 = \pm 1$$

$$1 = \frac{x_1 - 15}{3} \Rightarrow 3 = x_1 - 15 \Rightarrow x_1 = 18$$

$$-1 = \frac{x_2 - 15}{3} \Rightarrow -3 = x_2 - 15 \Rightarrow x_2 = 12$$

$$12 < x < 18$$

### حالت یازدهم

برعکس حالت هشتم است. به ما درصد می دهند و از ما نمره خام می خواهند. که ما درصد را تقسیم بر ۱۰۰ می کنیم، سطح را به دست می آوریم. از سطح  $Z$  را به دست می آوریم. و از  $Z$  نمره خام را به دست می آوریم.

مثال: اگر  $\bar{x} = 15$  و  $S = 3$  باشد، حداقل نمره ۱۶٪ بالایی چقدر است؟

$$16\% \div 100 = 0.16 \Rightarrow z = 1$$

$$z = \frac{x - \bar{x}}{S} \Rightarrow 1 = \frac{x - 15}{3} \Rightarrow 3 = x - 15 \Rightarrow x = 18$$

### حالت دوازدهم

برعکس حالت نهم است. تعداد را می دهند، نمرات خام را می خواهند.

مثال: در نمونه‌ای به حجم ۵۰۰ نفر، نمره ۴۷۵ نفر وسطی چیست؟ به شرط آن که  $S = 3$  و  $\bar{x} = 15$  باشد.

تعداد را تقسیم بر  $n$  می کنیم، سطح به دست می آید، از روی سطح  $Z$  به دست می آید و از روی  $Z$  نمره خام را به دست می آوریم.

$$\frac{475}{500} = 0.95$$

$$z = \pm 2$$

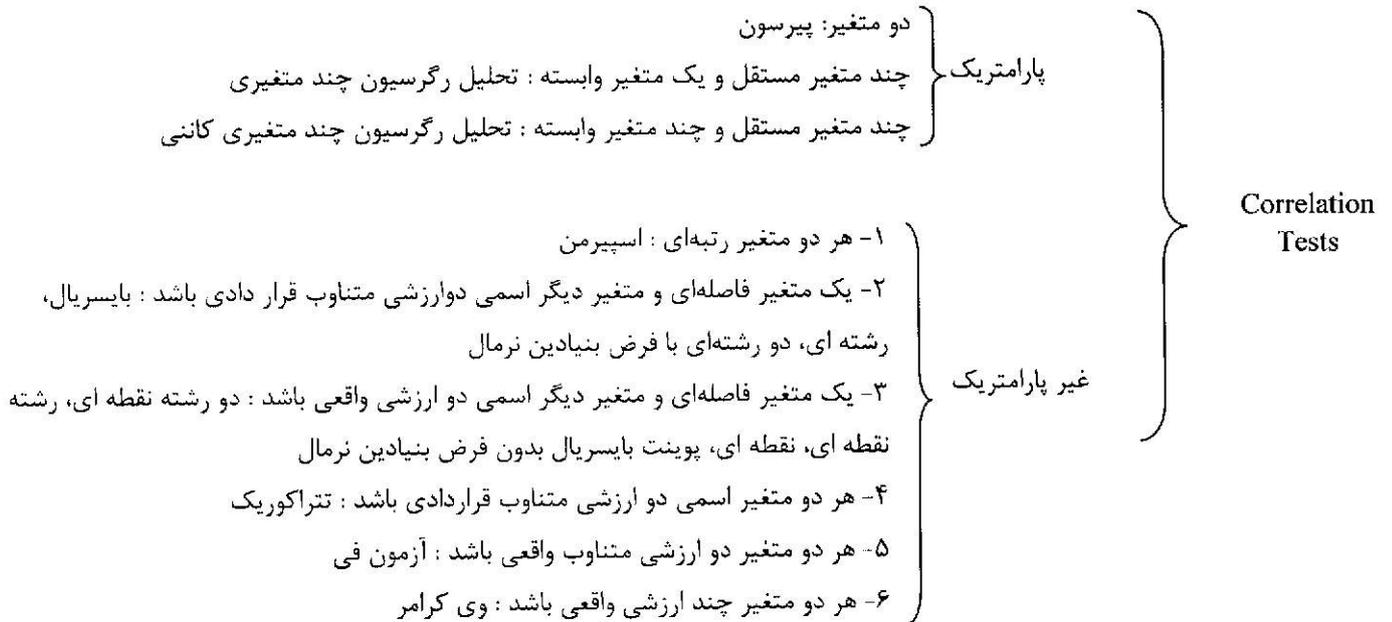
$$z = \frac{x - \bar{x}}{S}$$

$$2 = \frac{x - 15}{3} \Rightarrow 6 = x - 15 \Rightarrow x = 21$$

$$-2 = \frac{x - 15}{3} \Rightarrow -6 = x - 15 \Rightarrow x = 9$$

## مباحث آزمون‌های همبستگی

### همبستگی Correlation



سؤال: اگر بخواهیم همبستگی نمره هوشی و معدل را به دست آوریم. و قرار داد کنیم که از ۱۵ به بالا ← زرنگ و از ۱۵ به پائین ضعیف می‌باشد، باید از چه آزمونی استفاده کنیم؟ جواب ← بایسریال هوش ← فاصله ای

اسمی دو ارزشی } از ۱۵ به بالا ← زرنگ  
متناوب قرار دادی } از ۱۵ به پائین ← ضعیف  
متناوب ← یعنی یا این باشد یا آن، غیر از این قبول نیست.

نکته: در آزمون بایسریال ← فرض نرمال بودن ضروری است ← در مورد متغیر اسمی دو ارزشی متناوب. اگر بدون فرض بنیادی نرمال باشد. ← پوینت بایسریال می‌شود.

### آزمون پیرسون

با توجه به طبقه بندی آزمون‌های همبستگی در صورتی که داده‌های ما پارامتریک باشند و ما قصد بررسی میزان رابطه بین یک متغیر مستقل و یک متغیر وابسته را داشته باشیم از این آزمون استفاده می‌کنیم.

مهم: تعریف همبستگی از دیدگاه پیرسون ← میانگین حاصل ضرب نمرات استاندارد دو متغیر.

$$r = \frac{\sum z_x \cdot z_y}{n}$$

$$z_x = \frac{x - \bar{x}}{S_x}$$

$$z_y = \frac{y - \bar{y}}{S_y}$$

$$r = \frac{\sum \left( \frac{x - \bar{x}}{S_x} \right) \cdot \left( \frac{y - \bar{y}}{S_y} \right)}{n} \Rightarrow r = \frac{\sum (x - \bar{x}) \cdot (y - \bar{y})}{S_x \cdot S_y \cdot n} \Rightarrow r = \frac{\sum (x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{S_x^2 \cdot S_y^2}}$$

$$r = \frac{\text{cov}}{\sqrt{\text{دوتا واریانس}}}$$



× کوواریانس ← یعنی هم تغییری

کوواریانس میزان پراکندگی نمرات آزمودنی‌ها را نسبت به میانگین دو متغیر X و Y نشان می‌دهد.

بنابراین فرمول نهایی همبستگی به این قرار است:

$$r = \frac{\sum xy - n \bar{x} \bar{y}}{\sqrt{\left( \sum x^2 - n \bar{x}^2 \right) \left( \sum y^2 - n \bar{y}^2 \right)}}$$

تمرین: آیا بین اضطراب و تورم رابطه وجود دارد؟

x = اضطراب	y = تورم	xy	x <sup>2</sup>	y <sup>2</sup>
۱	۶	۶	۱	۳۶
۲	۷	۱۴	۴	۴۹
۳	۸	۲۴	۹	۶۴
۴	۹	۳۶	۱۶	۸۱
۵	۱۰	۵۰	۲۵	۱۰۰
Σ		۱۳۰	۵۵	۳۳۰
$\bar{x} = 3$	$\bar{y} = 3$			$\bar{xy} = 3 \times 8 = 24$

$$r = \frac{\frac{130}{5} - 24}{\sqrt{2 \times 2}} = \frac{26 - 24}{2} = \frac{2}{2} = 1$$

r = 1 ← یعنی همبستگی کامل و مستقیم است.

در بایسریال بعضی اوقات همبستگی کمی از ۱ بیشتر می‌شود.

چرا که بایسریال برآورد است.

در کل ضریب همبستگی بین  $-1 \leq r \leq +1$  قرار دارد.

اگر  $0 \leq r \leq +1$  ← همبستگی ناقص مستقیم است.

$r = 0$  ← همبستگی وجود ندارد.

$-1 < r < 0$  ← همبستگی ناقص معکوس است.

$r = +1$  ← همبستگی کامل و مستقیم است.

$r = -1$  ← همبستگی کامل معکوس است.

مهم: همیشه کوواریانس، جهت همبستگی را تعیین می‌کند. در ضریب همبستگی مخرج همیشه مثبت است.

نکته: اگر واریانس‌ها مساوی باشند ← ضریب همبستگی با ضریب زاویه مساوی است.

اگر داده‌ها را استاندارد بکنیم، در این صورت: کوواریانس = ضریب همبستگی می‌شود.

## ضریب تعیین

می‌گویند که چند درصد از متغیر اول را متغیر دوم تبیین می‌کند یا برعکس.

$$r^2 \times 100 = \text{ضریب تعیین یا واریانس مشترک}$$

$$r^2 \times 100 = \text{ضریب تعیین} = \text{درصد واریانس مشترک}$$

تمرین: بین پیشرفت تحصیلی و هوش در یک تحقیق همبستگی ۰,۲۲ به دست آمده است؛ چند درصد از پیشرفت تحصیلی وابسته به هوش است؟

$$r = 0.22 \Rightarrow \text{ضریب تعیین} = (0.22)^2 \times 100 = 4.48$$

رابطه بین انگیزه پیشرفت و پیشرفت تحصیلی چقدر است؛ اگر بدانیم:

$$r = 0.5 \Rightarrow \text{ضریب تعیین} = (0.5)^2 \times 100 = 25$$

اگر بدانیم  $R = 0.45$  است رابطه بین امکانات و پیشرفت تحصیلی، چقدر است؟

$$r = 0.45 \Rightarrow \text{ضریب تعیین} = (0.45)^2 \times 100 = 20.25$$

## گام هفتم

تحلیل سؤال به این دلیل استفاده می‌گردد تا سؤال‌های نامربوط را حذف کنیم.

## روش‌های تحلیل سؤال

(۱) درجه دشواری

(۲) ضریب تمیز

(۳) روش لوپ

(۴) روش تحلیل عوامل

(۵) روش مبتنی خصیصه مکنون (تئوری سوال پاسخ)<sup>14</sup>

روش‌های درجه دشواری و ضریب تمیز جزء روش‌های کلاسیک می‌باشند.

روش‌های درجه دشواری، ضریب تمیز، روش لوپ و روش تحلیل عوامل، جزء روش‌های وابسته به پاسخگو می‌باشند و بنابراین عناصر فرهنگی پاسخگویان در آن‌ها مؤثر است.

روش مبتنی بر خصیصه مکنون، جزء روش‌های وابسته به پاسخگو و بنابراین، وابسته به فرهنگ می‌باشد.

### محاسبه درجه دشواری

در اصل درجه آسانی یا سهولت است؛ چون از محاسبه ی تعداد پاسخ دهندگان صحیح نسبت به کل پاسخ دهندگان به دست می‌آید. در آزمون‌های نگرش و عاطفی ضریب مقبولیت نام دارد.

### محاسبه درجه دشواری در آزمون‌های عینی با مقیاس صفر و یک

اولین قدم تعیین ۲۷ درصد بالا و پایین است. یعنی ۲۷ درصد از افرادی را که در آزمون نمره بالایی گرفته اند انتخاب می‌کنیم. همینطور ۲۷ درصد از افرادی را که در آزمون نمره پایینی گرفته اند انتخاب می‌کنیم و سپس برای هر سؤال بصورت جداگانه درجه دشواری را با استفاده از فرمول زیر محاسبه می‌کنیم:

$$P = \frac{\text{تعداد صحیح بالا} + \text{تعداد صحیح پایین}}{\text{کل افراد بالا} + \text{کل افراد پایین}} \times 100$$

### محاسبه درجه دشواری در آزمون‌های تشریحی

چون در آزمون‌های تشریحی نمره به صورت پیوستار می‌باشد (مثلاً صفر تا ۳،۵ نمره) از این فرمول استفاده می‌کنیم:

$$P = \frac{\text{میانگین نمرات همه افراد در یک سوال}}{\text{بارم آن سوال}} \times 100$$

اگر  $P > 70$  باشد، سؤال آسان تلقی می‌شود و اگر  $P < 30$  باشد، سؤال سخت است.

مناسب‌ترین ضریب دشواری ۵۰ است و دامنه تغییر مناسب برای درجه دشواری  $30 \leq P \leq 70$  است.

نکته: در صورتی که آزمودنی‌های یک آزمون از جهت خصیصه مورد اندازه‌گیری متجانس باشند، انتظار می‌رود دامنه تغییر درجه دشواری سؤال‌ها محدود باشد.

هرچه درجه دشواری متوسط باشد واریانس آن سؤال بزرگتر است.

$$P = 0.5$$

$$S^2 = 0.5 \times 0.5 = 0.25$$

$$P = 0.3$$

$$S^2 = 0.3 \times 0.7 = 0.21$$

$$P = 0.8$$

$$S^2 = 0.8 \times 0.2 = 0.16$$

$$S^2 = P \cdot (1 - P) \quad \text{واریانس سؤال}$$

نکته: با افزایش سطح دشواری سؤال، واریانس سؤال ابتدا افزایش و سپس کاهش می‌یابد.

نکته: سطح دشواری ۰،۵ برای سؤال‌های کوتاه پاسخ مناسب است. در سؤالات عینی در تعیین سطح دشواری باید عامل حدس را در نظر گرفت.

## محاسبه ضریب تمیز

ضریب تمیز، قدرت سؤال را در تمایز گذاری یا تشخیص بین گروه قوی و ضعیف نشان می‌دهد. بنابراین، ضریب تمیز نوعی ضریب توافق سؤال با نمره کل آزمون است.

## محاسبه ضریب تمیز در سؤالات عینی

$$D = \frac{\text{تعداد صحیح گروه پایین} - \text{تعداد صحیح گروه بالا}}{\text{کل تعداد افراد گروه بالا یا پایین}}$$

روش دیگر: محاسبه ضریب همبستگی پوینت بایسریال ( دو رشته‌ای نقطه‌ای) بین نمره سؤال و نمره کل آزمون).

## محاسبه ضریب تمیز در سؤالات تشریحی

چون در آزمون‌های تشریحی نمره به صورت پیوستار می‌باشد، بنابراین از ضریب همبستگی پیرسون استفاده می‌کنیم. اگر همبستگی سؤال با نمره کل آزمون مثبت و معنی دار باشد، بنابراین سؤال قدرت تمیز خوبی دارد. اگر همبستگی سؤال با نمره کل آزمون منفی باشد و یا معنی دار نباشد، بنابراین سؤال قدرت تمیز ندارد. برای تعیین معنی داری قدرت تمیز به صورت زیر عمل می‌کنیم.

$$H_0 : r = 0$$

$$H_A : r \neq 0$$

$$d = n - 2$$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

$$r = \text{ضریب تمیز}$$

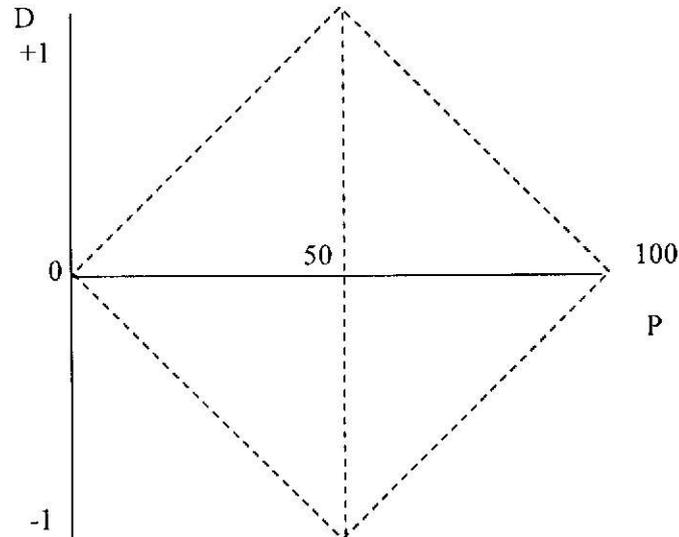
$$n = \text{تعداد آزمون شوندگان}$$

اگر  $t_c \geq |t_p|$  باشد فرض صفر رد می‌شود یعنی قدرت تمیز معنی دار است.

اگر  $t_c < |t_p|$  باشد فرض صفر را نمی‌توان رد کرد، یعنی قدرت تمیز معنی دار نیست. |۲|

## رابطه بین ضریب تمیز و درجه دشواری

در نقطه ۵۰ درصد بهترین ضریب تمیز را داریم. رابطه بین ضریب تمیز و درجه دشواری را به صورت شکل زیر می‌توان نمایش داد.



همانگونه که شکل فوق نمایش می‌دهد با افزایش درجه دشواری ضریب تمیز افزایش می‌یابد تا در نقطه ۵۰ درصد درجه دشواری، بالاترین ضریب تمیز را داریم، و پس از آن با افزایش درجه دشواری ضریب تمیز کاهش می‌یابد تا به نقطه ۱۰۰ درصد ضریب تمیز صفر می‌شود.

نکته: توزیع نمرات آزمودنی که دارای سؤالات با درجه دشواری ۰،۵ و ضریب تمیز ۱ می‌باشند، هنجار است.

نکته: در صورتی که هیچ یک از آزمودنی‌ها به یک سؤال آزمون پاسخ ندهند، مقدار درجه دشواری و ضریب تمیز آن سؤال، صفر خواهد بود.

نکته: در صورتی که تمام آزمودنی‌ها به یک سؤال آزمون پاسخ صحیح بدهند، مقدار درجه دشواری ۱۰۰ و ضریب تمیز آن سؤال، صفر خواهد بود.

نکته: هر اندازه قدرت تمیز یک آزمون بیشتر باشد، نمرات حاصل از اجرای آن به همان اندازه یکدست و یکنواخت‌تر خواهد بود.

نکته: در صورتی که درجه دشواری ۵۰ درصد باشد ضریب تمیز ۱ خواهد بود، به عبارت دیگر درجه دشواری ۵۰ درصد اطلاعات افتراقی بیشتری به دست می‌دهد.

مثال: در جدول مقابل، اگر گزینه الف صحیح باشد، درجه دشواری و قدرت تمیز سؤال به ترتیب چقدر خواهد شد؟

	الف	ب	ج	د
گروه قوی	۱۵	۲	۰	۳
گروه ضعیف	۲	۱۰	۰	۸

$$P = \frac{15+2}{20+20} \times 100 = 42.5$$

$$D = \frac{15-2}{20} = 0.65$$

### تحلیل گزینه‌های انحرافی

هدف گزینه‌های انحرافی، منحرف کردن کسانی است که جواب سؤال را کاملاً نمی‌دانند، این کار بسیار مشکل است و باید از روی اشتباهات یادگیرندگان این گزینه‌ها را طرح کرد.

### قاعده کلی تحلیل گزینه‌های انحرافی

هر گزینه انحرافی باید دست کم یک نفر از افراد گروه ضعیف را به خود جلب نماید و اگر هر دو گروه را جلب کرد تعداد افراد گروه ضعیف باید بیشتر باشد چون افراد گروه ضعیف در مجموع اطلاعات کمتری از مطالب آزمون را دارند. بنابراین گزینه ب و د در مثال فوق گزینه‌های انحرافی مناسبی است و ب بهتر از د است در حالی که گزینه انحرافی ج بسیار نامناسب است زیرا هیچکس را به خود جلب ننموده است.

### تحلیل سؤال‌های آزمون‌های عملکردی و عاطفی

$$P = \frac{\sum H + \sum L - (2N \text{score min})}{2N(\text{score max} - \text{score min})}$$

درجه دشواری

$$D = \frac{\sum H - \sum L}{N(\text{score max} - \text{score min})}$$

ضریب تمیز

$\sum H$  : مجموع نمره‌های گروه بالا

$\sum L$  : مجموع نمره‌های گروه پایین

N : تعداد افراد گروه بالا یا پایین

Score max : بالاترین نمره سؤال

Score min : پایین‌ترین نمره همان سؤال [۳]

### محاسبه ضریب دشواری و تمیز در پیش و پس آزمون

$$P = \frac{\text{تعداد افرادی که در پس آزمون به سوال پاسخ صحیح داده‌اند} + \text{تعداد افرادی که در پیش آزمون به سوال پاسخ صحیح داده‌اند}}{\text{تعداد کل پاسخ‌دهندگان} \times 2} \times 100$$

$$D = \frac{\text{تعداد افرادی که در پیش آزمون به سوال پاسخ صحیح داده‌اند} - \text{تعداد افرادی که در پس آزمون به سوال پاسخ صحیح داده‌اند}}{\text{تعداد کل پاسخ‌دهندگان}} \times 100$$

## درجه دشواری و ضریب تمیز در آزمون‌های وابسته به ملاک (در حد تسلط بودن)

$$P = \frac{\text{تعداد پاسخ‌دهندگان صحیح در گروه غیر حد تسلط} + \text{تعداد پاسخ‌دهندگان صحیح در گروه در حد تسلط}}{\text{تعداد کل پاسخ‌دهندگان}} \times 100$$

$$D = \frac{\text{تعداد پاسخ‌دهندگان صحیح در گروه [۲] حد تسلط}}{\text{گروه در حد تسلط } n} - \frac{\text{تعداد پاسخ‌دهندگان صحیح در گروه غیر حد تسلط}}{\text{گروه غیر حد تسلط } n} \times 100$$

با توجه به جدول زیر قدرت تمیز این سؤال چقدر است؟

	عدم تسلط	تسلط
پاسخ درست	۲	۹
پاسخ نادرست	۵	۲

$$D = \frac{9}{11} - \frac{2}{7} = 0.52$$

نکته: مهم‌ترین ملاک برای تصمیم‌گیری در مورد حذف یا نگهداری سؤال در آزمون، ضریب تشخیص (تمیز) می‌باشد.

### روش لوپ

محاسبه پایایی کلی را انجام می‌دهیم، عددی که به دست می‌آید نشان دهنده هماهنگی درونی سؤالات است. اگر یک سؤال را از مجموعه سؤال‌ها کم کنیم و پایایی را دوباره حساب کنیم، در صورتی که پایایی پایین بیاید، نشان دهنده با ارزش بودن سؤال است، ولی اگر با حذف سؤال، پایایی بقیه سؤالات بالا برود، نشان دهنده نامناسب بودن سؤال می‌باشد. [۷]

### روش تحلیل عامل

تحلیل عامل؛ یک مجموعه محاسبات ریاضی است با دو هدف اکتشافی و تأییدی. اگر در آزمون پس از انجام تحلیل عامل K عامل به دست بیاید، بنابراین، سؤال مناسب، سؤالی است که حداقل در یک عامل نقش داشته باشد. در صورتی که سؤال در هیچ عاملی نقش نداشته باشد، در این صورت سؤال نامناسبی است. [۸]

### روش IRT

تحلیل سؤال به کمک نظریه IRT (تئوری سؤال پاسخ) را در مورد آزمون‌های توانایی عینی (صفر و یک) می‌توانیم انجام دهیم. نظریه فوق بر این فرض استوار است که یک صفت یا ویژگی زیر بنایی وجود دارد که به شخص امکان می‌دهد تا در یک تکلیف شناختی معین موفقیت کسب نماید. [۸]

**برای مثال:** قدبک صفت مکنون تک بعدی است، چون نمی‌توان مفهوم قد را دید به آن مکنون می‌گویند. اما خصیصه‌های آن را می‌توان اندازه گرفت.

در نظریه سؤال - پاسخ به کاربرد نظریه یا الگوی صفت مکنون گفته می‌شود، در نظریه سؤال - پاسخ به جای تأکید بر نمرات کل آزمون، بر پاسخ‌های آزمودنی‌ها به تک تک سؤالات آزمون تأکید می‌شود، و با استفاده از مدل‌های ریاضی پیچیده می‌توان احتمال پاسخ درست به یک سؤال را به عنوان تابعی از توانایی آزمون‌شونده و همچنین برخی ویژگی‌های سؤال را پیش‌بینی کرد. [۳]

ویژگی‌های سؤال به دست آمده توسط این نظریه را به صورت نمودار منحنی ویژگی سؤال (ICC)<sup>۱۵</sup> نشان می‌دهند.

برای تهیه منحنی ویژگی یک سؤال، نسبت یا درصد آزمون‌شوندگانی که آن سؤال را درست جواب داده‌اند در رابطه با نوعی ملاک یا معیار (مثلاً نمره کل آزمون) رسم می‌شود. یعنی، بر روی محور افقی نمره کل آزمون و بر روی محور عمودی نسبت آزمون‌شوندگانی که به سؤال پاسخ درست داده‌اند مشخص می‌شود و از روی این منحنی می‌توان ضریب دشواری و تمیز را به دست آورد.

**ضریب دشواری:** نمره معیار یا ملاکی است که در آن ۵۰ درصد آزمون‌شوندگان سؤال را درست پاسخ داده‌اند.

**ضریب تمیز:** شیب منحنی ویژگی سؤال، هر چه شیب بیشتر باشد، ضریب تمیز قوی‌تر است. [۳]

بنابراین سؤال مناسب، سؤالی است که منحنی ویژگی آن در نقطه برش حداکثر شیب را دارد.

**نکته:** خط افقی در منحنی ویژگی سؤال، همه جواب غلط داده‌اند یا همه جواب صحیح داده‌اند.

**نکته:** آستانه، نقطه عطف منحنی است زیرا در این نقطه جهت منحنی تغییر می‌کند و در نقطه برش ۵۰ درصد است.

### برازنده کران منحنی ویژگی سؤال

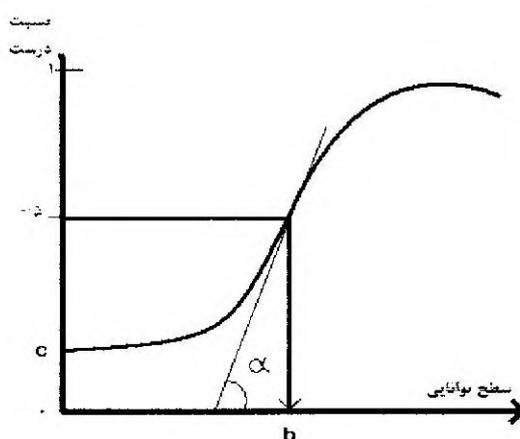
$\alpha = \lg \alpha$  نشان دهنده شیب منحنی برای قدرت تمیز

b معرف آستانه یا همان ضریب دشواری سؤال

c پارامتر حدس و تصادف است.

هر چه احتمال کسب نمره از راه حدس برای سؤالی بیشتر باشد مقدار c برای آن سؤال بیشتر است.

شکل ۳. برازنده کردن منحنی ویژگی سؤال



## سؤالات ۵۰ درصد اول

- ۱ - توانایی در ارائه یک طرح تحقیقی جزء کدام یک از طبقات حوزه شناخت است؟  
 الف) کاربستن      ب) تحلیل      ج) ترکیب      د) ارزشیابی
- ۲ - کدام یک از سیستم‌های نمره گذاری، هم آسان‌تر انجام می‌گیرد و هم اطلاعات دقیق تری در اختیار می‌گذارد؟  
 الف) صفر تا ۲۰      ب) صفر تا ۱۰۰      ج) خوب، متوسط، ضعیف      د) الف، ب، ج، د
- ۳ - اگر شماره کارت ملی ایرانیان از سن آنان اطلاع دهد، این شماره در کدام مقیاس اندازه‌گیری بیان خواهد شد؟  
 الف) اسمی      ب) ترتیبی      ج) فاصله ای      د) نسبی
- ۴ - در پاسخ گویی به آزمون ها، کدام عامل در کاهش حدس و گمان تأثیر گذار است؟  
 الف) دشوار بودن گزینه ها      ب) کاهش تعداد سؤال ها  
 ج) زیاد بودن تعداد گزینه ها      د) قدرت تشخیص بالای سؤال ها
- ۵ - سطح دشواری ۵۰٪ برای کدام نوع از سوال ها مناسب تر است؟  
 الف) تشریحی      ب) صحیح - غلط      ج) چهار گزینه‌ای      د) کوتاه پاسخ
- ۶ - در ارزشیابی ملاک - مرجع، کارکرد یادگیرنده با ..... سنجیده می‌شود.  
 الف) جدول نرم      ب) یک معیار      ج) با گروه خودش      د) مقیاس نمره گذاری
- ۷ - در یک آزمون ۴۰ سؤالی پنج گزینه‌ای، که به پاسخ‌های غلط نمره منفی تعلق می‌گیرد، دانش آموزی به ۳۲ سؤال پاسخ صحیح، و به ۱۲ سؤال، پاسخ غلط داده است. اگر هر سؤال یک نمره داشته باشد، نمره اصلاح شده وی کدام است؟  
 الف) ۲۰      ب) ۲۷      ج) ۲۸      د) ۲۹
- ۸ - سؤال خوب، سؤالی است که منحنی ویژگی آن در نقطه برش ..... شیب را دارد؟  
 الف) حداقل      ب) حداکثر      ج) ۵۰٪      د) ۶۸٪
- ۹ - در گزارش نتایج یک آزمون، مقدار انحراف استاندارد «۱.۵ -» ذکر شده است. بنابراین، می‌توان گفت:  
 الف) دامنه نمرات خیلی زیاد بوده است.      ب) واریانس نمرات منفی بوده است.  
 ج) محاسبات انجام شده غلط بوده است.      د) جمع نمرات مثبت و منفی صفر نبوده است.
- ۱۰ - در جدول مقابل، اگر گزینه «الف» صحیح باشد، درجه دشواری و قدرت تمیز سؤال به ترتیب چه قدر خواهد بود؟
- |     |    |   |   |
|-----|----|---|---|
| الف | ب  | ج | د |
| ۱۵  | ۲  | ۰ | ۳ |
| ۲   | ۱۰ | ۰ | ۸ |
- الف) ۴۲،۵٪ و ۰،۶۵      ب) ۰،۶۵ و ۴۲،۵٪      ج) ۰،۸۵ و ۰،۸۵      د) ۰،۸۵ و ۰،۸۵



- ۲۰ - در ارتباط با تهیه و اجرای آزمون‌های تشریحی، کدام روش توصیه نشده است؟  
 الف) از کلماتی مثل، چه کسی، چه وقت، و هرگز استفاده نکردن.  
 ب) زمان لازم برای پاسخگویی به هر سؤال را جداگانه در نظر گرفتن.  
 ج) حق انتخاب چند سؤال از بین تعدادی سؤال را به آزمودنی‌ها دادن.  
 د) به جای ورقه تصحیح کردن، سؤال به سؤال تصحیح کردن.
- ۲۱ - اگر طول پرسش‌ها و پاسخ‌های سؤالات جورکردنی کوتاه باشد، انجام دادن کدام یک از موارد زیر راحت‌تر خواهد شد؟  
 الف) تضمین روایی محتوا  
 ب) ازدیاد پایایی آزمون  
 ج) رعایت تجانس دو ستون  
 د) اندازه‌گیری اهداف سطح بالا
- ۲۲ - در آزمونی که ۶۰ نفر در آن شرکت کرده اند، ۱۵ نفر به سؤالی پاسخ غلط می‌دهند. ضریب دشواری این سؤال کدام است؟  
 الف) ۰،۱۵      ب) ۰،۲۵      ج) ۰،۴۵      د) ۰،۷۵
- ۲۳ - در طرح سؤالات چندگزینه‌ای، تنها محدودیتی که توصیه شده عبارت است از:  
 الف) استفاده نکردن از ذوق و سلیقه شخصی  
 ب) نوشتن تنه سؤال به صورت جمله سؤالی  
 ج) استفاده نکردن از جملات منفی و دشوار  
 د) حفظ ویژگی اصل سؤال (تنه و تعدادی گزینه)
- ۲۴ - در طرح سؤالات جورکردنی، تنظیم فهرست پاسخ‌ها به طور منطقی چه امتیازی دارد؟  
 الف) در وقت دانش آموزان صرفه جویی می‌شود.  
 ب) انسجام بین پاسخ‌های سؤالات تضمین می‌شود.  
 ج) پاسخ‌ها و پرسش‌ها با یکدیگر جور می‌شوند.  
 د) یک پاسخ معین با چند پرسش جور می‌شود.
- ۲۵ - اگر ضریب تشخیص (قوه تمیز) یک سؤال ۱۰۰ درصد باشد، درجه دشواری سؤال چند درصد خواهد بود؟  
 الف) صفر      ب) ۲۵      ج) ۵۰      د) ۱۰۰
- ۲۶ - در کدام یک از روش‌های ارزشیابی زیر، معلم، ضمن ارزشیابی، آموزش نیز می‌دهد؟  
 الف) پورت فولیو      ب) واقعه نگاری      ج) نمونه کار      د) چند گزینه ای
- ۲۷ - ضریبی که در مقابل هر یک از قسمت‌های جدول طرح آزمون نوشته می‌شود، چه چیزی را نشان می‌دهد؟  
 الف) ارزش محتوای درس      ب) تعداد سؤالات هر فصل  
 ج) اهمیت نسبی هر مورد      د) اهمیت هدف‌های آموزشی
- ۲۸ - در مراحل تحلیل سؤال‌های آزمون، مهم‌ترین ملاک برای تصمیم‌گیری در مورد حذف یا نگهداری سؤال در آزمون جدید چیست؟  
 الف) ضریب تشخیص      ب) واریانس سؤال      ج) سطح دشواری      د) کارایی گزینه‌ها
- ۲۹ - کدام یک از هدف‌های زیر برای جدول طرح آزمون مناسب‌تر خواهد بود؟  
 الف) مفاهیم کلیدی را درست تعریف کند.  
 ب) طبقه بندی آزمون‌ها را بشناسد.  
 ج) طرز تهیه جدول دو بعدی را درست بفهمد.  
 د) تفاوت میان اعتبار و روایی را بدانند.

۳۰ - کدام یک از فرآیندهای ذهنی در قلمرو « درک و فهم » دانش آموز قرار می‌گیرد؟

- (الف) تسلط بر دانش اصول و قواعد کلی  
 (ب) توانایی در تجزیه و تحلیل یک مطلب  
 (ج) قدرت ابتکار و نوآوری مطلب جدید  
 (د) توانایی در تفسیر نکات مهم یک مطلب

۳۱ - به منظور کار عملی، از دانشجویان خواسته شده است که یکی از مسائل آموزش و پرورش را بررسی کنند و راه حل‌های پیشنهادی خود را ارائه دهند. این تکلیف کدام سطح از یادگیری را ارزیابی می‌کند؟

- (الف) کاربرد (ب) تجزیه (ج) ترکیب (د) ارزشیابی

۳۲ - نمره استاندارد (Z) دانش آموزی ۰،۵ است. اگر نمره خام این دانش آموز ۱۵ و میانگین کلاس ۱۴ باشد، انحراف معیار نمرات کلاس چند خواهد بود؟

- (الف) ۱ (ب) ۱/۵ (ج) ۲ (د) ۲/۵

۳۳ - کدام درجه دشواری سؤال، اطلاعات افتراقی بیشتری را به دست می‌دهد؟

- (الف)  $P = ۰,۲۵$  (ب)  $P = ۰,۵۰$  (ج)  $P = ۰,۷۵$  (د)  $P = ۱,۰۰$

۳۴ - نسبت پاسخ درست آزمودنی‌های گروه ملاک بالا (قوی) به یک سؤال آزمون ۰،۲۷ و نسبت پاسخ درست آزمودنی‌های گروه ملاک پایین (ضعیف) ۰/۳۷ است. ضریب تمیز این سؤال کدام است؟

- (الف) -۰،۰۵ (ب) -۰،۱۰ (ج) +۰،۳۲ (د) +۰،۶۴

۳۵ - در صورتی که آزمودنی‌های یک آزمون از جهت خصیصه مورد اندازه‌گیری متجانس باشند، انتظار می‌رود دامنه تغییر درجه دشواری سؤال‌ها چگونه باشد؟

- (الف) محدود (ب) گسترده (ج) در حد متوسط (د) در حد بالا

۳۶ - برای محاسبه رابطه بین نمره‌های درس دستورزبان و درک زبان انگلیسی که هر یک توزیعی دو ارزشی دارند، کدامیک از ضرایب همبستگی ارجح‌تر است؟

- (الف) ضریب دو رشته ای  
 (ب) ضریب فی  
 (ج) ضریب تترا کوریک  
 (د) ضریب دو رشته‌ای نقطه ای

۳۷ - جدول زیر نشان دهنده پاسخ‌های درست و نادرست دو گروه از دانش آموزان به یک سؤال است. قدرت تمیز این سؤال چقدر است؟

	عدم تسلط	تسلط
پاسخ درست	۲	۹
پاسخ نادرست	۵	۲

- (الف) ۰،۴۳ (ب) ۰،۵۲ (ج) ۰،۶۳ (د) ۰،۷۳

۳۸ - در یک آزمون ۱۰۰ سؤالی با توزیع نرمال و انحراف معیار ۱۵ و میانگین ۵۰، تقریباً چند درصد افراد نمره‌ای بین ۵ و ۲۰ دارند؟

الف) ۲ درصد      ب) ۵ درصد      ج) ۸ درصد      د) ۱۶ درصد

۳۹ - در هنگام آموزش یک هدف عاطفی، دانش آموزان از معلم می‌خواهند که توضیحات بیشتری ارائه نماید. این دانش آموزان در کدام سطح از یادگیری عاطفی قرار دارند؟

الف) دانش      ب) ارزش گذاری      ج) سازماندهی ارزشها      د) توجه (دریافت)

۴۰ - کدامیک از انواع ارزشیابی تقریباً در حاشیه قرار دارد و مکمل عمل آموزشی نیست؟

الف) تشخیصی      ب) تکوینی      ج) مستمر      د) مجموعی

## گام هشتم

گام هشتم فرآیند ارزشیابی بررسی شواهد پایایی<sup>۱۶</sup> است.

پایایی یک وسیله اندازه‌گیری عمدتاً دقت نتایج حاصل را می‌سنجد.

بنا به تعریف؛ پایایی به دقت، اعتمادپذیری، همسانی در اندازه‌گیری، ثبات یا تکرار پذیری نتایج آزمون اشاره می‌کند. [۳]

از دیدگاه کرونباخ تعمیم‌پذیری نتایج آزمون به پایایی اشاره می‌کند. [۲]

نکته: پایایی آزمون با درجه دشواری متوسط و ضریب تمیز بالا ارتباط دارد.

برای اینکه یک آزمون روا<sup>۱۷</sup> باشد باید نخست پایا باشد یعنی پایایی شرط روایی<sup>۱۸</sup> است اما روایی برای پایایی ضروری نیست.

نکته: اگر آزمونی در هربار اجرا بر روی تعدادی دانش‌آموز نتایج مختلفی به دست بدهد، آن آزمون پایا نیست.

نکته: اگر آزمونی چیزی را بسنجد، پایایی آن بالاست.

## روش‌های تعیین پایایی

روش‌های تعیین پایایی را به قرار زیر می‌توان طبقه‌بندی نمود:

(۱) پایایی مصححان

(۲) روش بازآزمایی

(۳) روش فرم‌های هم‌ارز

(۴) روش همسانی درونی

روش دو نیمه کردن

تعیین بر آورد

ضریب آلفای کرونباخ

کودر- ریچاردسون ۲۰

کودر- ریچاردسون ۲۱

<sup>16</sup> Reliability

<sup>17</sup> Valid

<sup>18</sup> Validity

## پایایی مصححان

میزان همبستگی بین چند مصحح که یک برگه امتحانی را تصحیح می‌کنند و این پایایی به تصحیح کنندگان برگه‌های امتحانی مربوط است و نباید آن را با پایایی آزمون که مربوط به سؤالات آزمون است، اشتباه کرد. روش محاسبه آن به قرار زیر است:

$$\text{ضریب پایایی [۳]} = \frac{\text{تعداد توافقی‌ها}}{\text{تعداد توافقی‌ها} + \text{تعداد عدم توافقی‌ها}} \times 100$$

مثال: نتایج زیر به دست آمده است پایایی تصحیح را به دست آورید.

مردود توسط همه معلمان	قبول توسط همه معلمان	قبول توسط عده‌ای و مردود توسط عده‌ای دیگر
۲۵	۳۵	۴۵

$$\text{ضریب پایایی} = \frac{60}{100} \times 100 = 60\%$$

## روش بازآزمایی

ساده‌ترین روش تعیین پایایی یک آزمون، بازآزمایی می‌باشد. یک آزمون را در دو نوبت به گروه واحدی از آزمون شونده‌گان می‌دهند و نمرات حاصل را باهم مقایسه می‌کنند. ضریب همبستگی بین نمرات دوبار اجراء، ضریب پایایی آزمون است.

این روش تنها زمانی قابل استفاده است که ویژگی مورد اندازه‌گیری در طول زمان تغییر نکند: مثل هوش [۳] به عبارت دیگر، در این روش تعیین پایایی، میزان تغییر در طول زمان به عنوان منبعی از خطا در نظر گرفته می‌شود.

## معایب روش بازآزمایی

- (۱) آشنایی با سؤالات در نوبت اول آزمون
- (۲) چون نمونه سؤال‌ها تکراری است، نمرات حاصل از بازآزمایی هیچ اطلاعاتی مبنی بر این که اگر نمونه سؤالات تغییر کند، چه تغییری حاصل خواهد شد، به دست نمی‌دهد.
- (۳) اگر فاصله زمانی دو اجراء زیاد شود، احتمال دارد مقدار خطای اندازه‌گیری  $S^2_e$  با نمره واقعی  $S^2_T$  در هم آمیزد.
- (۴) اجراء مجدد آزمون، مطلوب معلم و شاگرد نیست و فقدان علاقه در نمرات اجراء دوم تأثیر می‌گذارد. [۳]

## روش فرم‌های هم ارز

تهیه دو آزمون موازی برای یک مطلب و اجرای آن‌ها روی یک گروه واحد از آزمون‌شوندگان در فاصله زمانی کوتاه. دو آزمون در صورتی موازی هستند که نسبت واریانس نمره واقعی آن‌ها به واریانس نمره مشاهده شده باهم برابر باشد  $\frac{S^2_T}{S^2_x} = \frac{S^2_{T'}}{S^2_{x'}}$ ، همینطور نمره‌های واقعی آن‌ها مساوی باشد  $x_T = x_{T'}$  و نمرات خطای آن‌ها مساوی باشد  $x_E = x'_E$ . همینطور واریانس نمرات مشاهده شده آن‌ها باهم برابر باشد  $S^2_x = S^2_{x'}$ . در این صورت کوواریانس مشاهده شده، همان واریانس نمرات حقیقی است.  $S^2_T = \text{Cov}_{xx'}$ . اگر  $r^2_{xx'} = 1$  در این صورت  $x_T = x_0$ . یعنی اگر ضریب پایایی مساوی یک باشد در این صورت نمره مشاهده شده همان نمره واقعی است.

اگر  $r^2_{TT'} = 1$  در این صورت دو آزمون موازی هستند. یعنی اگر همبستگی بین نمرات واقعی دو آزمون برابر یک باشد، در این صورت آن دو آزمون موازی هستند.

اگر  $r_{xx'} = 0$  باشد در این صورت  $S^2_x = S^2_E$ . یعنی اگر همبستگی بین دو آزمون مساوی صفر باشد، در این صورت واریانس نمره مشاهده شده برابر با واریانس نمرات خطا می‌باشد.

اگر  $r^2_{xx'} = 1$  در این صورت  $S^2_x = S^2_T$ . یعنی اگر همبستگی بین دو آزمون موازی مساوی یک باشد، در این صورت واریانس نمره مشاهده شده برابر با واریانس نمرات خطا می‌باشد.

اگر  $0 < r^2_{xx'} < 1$  در این صورت  $S^2_x = S^2_T + S^2_E$ . یعنی اگر همبستگی بین دو آزمون موازی عددی بین یک الی صفر باشد در این صورت واریانس نمره مشاهده شده برابر با مجموع واریانس‌های نمرات واقعی و خطا می‌باشد. [۸]

## مشکل عمده روش هم ارز

ساخت فرم هم ارز برای یک آزمون کار مشکلی است.

در بعضی موارد آزمون‌شوندگان در پاسخ به سؤال‌های فرم اول آزمون با اصولی آشنا می‌شوند که عملکرد آنان را در فرم دوم تحت تأثیر قرار می‌دهد. [۳]

## روش‌های همسانی درونی

در این روش‌ها همسانی یا یکنواختی درون آزمون (ماده‌ها یا سؤالات تشکیل دهنده آزمون) را می‌سنجد. بنابراین اگر آزمونی دارای سؤال‌های متجانس و همگون باشد از طریق همسانی درونی یا ثبات درونی، بهتر می‌توان پایایی را برآورد کرد. [۳]

## روش دو نیمه کردن

روش دو نیمه کردن اولین روش همسانی درونی می‌باشد. در این روش آزمون را یک بار با گروه واحدی اجرا می‌کنیم و پس از اجرا آن را به دو نیمه تقسیم می‌نماییم. ضریب همبستگی حاصل از نمرات دو نیمه آزمون، ضریب پایایی هر یک از دو نیمه خواهد بود. برای تعیین پایایی کل آزمون از فرمول تصحیح اسپیرمن- براون استفاده می‌کنیم که به قرار زیر است:

$$r_{ii} = \frac{K \cdot r_{1/2/2}}{1 + (k-1)r_{1/2/2}}$$

$r_{ii}$  = ضریب پایایی کل آزمون

$r_{1/2/2}$  = ضریب همبستگی بین نیمه‌های آزمون

$$K = \frac{\text{تعداد سوالات جدید}}{\text{تعداد سوالات قبلی}} = K \quad (\text{در دو نیمه آزمون ۲ می‌شود}) [۳]$$

مثال: ضریب همبستگی دو نیمه آزمون  $r_{1/2/2} = 0.6$ ، ضریب پایایی کل آزمون چقدر است؟

$$K = 2$$

$$r_{ii} = \frac{2 \times 0.6}{1 + (2-1)0.6} = \frac{1.2}{1.6} = 0.75$$

مثال: ضریب پایایی برای یک آزمون ۱۰۰ سؤالی برابر ۰.۹ است. اگر طول آزمون را به  $\frac{3}{4}$  تقلیل دهیم ضریب پایایی چقدر خواهد بود؟

$$K = \frac{3}{4}$$

$$r_{ii} = \frac{\frac{3}{4} \times 0.9}{1 + (\frac{3}{4} - 1) \times 0.9} = \frac{0.675}{0.775} = 0.87$$

مثال: ضریب پایایی یک آزمون n سؤالی برابر ۰.۸ است. برای آن که ضریب پایایی به ۰.۹ برسد طول آزمون چقدر باید باشد.

$$K = \frac{x}{n}$$

$$0.9 = \frac{\frac{x}{n} \times 0.8}{1 + (\frac{x}{n} - 1) \times 0.8} = \frac{\frac{0.8x}{n}}{\frac{n + (x-n)0.8}{n}} = 0.9 \Rightarrow 0.8x = 0.9n + 0.72(x-n) \Rightarrow 0.8x = 0.9n + 0.72x - 0.72n$$

$$0.08x = 0.18n$$

$$x = 2.25n$$

## روش ضریب آلفای کرونباخ

در این روش اجزاء یا قسمت‌های آزمون برای سنجش ضریب پایایی آزمون به کار می‌روند:

$n$  = تعداد بخش‌های آزمون

$$\alpha = \left( \frac{n}{n-1} \right) \left( 1 - \frac{\sum S_i^2}{S^2} \right)$$

$S_i^2$  = واریانس بخش‌های آزمون

$S^2$  = واریانس کل آزمون

مثال: آزمونی با ۱۰ بخش با مجموع واریانس بخش‌ها  $\sum S_i^2 = 15$  و واریانس نمرات ۲۰، ضریب پایایی چقدر است؟

$$\alpha = \left(\frac{10}{9}\right) \left(1 - \frac{15}{20}\right) = 0.27$$

نکته: ضریب آلفای کرونباخ برای آزمون‌هایی با سؤالات تشریحی نیز به کار می‌رود و هر سؤال یک بخش تلقی می‌شود.

مثال: واریانس‌های هریک از ۴ سؤال یک آزمون انشایی به ترتیب ۹، ۴، ۸، ۲، ۱۰ و ۱۶ محاسبه شده است، اگر واریانس کل سؤال برابر ۱۰۰ باشد، ضریب پایایی آزمون چقدر خواهد بود؟

$$\alpha = \left(\frac{4}{3}\right) \left(1 - \frac{40}{100}\right) = 0.8$$

نکته: مقدار  $\alpha$  کرونباخ چون برآورد است، ممکن است بیشتر از یک بشود.

نکته: مقدار  $\alpha$  کرونباخ وابسته به مقدار  $n$  (تعداد خرده آزمون‌ها) می‌باشد. هر چقدر  $n$  بیشتر باشد برآورد دقیق‌تر است. همانطور که اشاره گردید،  $\alpha$  کرونباخ یک برآورد است. این روش میانگین همه همبستگی‌های دو نیمه‌های یک آزمون است که با فرمول اسپیرمن - براون اصلاح گردد.

$$[\lambda] \alpha = r_{ii} = \frac{K\bar{r}}{1 + (K-1)\bar{r}}$$

مثال: در یک آزمون دارای ۴ سؤال بین تمام سؤالات به صورت دو به دو همبستگی گرفتیم و میانگین همبستگی‌ها را به دست آوردیم،  $\bar{r} = 0.7$ ، میزان پایایی آزمون چقدر است؟

$$\alpha = r_{ii} = \frac{2(0.7)}{1 + (2-1)0.7} = \frac{1.4}{1.7} = 0.82$$

## روش کودر - ریچاردسون ۲۰ و ۲۱

در صورتی که سؤالات دو ارزشی (صفر و یک) باشند و درجه دشواری سؤالات متفاوت باشد، در این صورت، از فرمول کودر - ریچاردسون ۲۰ استفاده می‌کنیم:

$$KR_{20} = \left(\frac{n}{n-1}\right) \left(1 - \frac{\sum p_i q_i}{S_i^2}\right)$$

$n$  = تعداد سؤالات

$S_i^2$  = واریانس نمرات آزمون

$p$  = نسبت آزمون شوندگانی که به هر سؤال پاسخ صحیح داده اند.

$q$  = نسبت آزمون شوندگانی که به هر سؤال پاسخ غلط داده اند.

اگر همه سؤالات دارای درجه دشواری متوسط و شبیه به هم باشند از  $KR_{21}$  استفاده می‌کنیم:

$$[\lambda] KR_{21} = \left(\frac{n}{n-1}\right) \left(1 - \frac{npq}{S_i^2}\right)$$

نکته: در روش‌های  $\alpha$  کرونباخ، کودر-ریچاردسون ۲۰ و ۲۱ برای تعیین ضریب پایایی به تک تک سوالات توجه می‌شود.  
 مثال: میانگین و انحراف معیار نمره‌های دانش آموزان یک کلاس ۳۱ نفری در آزمون ۵۰ سوالی به ترتیب برابر با ۳۰ و ۶ شده است، ضریب پایایی این آزمون چقدر است؟

$$\bar{p} = \frac{30}{50} = 0.6$$

$$\bar{q} = 1 - 0.6 = 0.4$$

$$S_i^2 = 36$$

$$n = 50$$

$$KR_{21} = \left( \frac{50}{49} \right) \left( 1 - \frac{50 \times 0.6 \times 0.4}{36} \right) = 0.68$$

نکته: در صورتی که فقط سه جزء تعدادسؤال‌ها، میانگین و واریانس را داشته باشیم،  $KR_{21}$  از فرمول زیر محاسبه می‌شود:

$$[۳] \quad KR_{21} = \left( \frac{n}{n-1} \right) \left( 1 - \frac{\bar{x}(n-\bar{x})}{nS_i^2} \right)$$

مثال: فوق را با این فرمول نیز محاسبه می‌کنیم:

$$\bar{x} = 30$$

$$S_i^2 = 36$$

$$n = 50$$

$$KR_{21} = \left( \frac{50}{49} \right) \left( 1 - \frac{20(50-30)}{50 \times 36} \right) = 0.68$$

نکته: مقادیر پایایی حاصل از دو روش  $KR_{20}$  و  $KR_{21}$  در حالتی برابر خواهند بود که دشواری همه سؤال‌های آزمون یکسان باشند. در غیر این صورت مقدار  $KR_{21}$ ، کوچکتر از مقدار  $KR_{20}$  خواهد شد.

## روش تعیین پایایی آزمون‌های سرعت

آزمون‌ها را می‌توان به دو دسته آزمون‌های قدرت و آزمون‌های سرعت طبقه بندی کرد. برای آزمون‌های قدرت تمامی روش‌های پایایی قابل استفاده است. ولی در آزمون‌های سرعت، فقط از روش‌های بازآزمایی استفاده می‌کنیم، چون در این آزمون‌ها تقریباً تمامی سؤالاتی را که وقت رسیدگی به آن‌ها را دارند، درست جواب می‌دهند و تنها تفاوت آن‌ها در تعداد سؤالاتی است که جواب می‌دهند. برای کاهش یادآوری دوبار اجرا در بازآزمایی، بهترین روش، اجرای دو فرم هم ارز یا موازی است که بلافاصله بعد از هم اجرا می‌شود.

## پایایی نمرات اختلاف

تفاوت بین نمرات دو آزمون (پیش آزمون و پس آزمون یا دو فرم موازی یک آزمون)، دارای پایایی اختلاف است که به قرار زیر به دست می‌آید:

$$r_d = \frac{r_{1,1} + r_{2,2} - r_{1,2}}{2 - r_{1,2}} \quad r_d = \text{پایایی اختلاف}$$

$$r_{1,1} = \text{پایایی آزمون اول}$$

$$r_{2,2} = \text{پایایی آزمون دوم}$$

$$r_{1,2} = \text{همبستگی آزمون اول و دوم}$$

مثال: دو آزمون دارای همبستگی ۰٫۸ است، ضریب پایایی آزمون اول ۰٫۹ و آزمون دوم ۰٫۸۵ است، ضریب پایایی نمره اختلاف چقدر است؟

$$r_d = \frac{0.9 + 0.85 - 0.8}{2 - 0.8} = 0.4$$

$$0.8 = r_{1,2}$$

$$0.9 = r_{1,1}$$

$$0.85 = r_{2,2}$$

نکته: هرچه همبستگی دو آزمون بیشتر باشد، پایایی نمره اختلاف دو آزمون کمتر می‌شود.

## خطای معیار اندازه‌گیری

از فواید ضریب پایایی، به دست آوردن خطای اندازه‌گیری است.

فقدان دقتی که از ناپایی آزمون ناشی می‌شود، در خطای اندازه‌گیری انعکاس می‌یابد. خطای معیار اندازه‌گیری به قرار زیر به دست می‌آید:

$SEM$ : خطای معیار اندازه‌گیری

$S_r$ : انحراف معیار نمرات

$r_{ii}$ : ضریب پایایی آزمون

مثال: انحراف معیار آزمونی برابر با ۸ و ضریب پایایی آن برابر با ۰٫۷۵ می‌باشد، خطای معیار اندازه‌گیری چقدر است؟

$$SEM = 8\sqrt{1 - 0.75} = 4$$

مثال: اگر ضریب پایایی یک آزمون با توزیع نرمال برابر ۰٫۸۴ باشد، خطای معیار اندازه‌گیری بر حسب نمرات  $Z$  چقدر خواهد بود؟

$$SEM = 1\sqrt{1 - 0.84} = 0.4$$

نکته: انحراف معیار بر حسب نمرات Z برابر یک است.

محاسبه خطای معیار در واقع برآورد خطا است، که به آن برآورد نقطه‌ای می‌گویند. اما جهت برآورد خطا به صورت فاصله‌ای به قرار زیر عمل می‌کنیم:

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

$$d = SEM \cdot Z$$

$$x_t = x_0 \pm d$$

مثال: اگر خطای معیار یک آزمون ۳ باشد، فراگیری که در آن نمره ۱۰ گرفته است، نمره واقعی وی به احتمال ۶۸ درصد بین کدام دو نمره قرار خواهد گرفت؟

$$SEM = 3$$

$$Z = 1$$

در سطح ۶۸ درصد اطمینان  $d = 3 \times 1 = 3$

$$x_t = 10 \pm 3$$

مثال: اگر خطای معیار اندازه‌گیری آزمونی برابر با یک باشد با ۹۵ درصد اطمینان نمره واقعی فردی که در این آزمون ۱۵ را به دست آورده است کدام است؟

$$SEM = 1$$

$$Z = 1.96$$

در سطح ۹۵ درصد اطمینان  $d = 1 \times 1.96 = 1.96$

$$x_t = 15 \pm 1.96$$

بنابراین، به هنگام تفسیر نمره یک فرد در یک آزمون لازم است به فکر خطاهای معیار اندازه‌گیری آزمون بود، بنابراین، بهتر است محدوده نمره در اطراف نمره مشاهده شده را درست کنید (با یک انحراف معیار) که به این محدوده نوار اطمینان می‌گویند، و هرچه خطای معیار اندازه‌گیری بزرگتر باشد نوار پهن‌تر خواهد بود. و هرچه خطای معیار اندازه‌گیری کوچک‌تر باشد، نوار اطمینان باریک‌تر خواهد بود.

نکته: افزایش سؤالات یک آزمون باعث کاهش خطای اندازه‌گیری خواهد شد.

## فرمول اصلاح کاهش

وقتی بخواهیم تأثیر افزایش یا کاهش پایایی دو آزمون X و Y را بر همبستگی بین آن‌ها اصلاح نمائیم.

از روش زیر استفاده می‌کنیم:

$$r'_{xy} = \frac{r_{xy} \sqrt{r'_{xx} \cdot r'_{yy}}}{\sqrt{r_{xx} \cdot r_{yy}}}$$

$r'_{xy}$ : ضریب همبستگی بین دو آزمون X و Y بعد از تغییر ضریب پایایی

$r_{xy}$ : ضریب همبستگی بین دو آزمون X و Y پیش از تغییر ضریب پایایی

$r'_{xx}$ : ضریب پایایی آزمون X پس از تغییر

- $r_{xx}$ : ضریب پایایی آزمون X پیش از تغییر
- $r'_{yy}$ : ضریب پایایی آزمون Y پس از تغییر
- $r_{yy}$ : ضریب پایایی آزمون Y پیش از تغییر [۳]

مثال: دو آزمون X و Y دارای ضریب پایایی ۰،۷ و ۰،۷۵ و ضریب همبستگی ۰،۴ هستند. اگر ضریب پایایی هر دو آزمون ۰،۹ شود، ضریب همبستگی چه تغییری می‌کند. افزایش پیدا می‌کند.

$$r'_{xy} = \frac{0.4\sqrt{0.9 \times 0.9}}{\sqrt{0.7 \times 0.75}} = 0.5$$

تمام روش‌های گفته شده برای آزمون‌های هنجار مرجع است. اگرچه برای آزمون‌های ملاک - مرجع نیز قابل استفاده می‌باشند. در آزمون‌های ملاک - مرجع به ویژه در آزمون‌های حد تسلط به این روش عمل می‌کنیم:

(۱) روش درصد توافق

(۲) روش کاپا

### روش درصد توافق

یک آزمون را دوبار اجرا می‌کنیم و اطلاعات زیر به دست می‌آید:

اجرای دوم		اجرای اول
غیرمسلط	مسلط	
b	a	مسلط
d	c	غیرمسلط

محاسبه پایایی به قرار زیر خواهد بود:

$$r_n = \frac{a+d}{\text{کل افراد}} \times 100$$

a: تعداد افراد مسلط در هر دو آزمون

d: تعداد افراد غیر مسلط در هر دو آزمون [۳]

### روش کاپا

محاسبه پایایی با روش کاپا به قرار زیر خواهد بود:

$$Kappa = \frac{\text{توافق تصادفی} - \text{توافق کلی}}{\text{توافق تصادفی 10}}$$

$$\text{توافق کلی} = \frac{a+d}{N}$$

$$\text{توافق تصادفی} = \frac{(a+b)(a+c) + (d+b)(d+c)}{N} \quad [۳]$$

## عوامل مؤثر در افزایش پایایی آزمون‌های ملاک مرجع

- ۱) آزمونی بسازید که بین میانگین نمرات آن و نمره حد تسلط تفاوت محسوس باشد.
- ۲) برای هر هدف آموزشی سؤال‌های زیادتری بنویسید.
- ۳) هدف‌ها را دقیق مشخص کنید.
- ۴) از روش عینی نمره گذاری کنید.

## عوامل مؤثر در افزایش پایایی آزمون‌های هنجار - مرجع

- ۱) افزودن تعداد سؤالات
- ۲) تجانس بیشتر سؤالات
- ۳) تجانس کمتر آزمودنی‌ها
- ۴) افزایش تعداد سؤالات با ضریب تمیز زیاد
- ۵) افزایش تعداد سؤالات با ضریب دشواری متوسط
- ۶) آزمون‌های سرعت ضریب پایایی بیشتری از آزمون‌های قدرت نشان می‌دهند.

### گام نهم

به دست آوردن شواهد اعتبار (روایی)

روایی یعنی آن چیزی را که آزمون می‌سنجد، همانی باشد که ما می‌خواهیم، یعنی به هدفی که آزمون برای تحقق بخشیدن به آن درست شده است، اشاره می‌کند. به عبارت دیگر روایی را می‌توان به صورت توافق بین نمره آزمون و کیفیتی که قرار است اندازه بگیرد، تعریف کرد. [۳]

مثال: اگر به طور نظری بپذیریم که استعداد ریاضی از خلاقیت متفاوت است، مشاهده همبستگی بالا بین نمرات یک آزمون استعداد ریاضی و آزمون خلاقیت نشان دهنده روایی پایین است.

### رابطه بین روایی و پایایی

در نظر بگیرید، قصد داریم یک آزمون هوش بسازیم که از خرده آزمون‌های زیر تشکیل شده: سیالی کلامی، سیالی عددی، هوش تصویری، ادراک کلامی. برای هر خرده آزمون تعدادی سؤال طرح می‌کنیم، هر کدام از این سؤال‌ها برای سنجش آن چیزی که می‌خواهیم دارای روایی بالایی است، ولی در تعریف پایایی گفتیم آزمون باید یک چیز را بسنجد، در حالی که، آزمون حاضر ۵ مقوله را می‌سنجد. بنابراین، همبستگی سؤال‌ها در هر خرده آزمون با خودشان بالاست ولی هر خرده آزمون با خرده آزمون دیگر همبستگی پایینی دارد. بنابراین، در آزمون‌هایی که فقط یک چیز را می‌سنجد، روایی و پایایی رابطه مستقیمی دارند. در چنین وضعیتی پایایی همواره بزرگ‌تر یا مساوی روایی می‌باشد.

اما در صورتی که آزمون بیش از یک چیز را بسنجند، که عموماً پدیده‌ها در پژوهش‌های رفتاری چند بعدی می‌باشند، بنابراین، در چنین موقعیت‌هایی روایی و پایایی رابطه معکوس دارند.

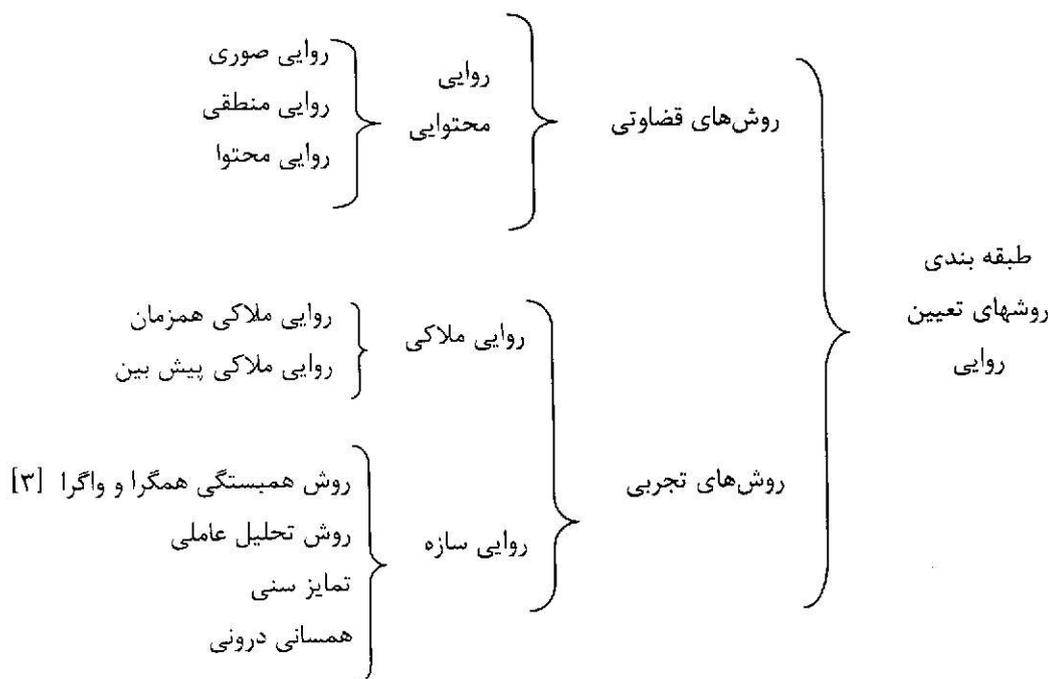
نکته: در پایایی پایین رابطه روایی و پایایی مستقیم است.

نکته: در پایایی بالا رابطه روایی و پایایی معکوس است.

نکته: بدون وجود پایایی، روای معنایی ندارد.

## انواع روایی

انواع روایی عبارتند از: روایی محتوایی<sup>۱۹</sup>، روایی ملاکی<sup>۲۰</sup> و روایی سازه<sup>۲۱</sup>. که به قرار زیر طبقه بندی می‌شوند:



## روایی محتوایی

به این مطلب اشاره دارد که نمونه سؤال‌های مورد استفاده در یک آزمون تا چه حد معرف کل جامعه سؤال‌های ممکن است که می‌توان از محتوا یا موضوع مورد نظر تهیه کرد.

بهترین روش برای این کار، تهیه جدول مشخصات محتوا - هدف می‌باشد. این روش تنها روشی که نیازی به محاسبات آماری ندارد و از قضاوت متخصصان در این باره که سؤال‌ها تا چه حد معرف محتوا و هدف‌های برنامه است، استفاده می‌شود. برای تهیه آزمون‌های پیشرفت تحصیلی مناسب‌ترین روش تعیین روایی، روایی محتوایی می‌باشد. [۳]

<sup>19</sup> Content Validity

<sup>20</sup> Criterion Validity

<sup>21</sup> Construct Validity

## روایی صوری

سؤال‌ها شبیه موضوعی باشند که برای اندازه‌گیری آن تهیه شده‌اند. پس در واقع نوعی ویژگی آزمون است که از لحاظ قابلیت پذیرش برای آزمون شوندگان دارای اهمیت است. و بر انگیزش آزمون شوندگان جهت پاسخگویی اثر می‌گذارد. محاسبه آماری ندارد و از نظر متخصصان استفاده می‌شود. [۳]

## روایی منطقی

شکل پیچیده‌تر روایی صوری است، نوعی ویژگی آزمون است که از لحاظ معقول بودن آزمون برای آزمون شوندگان دارای اهمیت است. محاسبه آماری ندارد و از نظر متخصصان استفاده می‌شود.

## روایی ملاکی

میزان ارتباط بین نمرات حاصل از یک آزمون با نمرات حاصل از ملاک می‌باشد. مهم‌ترین ویژگی متغیری که به عنوان ملاک در تعیین روایی یک آزمون مورد استفاده قرار می‌گیرد، مربوط بودن ملاک می‌باشد.

اگر هر دو آزمون در یک زمان روی یک گروه اجرا شود و همبستگی بین آن دو محاسبه شود، روایی ملاکی همزمان می‌باشد. اگر نمرات آزمونی که به صورت ملاک به کار خواهد رفت، پس از گذشت یک فاصله زمانی از اجرای آزمون اول، گردآوری شده و همبستگی این دو محاسبه گردد، به آن روایی ملاکی پیش بین می‌گویند.

معمولاً برای آزمون‌هایی که هدف آن پیش بینی آینده می‌باشد، مثل آزمون‌های استعداد از این روش استفاده می‌گردد. [۳] برای برآورد روایی ملاکی از دو روش استفاده می‌کنیم: ۱- استفاده از روش همبستگی بین دو آزمون. ۲- استفاده از روش استودنت بین دو دسته نمره که یکی حاصل از اجرای آزمون بر گروه عادی است و دیگری از گروه دارای خصیصه مورد اندازه‌گیری گردآوری شده است.

**مثال:** آزمونی که انگیزه درونی یادگیری را می‌سنجد، در موقعیت حذف پاداش‌های بیرونی، دانش آموزان با سطح بالای یادگیری را از دانش آموزان با سطح پایین تفکیک می‌کند، لذا می‌توان گفت که این آزمون دارای روایی ملاکی است.

## تفاوت روایی همزمان با روایی پیش بین

تفاوت اساسی روایی همزمان با روایی پیش بین در هدف‌های اندازه‌گیری است. روایی پیش بین به موقعیتی اشاره می‌کند که در آن داده‌های مربوط به ملاک در آینده به دست می‌آید. بنابراین نمره‌های آزمون واقعاً برای پیش بینی به کار می‌رود. اما در روایی همزمان هدف آن جانشین ساختن آزمون پیش بین با آزمون ملاک است. [۳]

## روایی سازه

تعیین تعداد و ماهیت ویژگی‌ها یا سازه‌هایی که زیر بنای مجموعه‌ای نمره آزمون را تشکیل می‌دهند که در اندازه‌گیری‌های مربوط به تئوری مطرح می‌شود. بنابراین روایی سازه، مربوط به تولید یا به کارگیری یک تئوری است. [۳] بنابراین اولین قدم در تعیین روایی سازه، تعریف سازه است.

**مثال:** مطابق برخی نظریه‌های یادگیری، درماندگی آموخته شده بر اثر شکست‌های پیاپی حاصل می‌شود، اگر بخواهیم برای درماندگی آموخته شده یک آزمون بسازیم، این آزمون باید دارای روایی سازه باشد.  
برای تعیین روایی سازه می‌توان راه‌های زیر را استفاده کرد:

### (۱) تعیین همبستگی با آزمون‌های همگرا و واگرا:

اگر یک آزمون با آزمون مشابه همبستگی مثبت نشان دهد و با آزمون نامربوط همبستگی نشان ندهد؛ در این صورت همبستگی با آزمون همگرا دارد و با آزمون واگرا همبستگی ندارد. [۲]

### (۲) روش تحلیل عاملی:

با این روش تعداد و ماهیت متغیرهایی را که یک آزمون اندازه می‌گیرد، مشخص می‌کنیم آزمون‌ها معمولاً دارای یک عامل عمومی و یک تا چند عامل اختصاصی دارند، یا دارای چند عامل گروهی می‌باشند. [۸]

**مثال:** اگر یکی از سؤال‌های آزمون با سایر سؤال‌ها همبستگی نداشته باشد، نتیجه تحلیل عاملی نشان دهند یک عامل عمومی و یک عامل اختصاصی است.

**مثال:** اگر ضریب همبستگی دو بخش یک آزمون نزدیک به صفر باشد ولی سؤالات هر بخش باهم رابطه داشته باشند؛ نتیجه تحلیل عاملی نشان دهنده دو عامل گروهی است.  
روش تحلیل عاملی به دو صورت کلی انجام می‌گیرد:

**الف: تحلیل عامل اکتشافی:** در این روش از عوامل موجود در آزمون اطلاعاتی نداریم و تازه می‌خواهیم آن‌ها را کشف نماییم.

**ب: تحلیل عامل تأییدی:** در این روش از عوامل موجود در آزمون خبر داریم و حالا می‌خواهیم مجدداً آن عوامل را ارزیابی کنیم.

### (۳) تمایز سنی:

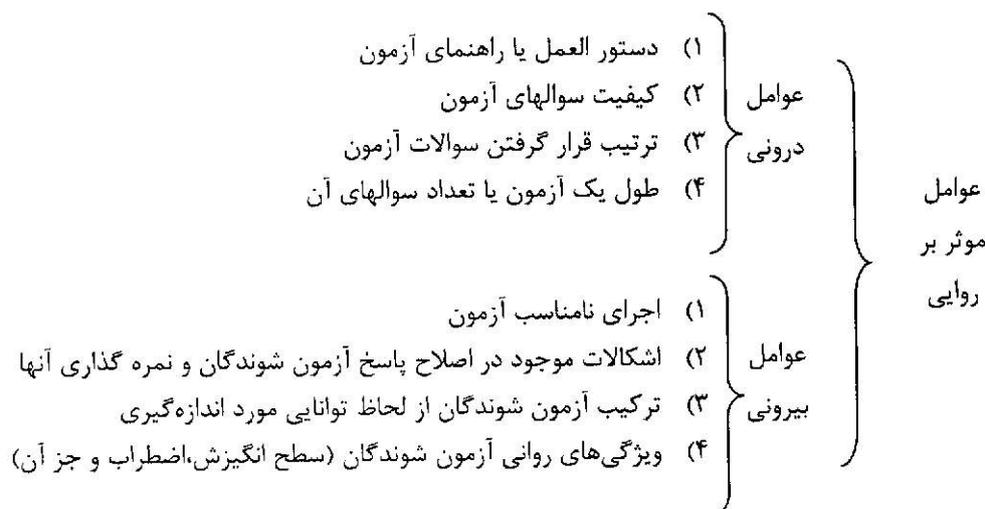
اگر یکی از ویژگی‌های انسان با بالا رفتن سن، تغییر نشان دهد. نمرات آزمون باید نشان دهنده این تغییرات باشد. (مثل آزمون‌های هوش و کسلر و بینه)  
تمام آزمون‌های توانایی هنجار مرجع این ویژگی را دارند.

### (۴) همسانی درونی:

خرده آزمون‌های یک آزمون باید با همبستگی کم و با نمره کل آزمون همبستگی بالایی داشته باشند.

**مثال:** زمانی که همبستگی نمرات درس ریاضی را با معدل محاسبه می‌نماییم، از روش همسانی درونی؛ روایی را تعیین کرده ایم.

**نکته:** اگر پایایی کل آزمون را بخواهیم، باید پایایی اختلاف را محاسبه کنیم.



## گام دهم

### هنجارها و نیم رخ ها

#### هنجارهای مربوط به نمرات معیار

اساس ارزشیابی مقایسه است، مقایسه با یک مرجع می‌شود ملاک مرجع و مقایسه با یک گروه می‌شود هنجار مرجع. پس اگر این مقایسه صورت بگیرد هنجاریابی صورت گرفته است.

هنجاریابی یکی از مراحل اصلی استاندارد کردن آزمون‌های وابسته به هنجار است.

نمرات معیار عملکرد هر دانش آموز را در یک آزمون بنا بر اختلاف عملکرد او از میانگین گروه، بر حسب واحد انحراف معیار نشان می‌دهد. بنابراین، هنجار عملکرد قابل انتظار است. و هنجاریابی نمرات را بر روی توزیع نرمال تطبیق دادن است.

مقیاس تمام تبدیل‌های خطی فاصله‌ای است چون همه بر مبنای  $Z$  که فاصله‌ای می‌باشد، است.

زمانی که ما نیاز به انتخاب یا تشخیص داشته باشیم تنها چاره ما روش هنجار - مرجع است.

در آزمون‌های شخصیت و عاطفی نیز تنها روش هنجاریابی روش هنجار - مرجع می‌باشد.

چگونه هنجار - مرجع را انجام می‌دهیم:

نمرات میانگین و انحراف معیار آزمودنی‌ها (گروهی بزرگ) را تبدیل به نمرات خطی می‌کنیم. تک تک نمرات را تبدیل به نمره  $Z$  می‌کنیم.

برای مثال در یک آزمون ۲۰ نمره‌ای، نمره‌های از صفر تا ۲۰ را تک تک به نمره  $Z$  تبدیل کرده و در جدول هنجار قرار می‌دهیم.

اما چون نمره  $Z$  اعشار زیاد دارد اغلب نمره  $Z$  را به نمرات خطی زیر تبدیل می‌کنند:

## تبدیل‌های خطی

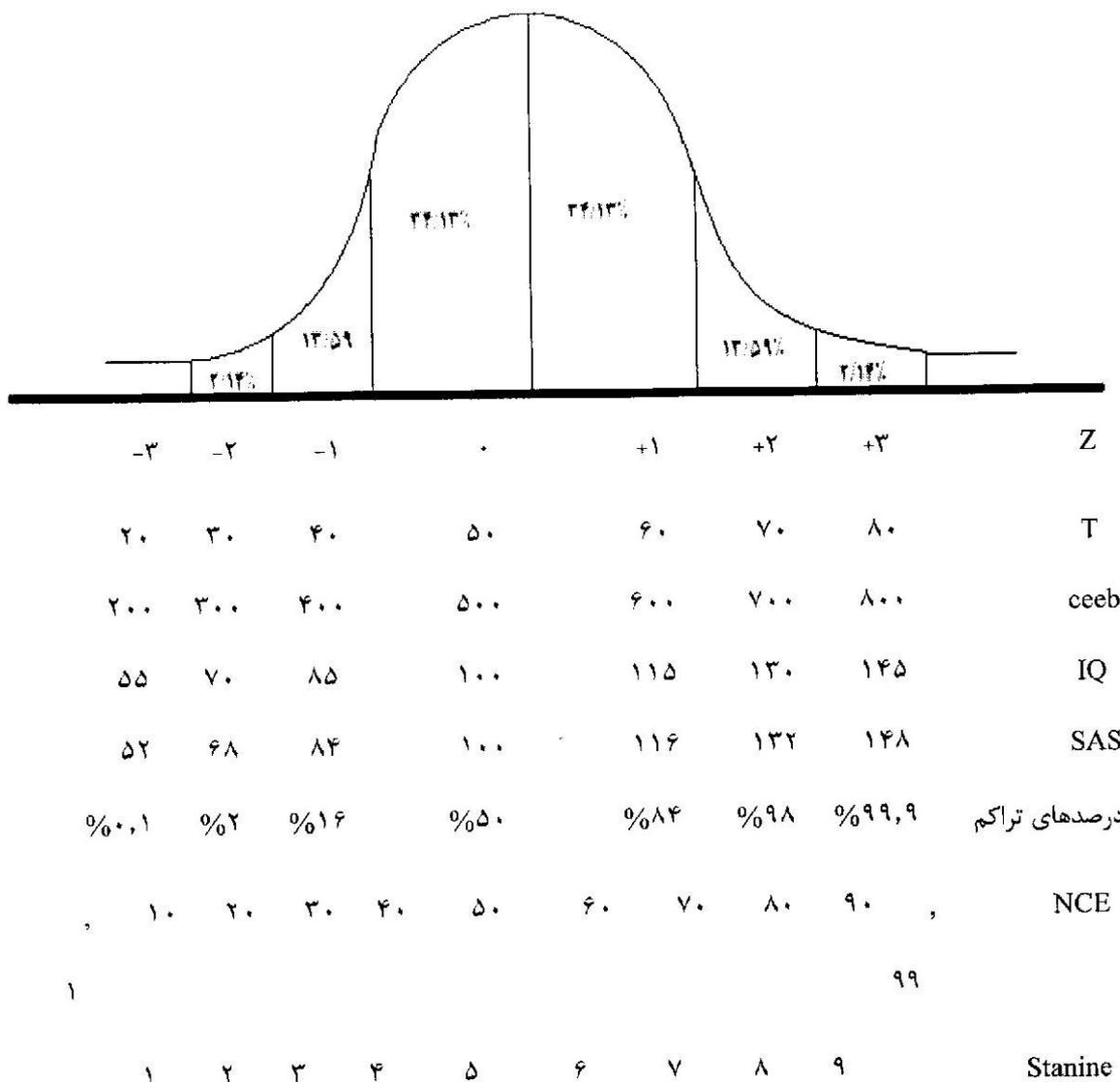
$Z = \frac{x - \bar{x}}{S}$	$\bar{x} = 0$	$S = 1$
$T = 10Z + 50$	$\bar{x} = 50$	$S = 10$
$SAT = 100Z + 500$	$\bar{x} = 500$	$S = 100$
$^{\text{۲۲}} S \text{ tanine} = 2Z + 5$	$\bar{x} = 5$	$S = 2$
$IQ = DIQ = 15Z + 100$	$\bar{x} = 100$	$S = 15$
$SAS = 16Z + 100$	$\bar{x} = 100$	$S = 16$
$^{\text{۲۳}} NCE = 21.06Z + 50$	$\bar{x} = 50$	$S = 16$

تبدیل‌های خطی تنها مربوط به آزمون‌های هنجار می‌باشد. در بهترین حالت در نمرات معیار موقعیت نسبی یک دانش آموز در گروهی که خود او عضوی از آن است نشان داده می‌شود.

<sup>22</sup> برای آزمون‌های تشریحی با ارزشیابی کیفی (مثل آزمون‌های خواندن و فهمیدن)

<sup>23</sup> معادل منحنی بهنجار (این نمرات در انتها بازتر و در وسط میانه مقیاس جمع ترند و از ۱ تا ۹۹ می‌باشند).

شکل ۴- انواع تبدیل‌های خطی



در تعیین هنجار برای یک آزمون باید سؤال زیر را مورد توجه قرار داد: آزمون در مورد چه گروهی به کار خواهد رفت؟

مثال: در یک توزیع بهنجار، با میانگین ۳۴ و انحراف استاندارد ۴، نمره خام ۴۴ معادل کدام نمره در مقیاس T است؟

$$Z = \frac{44 - 34}{4} = 2.5$$

$$T = 10 \times 2.5 + 50 = 75$$

مثال: در منحنی Z صدک ۳۵ بر روی کدام انحراف معیار قرار دارد؟ بین صفر و ۱-

مثال: رتبه درصدی دانش آموزی که نمره استاندارد او در توزیع نرمال برابر با ۱+ محاسبه شده است، چقدر می‌باشد؟ ۸۴ درصد

مثال: در آزمون هایی که طول آنها کوتاه و توزیع آنها نرمال باشد، برای آزمودنی هایی که دتبه درصد آنها در مرکز است، تفاوت زیاد نمره‌ها در مرکز توزیع، تفاوت کم در نقاط درصدی ایجاد می‌کند.

## تبدیل‌های غیر خطی

(۱) هنجار سنی

(۲) هنجار کلاسی

(۳) هنجار درصدی (رتبه درصدی)

برخی از آزمون‌های ملاک مرجع را مایل هستیم که از آن‌ها هنجار داشته باشیم یعنی بتوانیم با گروهی بزرگ مقایسه کنیم و این‌ها داده‌هایشان نیز نرمال نیست پس از تبدیل‌های غیر خطی استفاده می‌کنیم. در هر کدام از هنجارها ترتیب نمره از کم به زیاد است. (یعنی نمره کمتر ← بهتر)

### (۱) هنجار سنی

هنجار سنی یا معادل سنی بر میانگین یا میانه نمرات که دانش‌آموزان در سنین مختلف کسب می‌کنند و به صورت معادل‌های سنی نشان داده می‌شود.

بنابراین هنجار سنی مربوط به سنین مختلف در یک ویژگی خاص را (مثل قد)، از راه دادن آزمون مربوط به آن ویژگی به افراد نمونه سنین مختلف و محاسبه میانگین یا میانه نمرات آنان تعیین می‌کنند. مثلاً تعیین هنجار قد پسران دبستانی ← نمونه کاملی از پسران ۶-۱۱ سال ← اندازه‌گیری قد ← محاسبه میانگین یا میانه

محدودیت این روش: واحدهای نامساوی، چون تغییرات الگوهای رشد از یک توانایی به توانایی دیگر سبب این تفاوت واحدها می‌شود. بهترین مرحله کاربرد: سطح دبستان، چون رشد ذهنی و تحصیلی یکنواخت‌تر است.

اکثر آزمون‌های استاندارد پیشرفت تحصیلی دارای هنجار سنی هستند.

گروه مرجع ← گروه‌های مختلف سنی

### ۲-۱- هنجار سن ذهنی

همان سن عقلی بینه که در آزمون‌های هوشی به کار می‌رود.

### (۲) هنجار کلاسی

شبه هنجار سنی با تفاوت در گروه مرجع که در این جا گروه‌های کلاسی می‌باشد.

آزمون را به گروه‌های نمونه پایه‌ها یا کلاس‌های مختلف می‌دهند و میانگین یا میانه نمرات کلاس‌ها را به عنوان معادل کلاسی یا هنجار کلاسی به کار می‌برند و آن سطح مقدار توانایی ۷۰٪ بچه‌ها می‌باشد سپس برای هر سطح باید یک آزمون تهیه کرد. به سه شکل این کار انجام می‌شود.

سطحی‌ترین شکل ← برای هر کلاس یک سطح

شکل متوسط ← برای هر کلاس ۳ سطح یا ۲ سطح (تعداد ترم‌ها)

کامل‌ترین شکل ← برای هر کلاس ۹ سطح (تعداد ماه‌های سال تحصیلی)

و در کامل‌ترین حالت معادل کلاسی (هنجار کلاسی) را به صورت دو رقمی نشان می‌دهند که رقم اول سمت راست نشان دهنده سال و رقم دوم نشان دهنده ماه است که از صفر تا ۹ می‌باشد.

مثال: ۸.۳ ← پیشرفت در سطح ماه سوم از کلاس هشتم

محدودیت: واحدهای نامساوی، چون پیشرفت در موضوعهای درسی و در هر درس نسبت به درس دیگر فرق می‌کند. بهترین مرحله کاربرد: سطح دبستان و درس‌هایی که تأکید آموزشی یکسان بر آن‌ها می‌شود. مثل: حساب مهارت‌های زبان و خواندن

### ۱-۲- هنجار سنی نمایی

چون دامنه سنی دانش آموزان یک کلاس بزرگ است هنگام هنجاریابی سن‌های کوچک و بزرگ را برمی‌دارند و ۷۰٪ وسط را برمی‌دارند (۱۵٪ دو سر طیف را حذف می‌کنند).

### ۳) هنجار درصدی

در هنجار درصدی (رتبه درصدی PR) فرد با گروهی که منطقاً عضوی از آن است. مقایسه می‌شود و از رتبه‌های درصدی یا صدک‌ها استفاده می‌شود که وضعیت نسبی فرد در گروه را بر حسب کسانی که نمره پایین‌تر از او گرفته‌اند مشخص می‌کند. معادل هر نمره خام به صورت هنجاری ← رتبه درصدی آن نمره است. صدک وضعیت فرد را در روی توزیع و رتبه درصدی عملکرد نسبی فرد را نشان می‌دهد. در مورد صدک‌ها شمارش از پایین شروع می‌شود و هر اندازه صدک کوچک‌تر به همان نسبت جایگاه فردی که صدک معرف نمره خام اوست در گروه پایین‌تر است.

P = صدک شصت و پنجم ← یعنی رتبه درصدی ۶۵٪ ← یعنی عملکرد از ۶۵٪ افراد بهتر است.

صدک ۱۵ ← رتبه ۱۵٪ ← عملکرد از ۱۵٪ افراد بهتر است.

صدک ۵۰ = P ۵۰ = میانه

به چه صورت رتبه درصدی را به دست می‌آوریم؟

برای نمرات فراوانی نسبی می‌گیریم  $f_i = \frac{F_i}{n}$  و درصد ضرب کرده فراوانی درصدی به دست می‌آوریم  $(Fc_i = f_i \times 100)$  سپس

فراوانی تراکمی درصدی را محاسبه می‌کنیم  $(Pc_i = \frac{Fc_i}{n} \times 100)$  که این عدد به دست آمده نشان دهنده رتبه درصدی که نمره‌ای هنجار است، می‌باشد.

رتبه‌های درصدی داده‌های نرمال با نمره محاسبه شده توسط معادل منحنی بهنجار (NCE) یکسان است.

### عیب:

(۱) واحدهای نامساوی (تفاوت مدل‌های اواسط مقیاس کمتر از دو انتهای مقیاس)

(۲) برای رتبه درصدی نمی‌توان میانگین محاسبه کرد.

نکته: وقتی نمره فردی را در یک آزمون استاندارد شده پیشرفت تحصیلی به صورت رتبه درصدی ارائه می‌نماییم، معنی آن این است که فرد را با افراد گروه هنجار به صورت درصدی مقایسه می‌کنیم.

نکته: درصد عملکرد یعنی نسبت نمره به دست آمده از کل نمره ضرب در ۱۰۰.

مثال: فردی در یک آزمون ۶۰ سوالی نمره ۴۵ به دست آورده است. درصد عملکرد او در کل آزمون چقدر است:

$$\frac{45}{60} \times 100 = 75\%$$

## سؤالات ۵۰ درصد دوم

۱ - منظور از استاندارد بودن آزمون در روان سنجی، عبارت است از یکسان کردن:

- (الف) سؤالات و آزمودنی ها  
(ب) سؤالات آزمون بر اساس محتوا  
(ج) نمرات بر اساس نمره Z  
(د) مواد آزمون، اجرا و نمره گذاری

۲ - اولین قدم در تعیین روایی (Validity) سازه چیست؟

- (الف) تعریف سازه مورد نظر  
(ب) مشخص کردن نوع آزمون  
(ج) تعیین اندازه پذیر بودن سازه  
(د) تعیین روش اندازه گیری سازه

۳ - رتبه درصدی دانش آموزی که نمره استاندارد او در توزیع نرمال برابر با ۱+ محاسبه شده است، چقدر است؟

- (الف) ۱۶٪ (ب) ۳۲٪ (ج) ۵۰٪ (د) ۸۴٪

۴ - مطابق جدول زیر، درصد توافق معلمان در مورد اوراق زبان چقدر است؟

قبول از نظر همه معلمان	مردود از نظر همه معلمان	قبول از نظر گروهی دیگر
۱۵	۲۵	۶۰

- (الف) ۴۰٪ (ب) ۳۳٪ (ج) ۴۵٪ (د) ۶۰٪

۵ - در تعیین کدام یک از انواع روایی (Validity) از محاسبات آماری استفاده نمی شود؟

- (الف) همزمان (ب) پیش بین (ج) محتوا (د) سازه

۶ - منظور از پایایی آزمون کدام است؟

- (الف) همسانی در اندازه گیری  
(ب) کاهش خطای اندازه گیری  
(ج) بررسی هدف مورد اندازه گیری  
(د) یکسانی محتوای آزمون

۷ - پایایی یک آزمون از طریق دو نیمه کردن معادل ۰.۷۰ است. اعتبار کل آزمون چقدر است؟

- (الف) ۰.۳۵ (ب) ۰.۷۵ (ج) ۰.۸۲ (د) ۱.۰۴

۸ - تفاوت اساسی روایی همزمان و روایی پیش بین در چیست؟

- (الف) نوع آزمون  
(ب) هدفهای اندازه گیری  
(ج) نمره گذاری آزمون  
(د) زمان اجرای آزمون

۹ - در محاسبه کدام شاخص، از خطای استاندارد اندازه گیری استفاده می شود؟

- (الف) محاسبه ضریب تمرکز کل آزمون  
(ب) محاسبه ضریب دشواری آلفای کرونباخ  
(ج) محاسبه روایی (Validity)  
(د) محاسبه ضریب پایایی (Reliability)

۱۰ - اگر ضریب پایایی آزمون افزایش یابد، واریانس خطا:

- الف) کاهش می‌یابد  
ب) افزایش می‌یابد  
ج) تغییر نمی‌کند  
د) به همان میزان افزایش می‌یابد

۱۱ - یک آزمون ۱۸۰ سؤالی را به ۶ آزمون ۳۰ سؤالی تقسیم کرده ایم. اگر بخواهیم ضریب پایایی همسانی درونی این آزمون را حساب کنیم، بهترین روش کدام است؟

- الف)  $KR_{20}$       ب)  $KR_{21}$       ج) آلفای کرونباخ      د) فرمول اسپیرمن-براون

۱۲ - آزمون‌های پیشرفت تحصیلی بیشتر با کدامیک از انواع روایی (Validity) ارتباط دارد؟

- الف) سازه      ب) محتوا      ج) همزمان      د) پیش بین

۱۳ - فرمول کودر-ریچاردسون ۲۰ زمانی مناسب‌ترین کارایی را دارد که .....

- الف) سطح دشواری آزمون مناسب باشد  
ب) ضریب تمیز آزمون مناسب باشد  
ج) اختلاف بین دشواری سؤال‌ها زیاد باشد  
د) سؤال‌ها حداقل چهار ارزش داشته باشند

۱۴ - واریانس‌های هریک از چهار سؤال یک آزمون انشایی به ترتیب ۰.۹، ۰.۴۸، ۱.۰۲ و ۱.۶ محاسبه شده است. اگر واریانس کل سؤالات برابر ۱۰۰ باشد، ضریب پایایی آزمون چقدر خواهد شد؟

- الف) ۰.۶۵      ب) ۰.۷۵      ج) ۰.۸۰      د) ۰.۸۵

۱۵ - در کدام یک از روش‌های تعیین پایایی، واریانس خطای ناشی از حافظه وجود ندارد؟

- الف) دو نیمه کردن آزمون  
ب) اجرای مجدد آزمون  
ج) فرم‌های همتا با فاصله زمانی  
د) فرم‌های همتا بدون فاصله زمانی

۱۶ - افزودن تعدادی سؤال با ضریب دشواری ..... و ضریب تمیز ..... پایایی سؤالات آزمون را افزایش می‌دهد.

- الف) زیاد - متوسط  
ب) زیاد - بالا  
ج) متوسط - متوسط  
د) متوسط - بالا

۱۷ - چنانچه  $r_{xx'} = 0$  باشد، انتظار می‌رود که رابطه ..... برقرار باشد.

- الف)  $S_T^2 = S_E^2$       ب)  $S_X^2 = S_E^2$       ج)  $S_X^2 = S_T^2$       د)  $S_X^2 = S_T^2$

۱۸ - هرگاه در یک آزمون هوشی، انحراف معیار ۱۰، ضریب پایایی (Reliability) ۰.۷۵ و نمره احمد ۹۰ باشد، کدام یک از احکام زیر در تفسیر نمره صحیح است؟

- الف) به احتمال ۳۴ درصد نمره واقعی احمد بین ۸۵ و ۹۵ است.  
ب) به احتمال ۶۸ درصد نمره واقعی احمد بین ۸۰ و ۱۰۰ است.  
ج) به احتمال ۹۵ درصد نمره واقعی احمد بین ۷۵ و ۱۰۵ است.  
د) به احتمال ۹۹ درصد نمره واقعی احمد بین ۷۵ و ۱۰۵ است.

۱۹ - کدام هنجارها از راه تبدیل خطی به دست می‌آید؟

- الف) هوشبهر انحرافی و Z  
ب) نمرات معیار و صدک‌ها  
ج) هوشبهر انحرافی و هوشبهر نسبی  
د) معادل‌های کلاسی و هوشبهر نسبی

۲۰ - در منحنی Z صدک ۳۵ بر روی کدام انحراف معیار قرار دارد؟

الف) بین ۰ و ۱- (ب) بین ۰ و ۱+ (ج) بین ۱- و ۲- (د) بین ۲- و ۳-

۲۱ - در کدام یک از روشهای تعیین پایایی (Reliability)، تمرین بیشترین تاثیر را می تواند در پراش خطای آزمون داشته باشد؟

الف) تنصیف (ب) فرمهای همنا  
ج) باز آزمایی (د) کودر - ریچاردسون

۲۲ - برآورد پایایی (Reliability) به کمک ثبات درونی، در وضعیتی بهتر برآورد می شود که آزمون:

الف) دارای سوالهای متجانس و همگون باشد. (ب) به قسمت هایی موازی قابل تقسیم نباشد.  
ج) دارای سوالهای قابل تفکیک نباشد. (د) دارای سوالهای قابل تفکیک و مستقل باشد.

۲۳ - مقادیر پایایی حاصل از دو روش  $KR_{20}$  و  $KR_{21}$  در حالتی برابر خواهند بود که دشواری همه سوالهای آزمون ..... باشند در

غیر این صورت مقادیر  $KR_{21}$  ..... از مقدار  $KR_{20}$  خواهد شد.

الف) یکسان - کوچکتر (ب) یکسان - بزرگتر (ج) برابر یک - کوچکتر (د) برابر یک - بزرگتر

۲۴ - اگر یکی از سوالهای آزمون با سایر سوالها همبستگی نداشته باشد، نتیجه تحلیل عاملی سوالها چگونه خواهد بود؟

الف) دو عامل اختصاصی (ب) دو عامل عمومی و یک عامل گروهی  
ج) دو عامل گروهی (د) یک عامل عمومی و یک عامل اختصاصی

۲۵ - با افزودن نسبت گزینش داوطلبان یک شغل، دقت پیش بینی ..... و با افزایش روایی (Validity) آزمون، دقت پیش بینی ..... می یابد.

الف) کاهش - کاهش (ب) کاهش - افزایش (ج) افزایش - افزایش (د) افزایش - کاهش

۲۶ - در تبدیل خطی نمره های خام «شکل توزیع نمره های تبدیل شده» و «اندازه همبستگی» به ترتیب:

الف) تغییر می کند - تغییر می کند (ب) تغییر می کند - تغییر نمی کند  
ج) تغییر نمی کند - تغییر می کند (د) تغییر نمی کند - تغییر نمی کند

۲۷ - معرف بودن یک هنجار به ..... مربوط می شود.

الف) حجم و روش نمونه گیری (ب) تعریف دقیق و عینی هنجارها  
ج) تازگی و جدید بودن هنجار یابی (د) مربوط مناسب بودن گروه هنجاری

۲۸ - چنانچه  $0 < R_{XX'} < 1$  باشد، آنگاه کدام یک از رابطه های زیر صحیح است؟

الف)  $S_X^2 = S_E^2$  (ب)  $S_X^2 = S_T^2$  (ج)  $S_X^2 = S_T^2 + S_E^2$  (د)  $S_T^2 = S_X^2 + S_E^2$

۲۹ - خطای استاندارد اندازه گیری آزمونی برابر ۲/۵ است. دانش آموزی در این آزمون نمره ۸۴ کسب کرده است. نمره حقیقی این دانش آموز به احتمال ۹۵ درصد:

الف) همان ۸۴ است (ب) بین ۷۹ تا ۸۹ است  
ج) بین ۸۱،۵ و ۸۶،۵ است (د) بین ۷۶،۵ و ۹۱،۵ است

۳۰ - کدام یک از انواع روایی (Validity) شکل پیچیده تری از روایی صوری است؟

الف) منطقی (ب) ملاکی (ج) پیش بین (د) سازه

۳۱ - هر گاه یک آزمون، موضوعات متفاوتی را اندازه گیری کند استفاده از کدام روش احراز پایایی منجر به ضریبی کمتر از روش های دیگر خواهد شد؟

الف) تنصیفی (ب) باز آزمایی (ج) کودر - ریچاردسون (د) فرم های موازی

۳۲ - وقتی نمره های فردی را در یک آزمون استاندارد شده پیشرفت تحصیلی به صورت رتبه درصدی ارائه می نماییم، معنی آن این است که:

الف) درصد پاسخ های صحیح فرد را تعیین می کنیم.

ب) فرد را با افراد گروه هنجار به صورت درصدی مقایسه می کنیم.

ج) درصد افرادی را که از او نمره بیشتر کسب کرده اند، معین می کنیم.

د) درصد پرسش های درست او را نسبت به پرسش های نا درست تعیین می کنیم.

۳۳ - تعیین روایی پیش بین در کدام یک از انواع آزمون ها بهتر است؟

الف) شخصیت (ب) مهارت (ج) پیشرفت (د) استعداد

۳۴ - در کدام یک از گروه های زیر ضریب پایایی یک آزمون واحد بیشتر است؟

الف) پسران ۱۲ ساله (ب) دختران بالای ۱۵ سال

ج) مردان بالای ۲۵ سال (د) مردان و زنان بالای ۲۰ سال

۳۵ - اگر ضریب همبستگی دو نیمه یک آزمون نزدیک به صفر باشد، ولی سوالات هر نیمه با هم رابطه داشته باشند، نتایج تحلیل عاملی سوالات چگونه خواهد بود؟

الف) دو عامل گروهی

ب) دو عامل اختصاصی

ج) یک عامل مهم عمومی

د) یک عامل گروهی و یک عامل اختصاصی

۳۶ - برای مقایسه کدام یک از موارد زیر بهتر است از آزمون های استاندارد شده استفاده شود؟

الف) دو نظام یا شیوه آموزشی متفاوت

ب) دانش آموزان یک کلاس یا یکدیگر

ج) دانش آموزان دو کلاس مختلف

د) نتایج یک فرد در دو زمان مختلف

۳۷ - در تبدیل نمرات آزمون های شخصیت، کدام یک از مقیاس های ذیل بیشتر مورد استفاده قرار می گیرد؟

الف) میانگین ۵، انحراف معیار ۲

ب) میانگین ۵۰، انحراف معیار ۱۰۰

ج) میانگین ۱۰۰، انحراف معیار ۱۵

د) میانگین ۵۰، انحراف معیار ۱۰

۳۸ - در کدام یک از روش های تعیین پایایی (Reliability) آزمون تشریحی، به کارکرد افراد در تک تک سوالات توجه می شود؟

الف) باز آزمایی (ب) تنصیف (ج) فرم های همتا (د) آلفای کرونباخ

۳۹ - اگر ضریب پایایی (Reliability) آزمون برابر یک باشد، کدام یک از گزینه‌های زیر درست است؟

- الف) نمره مشاهده شده برابر نمره خطا است  
ب) نمره مشاهده شده بزرگتر از نمره واقعی است  
ج) نمره مشاهده شده برابر نمره واقعی است  
د) نمره واقعی برابر نمره خطا است

۴۰ - چه رابطه‌ای بین ضریب پایایی (Reliability) و روایی (Validity) آزمون وجود دارد؟

- الف) پایایی آزمون همواره بزرگتر یا مساوی روایی می‌باشد  
ب) پایایی آزمون همواره کوچکتر از ضریب روایی می‌باشد  
ج) پایایی آزمون همواره بزرگتر از روایی می‌باشد  
د) پایایی آزمون کوچکتر یا مساوی ضریب روایی می‌باشد.

پاسخنامه سوالات ۵۰ درصد اول							
سوال	پاسخ	سوال	پاسخ	سوال	پاسخ	سوال	پاسخ
۱	ج	۱۱	ب	۲۱	ج	۳۱	ج
۲	الف	۱۲	ج	۲۲	د	۳۲	ج
۳	ب	۱۳	الف	۲۳	الف	۳۳	ب
۴	ج	۱۴	ب	۲۴	الف	۳۴	ب
۵	د	۱۵	ج	۲۵	الف	۳۵	الف
۶	ب	۱۶	ج	۲۶	الف	۳۶	ب
۷	د	۱۷	الف	۲۷	د	۳۷	ب
۸	ب	۱۸	د	۲۸	الف	۳۸	الف
۹	ج	۱۹	ب	۲۹	الف	۳۹	الف
۱۰	الف	۲۰	ج	۳۰	د	۴۰	الف

پاسخنامه سوالات ۵۰ درصد دوم							
سوال	پاسخ	سوال	پاسخ	سوال	پاسخ	سوال	پاسخ
۱	د	۱۱	ج	۲۱	ج	۳۱	ج
۲	الف	۱۲	ب	۲۲	الف	۳۲	ب
۳	د	۱۳	ج	۲۳	الف	۳۳	د
۴	الف	۱۴	ج	۲۴	د	۳۴	الف
۵	ج	۱۵	د	۲۵	ب	۳۵	الف
۶	الف	۱۶	د	۲۶	د	۳۶	ج
۷	ج	۱۷	ب	۲۷	د	۳۷	د
۸	ب	۱۸	د	۲۸	ج	۳۸	د
۹	د	۱۹	ج	۲۹	ب	۳۹	ج
۱۰	الف	۲۰	الف	۳۰	الف	۴۰	الف

1) Colman , Andrew , 2003 , Oxford dictionary of psychology, Oxford university Press, New york

۲) هومن، حیدر علی، ۱۳۷۴، اندازه‌گیری‌های روانی و تربیتی و فن تهیه تست، ناشر مولف، تهران

۳) سیف، علی اکبر، ۱۳۸۲، اندازه‌گیری، سنجش، و ارزشیابی آموزشی، نشر دوران، تهران

۴) سیف، علی اکبر،

۵) بلوم، اس. انگلهارت، ام دی، فرست، ای جی و کراتول دی آر، ۱۳۶۸، طبقه بندی و هدفهای

پرورش، کتاب اول، حوزه شناختی، ترجمه علی اکبر سیف و خدیجه علی آبادی تهران، انتشارات رشد

۶) شعبانی، حسن، ۱۳۷۶، مهارت‌های آموزش و پرورش، چاپ ششم، انتشارات سمت، تهران

۷) صالح صدق پور، بهرام، ۱۳۷۲، هنجاریابی، پایاسازی، و رواسازی آزمون اضطراب مدرسه، پایان نامه

کارشناسی ارشد دانشگاه علامه طباطبائی

۸) آلن، ام جی، وین، دلبیو ام، ۱۳۷۴، مقدمه‌ای بر نظریه‌های اندازه‌گیری (روان سنجی) ترجمه علی

دلاور، تهران: سمت