

# فهرست مطالب

فصل اول. سابقه تاریخی سنجش و اندازه‌گیری و روان‌سنجی .....	۱
۱- پیدایش روان‌شناسی تجربی .....	۲
۲- مطالعه تفاوت‌های فردی .....	۲
۳- مطالعه بالینی افراد غیرعادی .....	۳
خود آزمائی (۱) .....	۷
پاسخ سوال‌های خودآزمائی (۱) .....	۷
فصل دوم. اندازه‌گیری و ارزشیابی در آموزش و پرورش .....	۱۱
ارزشیابی چیست؟ .....	۱۲
مراحل اندازه‌گیری .....	۱۲
فلسفه ارزشیابی .....	۱۵
خودآزمائی (۲) .....	۱۶
پاسخ خودآزمائی (۲) .....	۱۷
فصل سوم. امتحان .....	۲۳
۱- مفهوم امتحان .....	۲۳
۲- انواع امتحان و موارد کاربردی آنها .....	۲۳
۳- هدف‌های امتحان .....	۲۶
۴- مراحل انجام امتحان .....	۲۸
خودآزمائی (۳) .....	۴۶
پاسخ خودآزمائی (۳) .....	۴۷
فصل چهارم. تهیه پرسش‌های امتحان .....	۵۷
۱- پرسش‌های عینی .....	۵۷
۲- پرسش‌های انشائی .....	۶۹
خودآزمائی (۴) .....	۷۳
پاسخ سوال‌های خودآزمائی (۴) .....	۷۴
فصل پنجم. اجرای امتحان .....	۸۵
۱- تنظیم پرسش‌های آزمون .....	۸۵

۸۶	۲- اجرای آزمون.....
۸۸	۳- تصحیح و نمره‌گذاری پرسش‌های آزمون.....
۹۴	۴- تجزیه و تحلیل نتایج امتحان و استفاده از آن در کشف نارسائی‌های دانش‌آموزان.....
۹۸	خودآزمایی (۵).....
۹۹	پاسخ سؤال‌های خودآزمایی (۵).....
۱۰۷	<b>فصل ششم. آزمون و انواع آن</b> .....
۱۰۸	۱- هدف‌های کلی کاربرد آزمون‌ها.....
۱۱۱	۲- طبقه‌بندی آزمون‌ها.....
۱۱۶	۳- فرضیه‌های اساسی سنجش یا ارزشیابی.....
۱۱۸	خودآزمایی (۶).....
۱۱۹	پاسخ سؤال‌های خودآزمایی (۶).....
۱۲۵	<b>فصل هفتم. تجزیه و تحلیل پرسش‌های آزمون</b> .....
۱۲۵	مراحل تجزیه و تحلیل سؤال‌های آزمون.....
۱۲۶	الف) مراحل تجزیه و تحلیل پرسش‌های عینی.....
۱۳۷	ب) تجزیه و تحلیل پرسش‌های انشائی (باز پاسخ).....
۱۳۹	ج) تجزیه و تحلیل سؤال‌های چند ارزشی.....
۱۶۳	خودآزمایی (۷).....
۱۶۶	پاسخ خودآزمایی (۷).....
۱۸۷	<b>فصل هشتم. روائی آزمون</b> .....
۱۸۷	۱- تعریف روائی.....
۱۸۸	۲- شواهد روائی آزمون.....
۱۸۹	۳- انواع روائی.....
۲۰۱	خودآزمایی (۸).....
۲۰۲	پاسخ خودآزمایی (۸).....
۲۱۳	<b>فصل نهم. اعتبار آزمون</b> .....
۲۱۳	۱- مفهوم نظری اعتبار آزمون.....
۲۱۴	۲- تعریف اعتبار.....
۲۱۵	۳- منابع خطای اندازه‌گیری آزمون.....
۲۲۴	۴- اعتبار آزمون‌های سرعت.....
۲۲۴	۵- اعتبار نمره‌گذاری یا درجه‌بندی.....
۲۲۶	۶- اعتبار آزمون‌های ملاک مرجع.....

۲۲۷	۷- روش‌های افزایش ضریب اعتبار آزمون.....
۲۲۹	۸- خطای معیار اندازه‌گیری.....
۲۳۰	۹- برآورد نمره واقعی آزمودنی با استفاده از خطای معیار اندازه‌گیری.....
۲۳۰	۱۰- ضریب اعتبار بهینه.....
۲۳۱	۱۱- رابطه روائی و اعتبار آزمون.....
۲۳۲	۱۲- عملی بودن آزمون.....
۲۳۳	خودآزمایی (۹).....
۲۳۵	پاسخ سوال‌های خودآزمایی (۹).....
۲۶۳	<b>فصل دهم. نرم یا هنجار.....</b>
۲۶۳	۱- تعریف نرم.....
۲۶۴	۲- مشخصات مطلوب نرم‌ها.....
۲۶۴	۳- انواع نرم.....
۲۷۴	۴- نیمرخ روانی.....
۲۷۶	۵- مراحل تهیه آزمون استاندارد شده.....
۲۷۹	خودآزمایی (۱۰).....
۲۸۰	پاسخ سوال‌های خودآزمایی (۱۰).....
۲۹۱	<b>فصل یازدهم. آزمون‌های هوشی.....</b>
۲۹۱	الف) آزمون‌های فردی هوش.....
۳۰۰	ب) آزمون‌های گروهی هوش.....
۳۰۴	<b>فصل دوازدهم. آزمون‌های شخصیت.....</b>
۳۰۴	۱- پرسش‌نامه‌ها یا آزمون‌های عینی شخصیت.....
۳۰۹	۲- آزمون‌های شخصیتی فرافکن.....
۳۱۴	<b>آزمون‌های سال ۱۳۹۶-۱۴۰۰.....</b>
۳۷۹	<b>پیوست.....</b>
۳۸۶	<b>منابع و مأخذ.....</b>

## فصل ۱

### سابقه تاریخی سنجش و اندازه‌گیری و روان‌سنجی

امر استفاده از سنجش و اندازه‌گیری در زندگی روزمره انسان سابقه‌ای بس طولانی و مهم دارد. انسان‌های اولیه در مراسم گوناگون از قبیل جشن‌ها و مسابقات به منظور مقایسه توانائی و قابلیت افراد و قضاوت در مورد آنان (مثلاً قدرت بدنی، مهارت و خبرگی اشخاص) به صورتی بسیار ابتدائی و ذهنی در کلام و رفتار خود از مفاهیم اندازه‌گیری و ارزشیابی بهره‌مند می‌شدند. این اقدامات مشخص‌کننده وضع کلی اندازه‌گیری و سنجش تا اوائل قرن بیستم در ارتباط با زندگی بشر است. عده‌ای از روان‌شناسان معتقدند که ارزشیابی برای اولین بار در کشور چین حدود ۳۰۰۰ سال پیش انجام گرفته است و برخی نیز ظهور ارزشیابی را به سال ۱۶۵ پیش از میلاد نسبت می‌دهند و معتقدند که در این زمان افراد را تحت آزمایش قرار می‌دادند و به مشاغل مختلف می‌گماشتند. در قرون وسطی ارزشیابی استعدادها در بیشتر دانشگاه‌های اروپا رواج می‌یابد. کریستیان وُن وُلَف در ۱۷۳۲ رشته ریاضی روان‌شناسی را تحت عنوان روان‌سنجی دایر می‌کند. ویلهم وندت روان‌شناس آلمانی اولین کسی است که پیش از همه به روان‌سنجی علاقمند شده است. او در سال ۱۸۷۹ اولین آزمایشگاه روان‌شناسی را در شهر لایپزیک آلمان تأسیس کرد و سپس این نهضت در تمام دنیا گسترش یافت. به موازات این اقدام، روش‌های مشاهده و کنترل دقیق‌تر می‌شوند و وندت پدیده‌های مورد مطالعه را با وسایل و روش‌های دقیق کنترل می‌کند و از آن پس روان‌شناسان پژوهشگر زیادی از این آزمایشگاه روانشناسی بازدید به عمل آورده و در بازگشت این الگو را در دانشگاه‌های خود به کار گرفتند.

در سال ۱۸۸۴ فرانسیس گالتون آزمایشگاه مردم‌شناسی را که در آنجا برای بازدیدکنندگان آزمایش‌های لازم را به عمل می‌آوردند تأسیس کرد. او از نخستین کسانی بود که به مطالعه و اندازه‌گیری تفاوت‌های فردی پرداخت و عقیده داشت که بین توانایی ذهنی (هوش) و تمیز حسی رابطه وجود دارد. براین اساس بود که برای اندازه‌گیری قدرت تمیز حسی آزمون‌های مختلفی را ابداع کرد. از جمله این ابزارها می‌توان به سوت گالتون برای اندازه‌گیری قدرت شنوایی، رنگ سنج (فتومتر) برای سنجش دقت بصری آزمودنی‌ها و ... اشاره کرد.

خدمت مهم‌تر فرانسیس گالتون ابداع روش‌های آماری برای کمی ساختن نتایج حاصل از اجرای آزمون‌ها برای مطالعه تفاوت‌های فردی و تجزیه و تحلیل نتایج بود. روش آماری وی بعدها توسط کارل پیرسون توسعه پیدا کرد و به تدوین روش همبستگی گشتاوری منجر شد. جیمز کتل روان‌شناس امریکائی همانند گالتون به مطالعه تفاوت‌های فردی همت گماشت و برای نخستین بار اصطلاح «آزمون روانی» را به کار برد و این اصطلاح را وارد فرهنگ روان‌سنجی کرد. در ۱۸۹۵ تمامی آزمون‌های موجود و کارهای گالتون و کتل توسط آلفرد بینه مورد انتقاد قرار گرفت. او اظهار داشت که بین این آزمون‌ها و توانایی‌های عالی ذهن (هوش) همبستگی وجود ندارد و می‌بایست هوش براساس توانایی‌های عالی‌تر ذهنی مانند حافظه، تخیل، استعداد، توجه و ادراک مورد اندازه‌گیری قرار گیرد.

در کشف و گسترش روش‌های اندازه‌گیری در روان‌شناسی و علوم تربیتی سه عامل مهم موثر بوده است:

### ۱- پیدایش روان‌شناسی تجربی

در اوایل قرن نوزدهم فیزیولوژی تجربی، مرکز توجه برای انجام پژوهش‌های متعدد علمی در آزمایشگاه‌های تجربی کشورهای اروپائی (خصوصاً آلمان) گردید. در سال ۱۸۷۹ اولین آزمایشگاه روان‌شناسی توسط ویلهلم وندت در شهر لایپزیک آلمان تأسیس شد که مهمترین تأثیر این اقدام، قائل شدن ارزش و احترام برای روش‌های دقیق تجربی و روش‌های آماری و استفاده از آنها در حل مسائل روانی و تربیتی است.

### ۲- مطالعه تفاوت‌های فردی

در سال ۱۸۵۹ داروین اثر معروف خود را به نام «اصل انواع» منتشر ساخت و در آن اختلاف و تنوع میان افراد (تفاوت‌های فردی) را در انواع حیوانات مورد توجه قرار داد. متعاقب آن، فرانسیس گالتن تحت تأثیر آثار داروین به مطالعه توارث و خصوصیات فردی از نظر جسمانی و عقلانی پرداخت و اقدامات کارل پیرسون منجر به بوجود آمدن روش‌ها و تکنیک‌های مناسب آماری برای مطالعه تفاوت‌های فردی گردید.

## ۳- مطالعه بالینی افراد غیرعادی

نهضت علاقمندی نوع دوستانه نسبت به افراد معلول جسمی و روانی و بطور کلی کسانی که از سلامت کامل برخوردار نبودند در کشورهای مختلف بخصوص در فرانسه قوت گرفت و ضرورت تهیه وسیله‌ای یکسان و یکنواخت برای تشریح و توصیف نقائص آنان به خصوص از نظر هوش و استعداد عمومی، بیشتر احساس و آشکار شد و در نهایت منجر به تهیه مجموعه کاملی برای سنجش هوش توسط بینه و همکاران او گردید.

در یک بررسی اجمالی می‌توان فرازهای مهم تاریخچه شکل‌گیری روان‌سنجی و اندازه‌گیری را به ترتیب جدول زیر استنتاج و ارائه داد.<sup>۱</sup>

سال	اقدام انجام شده
۳۸۷ (ق.م)	افلاطون در آتن با استفاده از ارزشیابی برای رسیدن به یک «ایدال» آکادمی تأسیس کرد.
۱۶۵ (ق.م)	در چین ارزشیابی به منظور بررسی میزان صلاحیت کارگزاران در یک زمینه ملی به کار گرفته شد.
۱۶۳۶	در دانشگاه آکسفورد برای ارزشیابی آموخته‌های دانشجویان امتحان شفاهی در نظر گرفته شد.
۱۷۳۲	وُلف <sup>۲</sup> ، پیشنهاد کرد که شاخه‌ای از روان‌شناسی تحت عنوان «روان‌سنجی» دایر شود.
۱۷۹۹	ایتار <sup>۳</sup> با مطالعه کودک وحشی که در جنگل‌های جنوب فرانسه پیدا شده بود به مطالعه تفاوت بین رفتار بهنجار و نابهنجار پرداخت.
۱۸۳۴	وِبر <sup>۴</sup> تشخیص تفاوت‌های محرک‌ها را مورد بررسی قرار داد.
۱۸۳۷	سگن <sup>۵</sup> اولین موسسه اجتماعی مخصوص عقب مانده‌های ذهنی را تأسیس و در آزمایش‌های خود از صفحه سگن استفاده کرد.
۱۸۳۸	اسکیروول <sup>۶</sup> بین بیماری روانی و عقب ماندگی ذهنی تمیز قائل شد.
۱۸۶۰	فیخنر <sup>۷</sup> با کاربرد نتایج تحقیقات وِبر در ادراک در مسیر روان‌شناسی عملی گام نهاد و روان‌شناسی را با فیزیک مربوط کرد.

۱- نقل از کتاب روان‌سنجی، گنجی و ثابت

2- wolff

3- Itard

4- weber

5- SeGuin

6- Esquirol

7- Fechner

۱۸۶۹	گالتون آزمایشگاه مردم شناسی را دایر کرد و به مطالعه تفاوت‌های فردی پرداخت.
۱۸۷۹	وُنت اولین آزمایشگاه روان‌شناسی را که دریچه‌ای بر روی روان‌سنجی بود دایر کرد.
۱۸۸۳	فرانسیس گالتون به مطالعه و اندازه‌گیری تفاوت‌های فردی پرداخت، او عقیده داشت بین هوش و تمیز حسی رابطه وجود دارد.
۱۸۸۸	مک کین کتل، با دایر کردن آزمایشگاه آزمون‌ها در آمریکا به اندازه‌گیری تفاوت‌های فردی پرداخت و برای نخستین بار اصطلاح آزمون‌های روانی را به کار برد.
۱۸۹۷	ابینگه‌اوس، آزمون‌های حساب، حافظه و جملات ناقص را مطالعه کرد.
۱۹۰۴	اسپیرمن، نظریه دو عاملی هوش را ارائه داد و در ۱۹۰۵ مقاله‌ای تحت عنوان «هوش کلی» نوشت. پیرسون نظریه همبستگی را گسترش داد.
۱۹۰۵	بینه و سیمون اولین آزمون فردی هوش را ساختند. فرم تجدیدنظر دوم آن در سال ۱۹۰۸ و ویرایش سوم آن در ۱۹۱۲ منتشر شد.
۱۹۱۱	روسولیمو <sup>۱</sup> نتایج تعداد زیادی آزمون را به صورت نیمرخ نشان داد.
۱۹۱۲	اشترن <sup>۲</sup> ، مفهوم ضریب هوشی (سن عقلی تقسیم بر سن تقویمی) را پیشنهاد کرد.
۱۹۱۴	پُرتئوس <sup>۳</sup> آزمون مازها را در اتریش تهیه کرد.
۱۹۱۶	ترمن پیشنهاد می‌کند که ضریب هوشی در عدد ۱۰۰ ضرب شود او آزمون بینه را به انگلیسی ترجمه و به کمک دانشگاه استنفرد تراز کرد و بنام استنفرد بینه نامید.
۱۹۱۷	یرکس <sup>۴</sup> و همکارانش به کمک آزمون‌های گروهی آلفا و بتا تحقیقاتی انجام دادند.
۱۹۲۰	وود ورث <sup>۵</sup> اولین پرسشنامه شخصیتی را که نشانه‌های روان رنجوری را مورد سنجش قرار می‌دهد تهیه کرد.
۱۹۲۱	رُرشاخ، آزمون کلمه‌های مرکب را انتشار داد.
۱۹۲۲	شغل مشاور و راهنمایی در فرانسه بوجود آمد.
۱۹۲۳	کهِس <sup>۶</sup> آزمون مکعب‌های رنگی را ساخت.
۱۹۲۶	خانم گودیناف <sup>۷</sup> ، آزمون نقاشی آدمک را مطرح کرد.

1- Rossolimo

2- stern

3- porteus

4- yerkes

5- woodworth

6- kohs

7- goodenough

۱۹۲۷	پرسشنامه علائق شغلی استرانگ <sup>۱</sup> منتشر شد.
۱۹۳۵	مورای آزمون اندریافت موضوع را ارائه داد.
۱۹۳۶	ڈل <sup>۳</sup> مقیاس بلوغ اجتماعی را عرضه کرد.
۱۹۳۸	خانم بندر <sup>۴</sup> آزمون بینائی - حرکتی را ارائه داد. ماتریس‌های پیش رونده ریون <sup>۵</sup> ساخته شد. گزل <sup>۶</sup> ، مقیاس آزمون‌های رشد را ارائه داد.
۱۹۳۹	وکسلر، اولین آزمون هوشی بزرگسالان را ارائه داد.
۱۹۵۹	گیلفورد، نظریه ساخت ذهنی را مطرح کرد.
۱۹۶۲	هات وی و مک کین لی، پرسشنامه شخصیتی چند جنبه‌ای مینه سوتا ( <i>MMpi</i> ) را انتشار داد. آربی کتل <sup>۷</sup> ، نظریه هوش سیال و هوش متبلور را ارائه داد.
۱۹۷۲	مک کاتر، مقیاس اندازه‌گیری استعداد‌های کودکان را انتشار داد.
۱۹۸۶	ثردنایک هیگن و ساتلر، پنجمین ویرایش مقیاس هوش استنفرد بینه را انتشار دادند.
۱۹۸۹	دومین ویرایش پرسشنامه شخصیتی چند جنبه‌ای مینه سوتا ( <i>MMpi-2</i> ) تهیه شد.
۱۹۷۷	سومین ویرایش مقیاس هوش وکسلر برای بزرگسالان آماده شد.

پس از دایر شدن آزمایشگاه روان‌شناسی در سال ۱۸۷۹ به وسیله ویلهلم ونت روان‌سنجی وارد مرحله جدیدی شد، زیرا شاگردان وُندت که از تمام کشورهای دنیا بودند پس از مراجعت به وطن خود الگوی او را ترویج نمودند و در ۱۸۸۴ گالتون آزمایشگاه مردم‌شناسی را در لندن دایر کرد. بینه در سال ۱۹۰۰ کتاب مهمی تحت عنوان «دقت و انطباق»<sup>۸</sup> منتشر کرد که در آن خاصیت مهم هوش را دقت و توجه و انطباق و سازش با موقعیت تازه می‌داند که شخص با آن مواجه می‌شود. در سال ۱۹۰۴ وزارت فرهنگ فرانسه برای انتقال دانش‌آموزان عقب مانده از مدارس معمولی به مدارس خصوصی انجمنی را مأمور کرد و این انجمن تصمیم گرفت هیچ دانش‌آموزی را بدون

1- strong

2- Murray

3- Doll

4- Bender

5- Raven

6- Gesell

7- R.B. Cattell

8- Attention and Adaptation



آزمایش‌های هوشی به مدرسه مخصوص منتقل نکند و از آنجا که کار تهیه ابزار و روش سنجش هوش نیازمند یافتن راه و چاره علمی است بنابراین بینه و سیمون تلاش خود را بر آن داشتند تا جنبه‌های علمی این مسأله را حل کنند. نتیجه این تلاش در ۱۹۰۵ بصورت مقاله‌ای تحت عنوان «روش‌های نو برای شناختن اندازه هوش اطفال غیرعادی» توسط بینه به جهانیان عرضه شد. در این مقاله بینه معتقد است برای تمیز دادن دانش‌آموزان عقب مانده باید به سه روش متوسل شد. (یوسف اردبیلی، اصول علمی تهیه تست ۱۳۴۸)

۱- روش طی که شامل تحقیقات تشریحی مربوط به اندازه قد، وزن، دور سر، حواس و تنفس  
 ۲- روش تعلیمی که شامل اندازه‌گیری معلومات مدرسه‌ای و اطلاعات مکتسبه طفل در خارج از مدرسه  
 ۳- روش روان‌شناسی که عبارت از تهیه سلسله‌ای از مواد و وسائل که از لحاظ آسانی و دشواری یکسان نباشند. سوالات باید از آسان به دشوار مرتب شوند. بینه در کار هوش‌سنجی از اطفال عقب مانده به اندازه‌گیری هوش اطفال عادی و با هوش نیز رسید و در سال ۱۹۰۸ کتاب دیگری تحت عنوان «پرورش هوش بچه» منتشر ساخت.

کار پرارزش بینه را باید در آن دانست که از بررسی عناصر روان صرفنظر کرد و به آزمایش و فعل و انفعالات مدغم و پیچیده روح پرداخت. (همان منبع)

بینه تست‌های خود را در سال ۱۹۰۸ منتشر کرد و در ۱۹۱۱ بار دیگر مورد تجدیدنظر قرار داد، سلسله تست‌های بینه و سیمون در کشورهای آلمان، ایتالیا، هلند و امریکا مورد استقبال قرار گرفت. علاوه بر آزمون‌های هوش، آزمون‌های شخصیت نیز مورد توجه قرار گرفت. امیل کراپلین روان‌پزشک آلمانی «آزمون تداعی آزاد و اندیشه‌ها» را که قبلاً توسط گالتون ابداع شده بود تدوین کرده و در مورد بیماران به کار برد و بعدها کارل یونگ روش مشابهی را با عنوان «تداعی آزاد کلمات» برای تشخیص عقده‌های روانی بیماران مورد استفاده قرار داد.

به سبب شتاب‌زدگی در تدوین آزمون‌ها برای پاسخگویی به نیازهای موسسات آموزشی و تجاری بسیاری از آزمون‌های نامعتبر تهیه و انتشار یافت که نتایج حاصل از اجرای این گونه آزمون‌ها، مایوس‌کننده بود. در نتیجه بسیاری از شرکت‌های تجاری و سازمان‌های صنعتی در سال‌های میانی دهه ۱۹۲۰ استفاده از آزمون‌های روانی را کنار گذاشتند و بعدها مطالعات روان‌شناسان ژرفاندیش نشان داد که نارسائی‌های موجود در روان آزمائی به علت نارسائی در خود آزمونها است. اگر در تهیه، اجرا و تفسیر نتایج آزمون‌ها اصول و موازین علمی رعایت شود می‌توان از آنها به عنوان ابزارهای نسبتاً معتبری برای سنجش توانائی‌ها و خصائص افراد استفاده کرد. (اصول روان‌سنجی، شریفی)

## خود آزمائی (۱)

- ۱- سه عامل اساسی را که در کشف و گسترش روش‌های اندازه‌گیری در روان‌شناسی و علوم تربیتی موثر بوده‌اند توضیح دهید.
- ۲- بینه در کتاب خود تحت عنوان «دقت و انطباق» خاصیت هوش را در چه می‌داند؟
- ۳- بینه در مقاله‌ای تحت عنوان «روش‌های نو برای شناختن اندازه هوش اطفال عادی» برای تمیز دادن دانش‌آموزان عقب مانده از سه روش استفاده کرد، آنها را بیان و مختصراً توضیح دهید.
- ۴- اقدامات انجام شده در مورد سنجش خصائص شخصیت را در نیمه اول قرن بیستم شرح دهید.
- ۵- کارهای گالتون و کتل را مقایسه کرده و شباهت‌ها و تفاوت‌های آنها را بیان کنید.

## پاسخ سوال‌های خود آزمائی (۱)

- ۱- در کشف و گسترش روش‌های اندازه‌گیری در روان‌شناسی و علوم تربیتی سه عامل مهم مؤثر بوده‌اند که عبارتند از:
  - الف) پیدایش روانشناسی تجربی: تأسیس اولین آزمایشگاه روانشناسی در سال ۱۸۷۹ توسط ویلهلم ونت در شهر لایبزیگ آلمان تأثیر زیادی در زمینه پی بردن به ارزش واقعی استفاده از روش‌های دقیق تجربی و روش‌های آماری در حل مسائل روانی و تربیتی داشت و فیزیولوژی تجربی را مرکز توجه پژوهش‌های علمی در آزمایشگاه‌های تجربی قرار داد.
  - ب) مطالعه تفاوت‌های فردی: توجه داروین به موضوع تفاوت‌های فردی در انواع حیوانات در سال ۱۸۵۹، همچنین مطالعات فرانسیس گالتون در زمینه تأثیر توارث و خصوصیات فردی و بالاخره اقدامات کارل پیرسون منجر به کشف روش‌ها و تکنیک‌های مناسب آماری برای مطالعه تفاوت‌های فردی گردید.
  - ج) مطالعات بالینی افراد معلول جسمی و روانی در کشورهای مختلف به خصوص در فرانسه منجر به تهیه مجموعه کاملی برای سنجش هوش توسط بینه و همکارانش گردید.
- ۲- بینه در کتاب خود تحت عنوان «دقت و انطباق» که در سال ۱۹۰۰ منتشر کرد، خاصیت مهم هوش را دقت و توجه و انطباق و سازش باموقعیت تازه می‌دانست.
- ۳- در مقاله بینه برای تمیز دادن کودکان و دانش‌آموزان عقب‌مانده سه روش ذیل معرفی شده است.
  - الف) روش طبی که شامل مطالعات مربوط به اندازه قد، وزن، دوسر، حواس و تنفس
  - ب) روش تعلیمی که شامل اندازه‌گیری آموخته‌های کلاسیک مدرسه‌ای و غیر کلاسیک اجتماعی است.
  - ج) روش روانشناسی که عبارت است از تهیه سلسله‌ای از مواد و ابزارهای سنجش هوش برای کودکان عقب مانده، عادی و با هوش توسط بینه

۴- مهم‌ترین اقدامات انجام شده در زمینه سنجش خصائص شخصیتی در نیمه اول قرن بیستم شامل موارد زیر است:

الف) ابداع «آزمون تداعی آزاد اندیشه‌ها» توسط فرانسویس گالتون و تدوین آن توسط امیل کراپلین، روانپزشک آلمانی و استفاده از آن برای بیماران روانی  
ب) استفاده از روش «تداعی آزاد کلمات» توسط کارل یونگ برای تشخیص عقده‌های روانی بیماران  
ج) تهیه پرسشنامه مربوط به اطلاعات فردی برای تشخیص افراد نظامی روان رنجور جنگ جهانی اول توسط «رابرت وودورث» که بعدها به عنوان الگویی برای ساختن آزمون‌های شخصیت به کار گرفته شد.

۵- فرانسویس گالتون و جیمز مک‌کین کتل هر دو به مطالعه و اندازه‌گیری تفاوت‌های فردی پرداختند و معتقد بودند که با اندازه‌گیری‌های ساده تمیز حسی و زمان واکنش می‌توان توانایی‌های ذهنی را اندازه‌گیری کرد. اما خدمت مهم گالتون ابداع روش‌های آماری برای کمی ساختن نتایج حاصل از اجرای آزمون‌ها بود و کتل نیز متقابلاً برای نخستین بار اصطلاح آزمون‌های روانی را در سال ۱۸۹۰ وارد فرهنگ روان‌سنجی کرد.

**[سؤال‌های چهارگزینه‌ای فصل اول]**

- ۱- چه کسی اولین آزمون انفرادی هوش را ساخت؟  
(۱) بینه (۲) اشترن (۳) پیرسون (۴) براون (روان‌سنجی ۸۴)
- ۲- اجرای آزمون‌های خدمات اجتماعی از چه زمانی و در کجا آغاز شد؟  
(۱) انگلستان ۱۹۰۵ (۲) فرانسه ۱۹۰۵  
(۳) چین ۳۰۰۰ سال ق م (۴) امریکا بعد از جنگ جهانی دوم (روان‌سنجی ۸۴)

**[ پاسخنامه سؤال‌های چهارگزینه‌ای فصل اول ]**

۱- گزینه (۱).

۲- گزینه (۲).

## اندازه‌گیری و ارزشیابی در آموزش و پرورش

### فصل ۲

نقش اصلی معلم در فرآیند آموزش و پرورش فراهم ساختن شرایط و امکانات مناسب آموزشی و پرورشی و هدایت یادگیری دانش‌آموزان به منظور کمک به رشد و تکامل همه جانبه آنان است. معلم به همین منظور باید هر یک از دانش‌آموزان را بخوبی بشناسد.

شناخت دانش‌آموزان عبارت از کسب آگاهی‌های لازم در مورد جنبه‌های مختلف رفتار و ویژگی‌های آنان است. یعنی معلم باید از ویژگی‌های جسمانی، روانی، اجتماعی و عاطفی و همچنین علائق و توانائی‌ها و نارسائی‌های هر یک از دانش‌آموزان و رابطه این ویژگی‌ها با هم و تأثیر آنها در رفتار به خوبی آگاه باشد. شناخت دانش‌آموزان مستلزم جمع‌آوری و تحلیل اطلاعات لازم در مورد آنان براساس اصول و روش‌های علمی است. به دست آوردن اطلاعات لازم در مورد هر فرد را بر مبنای روش‌های علمی «اندازه‌گیری»<sup>۱</sup> می‌نامند.<sup>۲</sup>

اطلاعات حاصل از اندازه‌گیری ممکن است به صورت کمی (عددی) و یا به صورت توصیف (کیفی) باشد. که در صورت کمی می‌توان گفت «اندازه‌گیری عبارت است از اختصاص دادن اعداد به مقادیر مختلفی از صفات برحسب قواعد معین که صحت آنها را می‌توان از راه آزمایش بررسی کرد.<sup>۳</sup> و یا می‌توان گفت اندازه‌گیری یعنی فرآیند منظم برای کمی کردن یک متغیر پیوسته (روان‌سنجی، گنجی و ثابت) در این معنا اصطلاحاً روان‌سنجی<sup>۴</sup> نامیده می‌شود. اما باید دانست که همه ویژگی‌های آدمی را نمی‌توان به صورت کمیت‌های عددی بیان کرد،

1- Measurement

۲- نقل از کتاب اصول روان‌سنجی تألیف دکتر حسن پاشاشریفی

۳- نقل از کتاب مبانی آزمون‌های روانی، تألیف مگنوسون، ترجمه محمدتقی براهنی

4- Psychometric

چنان که تعریف بسیاری از صفات اخلاقی و معنوی به صورت کمی به آسانی میسر نیست، در چنین حالتی ویژگی مورد مطالعه باید به صورت رفتار قابل مشاهده توصیف شود مثلاً:

- وقت‌شناس و دقیق است.
- مقررات و قوانین رانندگی را رعایت می‌کند.
- از انجام کارهای عام‌المنفعه لذت می‌برد.
- اصول رفتار اجتماعی را رعایت می‌کند.

### ارزشیابی چیست<sup>۱</sup>؟

ارزشیابی عبارت است از «مقایسه اطلاعات حاصل از اندازه‌گیری با یک معیار معین و داوری در مورد آن» به عنوان مثال اگر نمره پیشرفت تحصیلی دانش‌آموزی را با میانگین نمره همکلاسانش مقایسه کنیم و نتیجه بگیریم که از نظر پیشرفت درسی از حد متوسط کلاس بالاتر است، عمل ارزشیابی انجام داده‌ایم. بنابراین می‌توان گفت که ارزشیابی عبارت است از آزمودن یک پدیده و قضاوت درباره ارزش، کیفیت، اهمیت، میزان یا درجه آن پدیده با توجه به ملاک‌های معین به منظور اقدام و تصمیم‌گیری. (اصول روان‌سنجی، شریفی)

**با توجه به مطالب مذکور می‌توان به این نتیجه رسید که:**

- ۱- اندازه‌گیری مقدمه و پیش‌نیاز ارزشیابی است.
  - ۲- ارزشیابی مستلزم یک عمل مقایسه و تقابل است.
  - ۳- ارزشیابی نیازمند یک ملاک یا معیار بیرونی و یا درونی است.
  - ۴- ارزشیابی مستلزم قضاوت علمی و داوری آگاهانه است.
- باید بدانیم که شرط مفید بودن اطلاعات حاصل از اندازه‌گیری برای ارزشیابی و شناخت دانش‌آموزان آن است که:

الف) اطلاعات با هدف ارزشیابی و شناخت کاملاً مربوط باشد. به عنوان مثال در انتخاب مناسب‌ترین افراد در مسابقات علمی، اطلاعاتی نظیر دانش و معلومات و توانایی علمی آنان به مراتب مفیدتر از اطلاعات مربوط به آمادگی‌های جسمانی و مهارت‌های بدنی آنان خواهد بود.

ب) اطلاعات به دست آمده تا حد امکان دقیق بوده و عاری از خطاهای اندازه‌گیری باشد.

### مراحل اندازه‌گیری

- ۱- **تعریف صفت یا خصوصیت مورد اندازه‌گیری:** در نخستین گام در اندازه‌گیری باید بدانیم چه چیزی را می‌سنجیم. متغیرها و خصوصیات مورد اندازه‌گیری را می‌توان به دو صورت تعریف کرد:

الف) تعریف مولفه‌ای<sup>۱</sup> (مفهومی) که متغیر با استفاده از مفاهیم دیگر تعریف می‌شود. مثلاً «هوش عبارتست از توانایی حل مسأله» و یا «اضطراب عبارت است از ترس ذهنیت یافته» در تعریف مفهومی مشخص نشده است که از شخص چه نوع رفتار قابل مشاهده سر بزند که نشانگر توانایی مورد نظر باشد و نیز چگونگی اندازه‌گیری مشخص نشده است. این نوع تعریف پایه نظری اندازه‌گیری متغیرها به شمار می‌رود.

ب) تعریف عملیاتی<sup>۲</sup>: در این معنا هر متغیر با عملیاتی که برای اندازه‌گیری و یا دستکاری آن لازم است یعنی به صورت رفتارهای قابل مشاهده و سنجش تعریف می‌شود. به نظر «کرلینجر» تعریف عملیاتی بر دو نوع است:

- سنجشی<sup>۳</sup>: در این نوع تعریف، شیوه اندازه‌گیری متغیر معلوم می‌شود. مثلاً «جرم عبارت از اقدام به عملی است که قانون آن را منع کرده است» و یا «خلاقیات عبارت است از رتبه درصدی فرد در یک آزمون میزان شده خلاقیات»

- آزمایشی<sup>۴</sup>: در این نوع تعریف، شیوه دستکاری و اعمال آن متغیر توسط پژوهشگر تعریف می‌شود. مثلاً «تقویت را می‌توان به صورت عملیاتی به تقویت کردن (پاداش دادن، ستایش کردن) و تقویت نکردن (پاداش ندادن، سرزنش کردن) به صورت رفتارهای مشخص شده تعریف کرد.

۲- تهیه ابزار یا وسیله اندازه‌گیری: برای اندازه‌گیری متغیرهایی که در علوم رفتاری مطرح می‌شوند وسیله باید طوری ساخته شود که بروز رفتارهای مورد اندازه‌گیری را به صورتی که تعریف شده‌اند موجب شود. مثلاً برای اندازه‌گیری توانایی محاسبه مساحت شکل‌های هندسی وسیله اندازه‌گیری آزمونی خواهد بود که هر یک از سوال‌های آن توانایی فرد را در یکی از زمینه‌های مساحت شکل‌های هندسی نشان دهد.

۳- تهیه مقیاس کمی: شرط اصلی در تهیه مقیاس کمی عبارت از برابری ارزش تمام واحدهای مقیاس با یکدیگر است. مقیاس‌های اندازه‌گیری برای توصیف متغیرها شامل موارد ذیل‌اند:

۱-۳- مقیاس اسمی (طبقه‌ای): در این مقیاس اشیاء را برحسب صفات مشترکشان به گروه‌های مختلف طبقه‌بندی کرده و برای هر طبقه نام و یا نشانه‌ای را قائل می‌شویم. در مقیاس طبقه‌ای می‌توان به هر یک از طبقات یک عدد (کُد) را نسبت داد. مثلاً در بررسی تعداد تصادفات یک جاده در ۴ فصل سال می‌توان فصل بهار را عدد (۱)، تابستان را عدد (۲)، پاییز و زمستان را نیز به ترتیب عدد (۳) و (۴) مشخص کرد. باید بدانیم که اعداد اسمی فاقد هرگونه ارزش عددی بوده و تابع قواعد ریاضی نیستند.

1- Constitutive

2- operational

3- Measured

4- Experimental



۲-۳- مقیاس ترتیبی: در این مقیاس بین اعداد یک رابطه منظم وجود دارد. مثلاً در درجه‌بندی پیشرفت تحصیلی دانش‌آموزان در مدارس، رتبه‌های ۱، ۲، ۳ و ... به ترتیب قوی‌ترین و ضعیف‌ترین آنها را نمایش می‌دهند. در این رتبه‌بندی نمی‌توان گفت که نفر اول چقدر از نفر دوم قوی‌تر است و نفر دوم از نفر سوم چقدر فاصله دارد. باید دانست که در این مقیاس صفر واقعی، یعنی نقطه فقدان ویژگی مورد اندازه‌گیری وجود ندارد در مثال رتبه‌بندی دانش‌آموزان برحسب پیشرفت تحصیلی، دشوار است که تصور کنیم اصلاً پیشرفتی وجود ندارد.

۳-۳- مقیاس فاصله‌ای: در این مقیاس اعداد وسعت تفاوت میان صفات افراد یا اشیاء مورد مطالعه را نشان می‌دهند. مقیاس فاصله‌ای علاوه بر اینکه همانند مقیاس رتبه‌ای دارای نظم و ترتیب منطقی است، میزان تقریبی تفاوت میان مقادیر مختلف صفت مورد اندازه‌گیری را نیز نشان می‌دهد. علاوه بر این در آن یک نقطه صفر وجود دارد که نشانه فقدان کامل یک ویژگی نیست، مثلاً صفر دماسنج به معنی عدم حرارت نیست زیرا درجه حرارت از این نقطه هم ممکن است پائین‌تر برود. بنابراین صفر در این مقیاس مطلق نبوده و به صورت قراردادی تعیین می‌شود. نمره‌های امتحانی و هوشبهر دانشجویان دارای مقیاس فاصله‌ای است. زیرا اولاً فواصل نمره‌ها با یکدیگر برابرند. یعنی می‌توان آنها را با هم مقایسه و تفاوت‌هایشان را مشخص کرد، ثانیاً نمره صفر به معنی عدم محض نیست و لذا عملیات ضرب و تقسیم را نمی‌توان در مورد آن به کار برد. اما جمع و تفریق امکان‌پذیر است. بنابراین در این مقیاس می‌توان از میانگین، انحراف معیار و ضریب همبستگی (تقریباً کلیه روش‌های آماری) استفاده کرد. بیشتر اعدادی که در علوم رفتاری به کار می‌روند از این نوعند.

۴-۳- مقیاس نسبی: این مقیاس دارای کلیه خصوصیات مقیاس فاصله‌ای است و علاوه بر آن دارای صفر مطلق نیز هست. بنابراین عملیات تجربی مورد لزوم برای داشتن چنین مقیاسی نه تنها شامل عملیات تعیین برابر، نابرابری، رتبه و فواصل با واحدهای مساوی است بلکه مستلزم برابر بودن نسبت‌ها نیز هست (عملیات ضرب، تقسیم و ...) اکثر اطلاعاتی که در نتیجه اندازه‌گیری با ابزارهای فیزیکی بدست می‌آید (از قبیل طول، سطح، حجم، وزن، زمان و ...) از این نوعند همه اعمال ریاضی را در این مقیاس می‌توان استفاده کرد.

باید بدانیم که نمره‌هایی که در آزمون‌های روانی به کار می‌روند از نوع مقیاس نسبی نیستند زیرا فاقد صفر مطلق‌اند. مثلاً نمره صفر در امتحان ریاضی به این معنی نیست که دانش‌آموز اصلاً ریاضی نمی‌داند. نمره بعضی از آزمون‌های روانی (از قبیل آزمون‌های شخصیت و رغبت) به عنوان مقیاس اسمی به کار می‌روند و نمره بعضی از آنها به صورت رتبه درصدی محاسبه می‌شود. نمره خام و نمره‌های تراز شده بیشتر آزمون‌های استاندارد شده نیز از نوع مقیاس فاصله‌ای هستند (شریفی، اصول روان‌سنجی، ۱۳۷۲)

## فلسفه ارزشیابی<sup>۱</sup>

آموزش و پرورش فرایندی است که برای تغییر و تعدیل رفتار دانش‌آموز براساس هدف‌های معین طرح‌ریزی شده است. یکی از عناصر تفکیک‌ناپذیر نظام و فرآیند آموزش و پرورش، ارزشیابی است که براساس آن اطلاعات لازم در مورد یادگیری یا آموخته‌های دانش‌آموز جمع‌آوری و تجزیه و تحلیل می‌شود تا بتوان در مورد میزان تحقق هدف‌های موردنظر داوری و قضاوت نمود.

### ارزشیابی در آموزش و پرورش بر دو دیدگاه فلسفی استوار است:

۱- **فلسفه روان‌سنجی:** طرفداران این نظریه معتقدند که هرگاه چیزی یا صفتی وجود داشته باشد، دارای اندازه معینی است که می‌توان برای آن کمیتی قائل شد و هرآنچه دارای کمیت باشد قابل اندازه‌گیری است. براین اساس وجود صفاتی مانند هوش، استعداد، معلومات و مهارت در یک کار را نمی‌توان در اشخاص منکر شد و چون سطح این خصائص در اشخاص مختلف متفاوت است، بنابراین می‌توان بر ای آنها کمیت قائل شد و آنها را اندازه‌گیری کرد. طرفداران نظریه روان‌سنجی به استفاده از آزمون‌های مختلف که نتایج آنها به صورت کمی بیان می‌شود اعتقاد دارند شاید مهمترین حسن روان‌سنجی آن است که بر روشی سازمان یافته و مشخص استوار است و اظهارنظر و قضاوتی که براین اساس انجام می‌گیرد قاطعانه‌تر است. باید بدانیم که با توجه به پیچیدگی انسان هرگز نمی‌توان درباره‌اش قاطعانه حکم کرد. بنابراین استفاده از روش چندی (کمی) در ارزشیابی همه رفتارها و خصوصیات انسان نمی‌تواند به کار رود.

۲- **فلسفه طرح و هیات کلی:** طرفداران این دیدگاه در پی آنند که تصویری جامع و توصیف‌کننده از خصائص فرد بدست آورند. آنان با استفاده از تمامی روش‌ها و امکاناتی که در مورد ویژگی‌های فرد، اطلاعات مهمی در اختیار پژوهشگر می‌گذارد و در نهایت تصویری کلی و جامع از وی فراهم می‌کند به توصیف و ارزشیابی می‌پردازند. در این روش چونی و چرائی ویژگی‌های آدمی بیش از چندی آن مورد توجه است.

در روش مبتنی بر نظریه طرح و هیات کلی به کلیت و وحدت فرد توجه می‌شود و بیشتر از روش‌هایی مانند مشاهده رفتار، داده‌های توصیفی و گزارش‌های شخصی استفاده به عمل می‌آید ولی این روش در حد روان‌سنجی، سازمان یافته و مشخص نبوده و قاطعیت ندارد.

با توجه به محاسن و محدودیت‌های مربوط به روش‌های مبتنی بر این دو نظریه، معلم و یا پژوهشگر باید سعی کند که ارزشیابی خود را به ترکیبی از این دو نظریه استوار سازد تا به گونه‌ای معتبرتر و دقیق‌تر بتواند در مورد آموخته‌ها و یا خصوصیات فرد قضاوت کند.

۱- نقل از کتاب شیوه‌های ارزشیابی از آموخته‌های دانش‌آموزان، تألیف دکتر حسن پاشا شریفی.

## دسته‌بندی ارزشیابی‌های آموزشی

صاحب‌نظران ارزشیابی آموزشی در دیدگاه‌هایشان نسبت به اینکه ارزشیابی چیست و چگونه باید آن را انجام داد اختلاف‌نظر دارند. این اختلاف‌نظرها منجر به رویکرد ارزشیابی آموزشی مختلفی شده است. این صاحب‌نظران ارزشیابی تمامی رویکردهای ارزشیابی آموزشی را در شش دسته به شرح زیر تقسیم کرده‌اند.

- ۱- رویکرد مبتنی بر هدف: در رویکرد مبتنی بر هدف، هدف‌های کلی و هدف‌های دقیق مورد تأکید قرار می‌گیرند و ارزشیابی آموزشی به دنبال این است که تعیین کند این هدف‌ها به چه میزان تحقق یافته‌اند.
  - ۲- رویکرد مبتنی بر مدیریت: هدف تشخیص و تحقق بخشیدن به نیازهای اطلاعاتی مدیران تصمیم‌گیرنده است.
  - ۳- رویکرد مبتنی بر مصرف‌کننده: برکسب اطلاعات مربوط به نتایج کلی آموزشی تأکید دارد که توسط مصرف‌کنندگان آموزشی از میان برنامه‌های درسی مختلف انتخاب و مورد استفاده قرار می‌گیرد.
  - ۴- رویکرد مبتنی بر نظر متخصصان: در این رویکرد تأکید اصلی بر کاربرد مستقیم نظر متخصصان در قضاوت درباره کیفیت فعالیت‌های آموزش و پرورش است.
  - ۵- رویکرد مبتنی بر اختلاف نظر متخصصان: در این رویکرد اختلاف‌نظرهای ارزشیابان مختلف (موافق و مخالف) مورد تأکید است.
  - ۶- رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان: در این رویکرد بررسی طبیعت‌گرایانه و کوشش مشارکت‌کنندگان در تعیین ارزش‌ها، ملاک‌ها، نیازها و داده‌های ارزشیابی مورد تأکید است.
- این شش رویکرد ورتن و سندرز بر تقسیم‌بندی دوگانه هاوس از ارزشیابی آموزشی استوار است. هاوس انواع ارزشیابی‌های آموزشی را به دو دسته تقسیم کرد. عین‌گرا در مقابل ذهن‌گرا یا فایده‌گرا در مقابل کثرت‌گرا-شهودگرا. در حقیقت این نوع تقسیم‌بندی پیوستاری می‌سازد که در یک کرانه آن عین‌گرا یا فایده‌گرا و در کرانه دیگر آن ذهن‌گرا یا شهودگرا - کثرت‌گرا قرار می‌گیرد. رویکرد مبتنی بر هدف و رویکرد مبتنی بر مدیریت با کرانه عین‌گرایی یا فایده‌گرایی منطبق هستند و رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان بر کرانه ذهن‌گرایی یا شهودگرایی - کثرت‌گرایی انطباق دارد و بقیه طبقات در حد فاصل این دو طبقه واقع هستند. در ارزشیابی فایده‌گرا، بیشترین خوبی آن است که به بیشترین افراد فایده برساند. فایده‌گرایی به این فکرند که شادمانی را در جامعه باید به حداکثر رساند. بنابراین در ارزش‌یابی‌های وابسته به رویکرد فایده‌گرایی

یا عینیت‌گرایی، متخصص ارزشیابی، با استفاده از میانگین نمرات آزمون یا شاخص‌های دیگر «خوبی» بر دستاوردهای کلی گروه تأکید می‌ورزند تا حداکثر خوبی را برای حداکثر افراد تعیین نمایند. در مقابل عینیت‌گرایی یا فایده‌گرایی، شهودگرایی یا کثرت‌گرایی قرار دارد. رویکرد شهودگرایی - کثرت‌گرایی به این فکرند که ارزش به تأثیر برنامه بر فرد فرد افراد وابسته است نه بر اکثریت آنان. در این رویکرد، احساس‌ها و دریافتهای فردی ملاک خوبی و درستی به حساب می‌آیند. به همین سبب از لحاظ فلسفی به این رویکرد شهودگرایی - کثرت‌گرایی، نام ذهن‌گرایی داده‌اند. بنابراین در رویکرد ذهن‌گرایی (کثرت‌گرایی) تأکید بر توزیع دستاوردها در میان همه افراد و خرده‌گروه‌ها مانند گروه‌های قومی است. و در رویکرد ذهن‌گرایی (شهودگرایی) متعادل ساختن و وزن دادن به ارزشیابی‌ها و ملاکهای گوناگون در این رویکرد عمدتاً جنبه شهودی دارد.

دسته‌بندی ارزشیابی‌های آموزشی در قالب الگوهای مختلف: حال که با رویکرد مختلف ارزشیابی آموزشی آشنا شدیم می‌خواهیم مدل‌ها یا الگوهای ارزشیابی را مورد بحث قرار دهیم. **الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر هدف:** در رویکرد ارزشیابی مبتنی بر هدف، ابتدا اهداف و مقاصد یک فعالیت آموزشی تعیین می‌شود و بعد از آن در ارزشیابی کوشش به عمل می‌آید تا نشان دهد که آن اهداف و مقاصد چه اندازه تحقق یافته‌اند. در آموزش و پرورش فعالیت ارزشیابی می‌تواند از یک روز آموزشی کلاسی تا تمامی نظام آموزشی را شامل شود. نتایج بدست آمده از ارزشیابی مبتنی بر هدف را می‌توان برای اصلاح و تجدید نظر هدف‌های فعالیت آموزشی، خود فعالیت آموزشی، روش‌ها و تدابیر به کار رفته برای تحقق اهداف به کار برد.

**الگوی ارزشیابی تایلر:** اولین الگوی ارزشیابی آموزشی الگوی ارزشیابی تایلر است. مفهوم ارزشیابی آموزشی تایلر این است که تعیین کند هدف‌های آموزشی برنامه درسی تا چه میزانی تحقق یافته‌اند. تایلر در الگوی ارزشیابی مراحل زیر را پیشنهاد داده است. ۱- تعیین هدف‌های کلی (غایت‌ها) و هدف‌های دقیق ۲- طبقه‌بندی هدف‌ها ۳- بیان هدف‌ها به صورت رفتاری ۴- یافتن موقعیتهایی که در آنها بتوان دستیابی به هدف‌ها را نشان داد ۵- تولید یا تهیه روش‌ها و فنون اندازه‌گیری ۶- گردآوری داده‌های مربوط به عملکرد یادگیرندگان ۷- مقایسه داده‌های مربوط به عملکرد یا هدف‌های رفتاری. الگوی ارزشیابی تایلر، از لحاظ منطقی درست، از نظر علمی قابل پذیرش و توسط متخصصان ارزشیابی به آسانی قابل استفاده است. و حوزه برنامه درسی را شدیداً تحت تاثیر قرار داده است.

رویکردهای دیگر ارزشیابی مبتنی بر هدف عبارتند از: پارادیم ارزشیابی متفلسل و مایکل، رویکرد ارزشیابی هاموند و الگوی ارزشیابی اختلاف پرووس. تبا یکی از همکاران نزدیک به تایلر، برای برنامه‌ریزی درسی یک الگوی مبتنی بر هدف شامل مراحل زیر را پیشنهاد کرده است:

۱- تعیین نیازها ۲- بیان هدف‌ها ۳- انتخاب محتوا ۴- سازمان دادن به محتوا ۵- انتخاب تجارب یادگیری ۶- سازمان دادن به تجارب یادگیری ۷- تعیین «چه چیزی» و «چگونگی» ارزشیابی.

**الگوی ارزشیابی برخاسته از رویکرد مبتنی بر مدیریت:** دسته‌ای از الگوهای ارزشیابی که از رویکرد مبتنی بر مدیریت برخاسته‌اند به قصد کمک به مدیران و تصمیم‌گیرندگان به وجود آمده‌اند. پدیدآورندگان این الگوها بر یک رویکرد سیستمی (نظام‌دار) تأکید دارند که در آن، در مورد درون‌داده‌ها، فرایندها و برون‌داده‌ها تصمیم‌گیری می‌شود، مخاطبین این الگوها تصمیم‌گیرندگان آموزش و پرورش هستند و جهت فعالیت ارزشیابی را خواسته‌های تصمیم‌گیرندگان، نیازهای اطلاعاتی، و ملاک‌های اثربخشی تعیین می‌کند.

الگوی ارزشیابی سیپ و الگوی ارزشیابی یوسی‌ال‌ای، جزو الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر مدیریت هستند. الگوی ارزشیابی سیپ توسط دانیل ال. استافل‌بیم و همکاران او پدید آمد. عنوان سیپ (*CIPP*) از حروف اول کلمات بافت یا موقعیت (*Context*)، درون داد (- *Input*)، فرایند (*Process*)، و فرآورده (*Product*) به دست آمده است، زیرا این الگوی ارزشیابی دارای چارچوبی است که مدیران و تصمیم‌گیرندگان را در چهار نوع تصمیم‌گیری زیر کمک می‌کند:

۱) **ارزشیابی از موقعیت یا بافت:** از نظر استافل‌بیم و همکارانش، این مهمترین نوع ارزشیابی است. هدف این ارزشیابی فراهم آوردن یک زمینه منطقی برای تعیین هدف‌های آموزشی است. این مرحله از ارزشیابی همچنین شامل کوشش‌هایی تحلیل‌گرایانه برای تعیین عناصر مربوط در محیط آموزشی و نیز کوشش در جهت شناسایی مشکلات، نیازها، و فرصت‌های موجود در این بافت یا موقعیت آموزشی است. به طور خلاصه، این مرحله از ارزشیابی به تصمیم‌گیری در زمینه طراحی کمک می‌کند. روش‌های ارزشیابی از بافت عمدتاً توصیفی و مقایسه‌ای هستند.

۲) **ارزشیابی از درون‌داد:** در این مرحله از ارزشیابی اطلاعات مورد نیاز درباره چگونگی استفاده از منابع به منظور دستیابی به هدف‌های برنامه جمع‌آوری می‌شوند. در ضمن ارزشیابی از درون‌داد، ماهیت توانایی‌های نظام آموزشی و استراتژی‌های بالقوه برای رسیدن به هدف‌هایی که در نتیجه ارزشیابی از بافت مشخص گردیدند تعیین می‌شود. همچنین در این مرحله از ارزشیابی به تصمیم‌گیرندگان کمک می‌شود تا روش‌هایی را انتخاب و طراحی کنند که برای دستیابی به هدف‌های ضروری تشخیص داده شده‌اند.

۳) **ارزشیابی از فرایند:** در این ارزشیابی کوشش می‌شود تا پاسخ پرسش‌هایی نظیر اینها تعیین گردند: آیا برنامه به خوبی در حال پیاده شدن است؟ چه موانعی بر سر راه موفقیت آن قرار

دارند؟ چه تغییراتی ضروری هستند؟ پاسخ به این پرسش‌ها به کنترل و هدایت شیوه‌های اجرایی کمک می‌کند.

۴) ارزشیابی از فرآورده: در این مرحله مشخص می‌شود که نتایج بدست آمده چه هستند. بعد نتایج حاصل با هدف‌های برنامه مقایسه می‌گردند و رابطه بین انتظارات و نتایج واقعی مشخص می‌شوند. پس از بررسی کامل نتایج اطلاعات لازم در اختیار تصمیم‌گیرندگان گذاشته می‌شود تا در صورت لزوم درباره ادامه برنامه، متوقف ساختن آن، یا تجدیدنظر در آن اتخاذ تصمیم کنند.

**الگوی ارزشیابی یو سی ال ای.** در این الگو، ارزشیابی عبارت است از فرایند تعیین انواع تصمیماتی که باید اتخاذ شود و نیز انتخاب، گردآوری و تحلیل اطلاعات مورد نیاز برای اخذ این تصمیمات؛ و گزارش کردن این اطلاعات به تصمیم‌گیرندگان.

الگوی ارزشیابی یو سی ال ای از پنج نوع ارزشیابی زیر تشکیل شده است:

۱) **سنجش نظام:** در این مرحله از ارزشیابی اطلاعات لازم را درباره حالت یا وضعیت نظام آموزشی گردآوری می‌کند. بسیار شبیه مرحله ارزشیابی بافت در الگوی سیپ است.

۲) **طراحی برنامه:** در این مرحله از ارزشیابی برنامه‌های ویژه‌ای انتخاب می‌شود که برای نیازهای آموزشی مفید هستند.

۳) **اجرای برنامه:** در این مرحله از ارزشیابی، اطلاعات ضروری درباره‌ای که برنامه انتخابی به گونه‌ای که در نظر بوده و به گروهی که می‌بایست داده شود معرفی گردیده است یا نه جمع‌آوری می‌گردد.

۴) **بهبودی برنامه:** در این مرحله از ارزشیابی، جمع‌آوری اطلاعات است درباره چگونگی کارکرد برنامه، چگونگی تحقق هدف‌های واسطه‌ای و بازده‌های پیش‌بینی‌نشده‌ای ظاهر شده‌اند یا نه. این مرحله شبیه به ارزشیابی فرایندی الگوی سیپ است.

۵) **تصدیق برنامه:** در این مرحله اطلاعات مورد نیاز برای قضاوت درباره ارزش برنامه و امکان استفاده از آن در موقعیت‌های دیگر فراهم می‌آیند.

**رویکرد مبتنی بر مصرف‌کننده:** این رویکرد عمدتاً برای کمک به مصرف‌کنندگان در انتخاب بهترین محصولات موجود در بازار فراهم آمده‌اند. در این رویکرد ارزشیابی برابر است با تعیین میزان مطلوب بودن یک فرآورده آموزشی.

الگوهای مهم رویکرد مبتنی بر مصرف‌کننده عبارتند از: الگوی ارزشیابی اسکرین و الگوی ارزشیابی هدف‌آزاد که هر دو توسط اسکرین پیشنهاد شده است.

**الگوی ارزشیابی اسکریون:** اسکریون معروف‌ترین شخصیتی است که در زمینه ارزشیابی آموزشی به قصد کمک به مصرف‌کنندگان کار کرده است. از اندیشه‌های مهم وی ایجاد تمایز بین ارزشیابی تکوینی و تراکمی است. ارزشیابی تکوینی در حین اجرای برنامه و زمانی که هنوز امکان اصلاح مشکلات و رفع نواقص موجود است به کار می‌رود، یعنی زمانی که برنامه در حال شکل‌گیری یا تکوین است؛ اما ارزشیابی تراکمی در پایان کار و صرفاً برای روشن شدن اینکه تا چه اندازه به اهداف مورد نظر رسیدیم انجام می‌شود. الگوی اسکریون که برخاسته از رویکرد مبتنی بر مصرف‌کننده است عمدتاً یک ارزشیابی تراکمی است.

ملاک‌های اسکریون برای ارزشیابی فراورده‌ها عبارتند از: ۱- شواهد حاکی از محقق شدن اهداف مهم آموزشی ۲- شواهد حاکی از تحقق یافتن اهداف مهم غیرآموزشی (اهداف اجتماعی) ۳- نتایج حاصل از پی‌گیری ۴- آثار جانبی پیش‌بینی نشده، مثل تأثیر بر معلم، همکاران معلم، سایر دانش‌آموزان، مدیران، والدین، مدرسه و غیره ۵- دامنه استفاده (یعنی برای چه کسانی مفید است) ۷- ملاحظات اخلاقی (استفاده نابجا از تنبیه یا محتوای بحث‌انگیز) ۶- هزینه‌ها

**الگوی ارزشیابی هدف آزاد:** یکی از پیشنهاد‌های اسکریون، الگوی ارزشیابی هدف‌آزاد در مقابل ارزشیابی مبتنی بر هدف است. در ارزشیابی مبتنی بر هدف، متخصص ارزشیابی هم مترصد کیفیت هدف‌های برنامه آموزشی و هم علاقه‌مند به این است که بدان تا چه میزانی آن هدف‌ها تحقق یافته‌اند. در الگوی ارزشیابی هدف‌آزاد پیشنهادی اسکریون، علاوه بر اینها، بر بازده‌های پیش‌بینی نشده برنامه آموزشی نیز تأکید می‌شود. اسکریون معتقد است که نقش عمده ارزشیابی هدف آزاد کاستن از سوگیری و افزایش عینیت است. با این حال، اسکریون ارزشیابی هدف‌آزاد را به عنوان جانشین برای ارزشیابی مبتنی بر هدف پیشنهاد نداده است، بلکه آن را مکمل ارزشیابی مبتنی بر هدف می‌داند.

**الگوی ارزشیابی برخاسته از رویکرد مبتنی بر متخصصان:** در این رویکرد، اثربخشی مؤسسات، برنامه‌ها، فراورده‌ها، و فعالیت‌های آموزشی توسط متخصصان و صاحبان فن صورت می‌پذیرد. ورتن و سندرز الگوهای ارزشیابی برخاسته از نظر متخصصان را در چهار طبقه دسته‌بندی کرده‌اند: ۱- نظام‌های بررسی تخصصی رسمی (۲) نظام‌های بررسی تخصصی غیررسمی (۳) بررسی توسط جلسات میزگرد ویژه (۴) بررسی توسط افراد خبره. موقعیت‌هایی که در آنها ارزشیابی مبتنی بر نظر متخصصان به کار می‌رود شامل امتحانات شفاهی دوره دکتری، ب بررسی پرونده‌های استخدامی و انجام مصاحبه‌های استخدامی، بررسی پرونده علمی استادان دانشگاه برای صدور احکام قطعی استخدامی و ارتقاء، بررسی مقالات رسیده برای چاپ در مجله‌های علمی توسط هیأت دبیران مجله. **الگوهای ارزشیابی برخاسته از رویکرد مبتنی بر اختلاف نظرهای متخصصان:** در این الگو فرض بر این است که تمام متخصصان ارزشیابی خواسته یا ناخواسته نظر‌ها و تعصبات خود را در جریان

ارزشیابی دخالت می‌دهند، لذا برای اینکه نتایج درست‌تری از ارزشیابی به دست بیاوریم باید از نظرهای موافق و مخالف چند تن از متخصصان و صاحب‌نظران استفاده کنیم. این رویکرد امیدوار است با دخالت دادن نظرهای مثبت و منفی در ارزشیابی سوگیریها را متعادل سازد و از این طریق عدالت را برقرار نماید.

**الگوهای ارزشیابی برخاسته از رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کنندگان:** این رویکرد بر تجارب دست اول در موقعیت‌ها و فعالیت‌های تربیتی و آموزشی تأکید دارد. این رویکرد، روش‌های پژوهش و بررسی مبتنی بر مشاهدات طبیعی را به روش‌های دقیق و کنترل‌شده آزمایشی ترجیح می‌دهند و از این رو به رویکرد آنها نام طبیعت‌گرایانه داده شده است. همچنین به جهت آنکه بر دخالت کسانی که در موضوع مورد ارزشیابی شرکت دارند نیز تأکید می‌ورزند نام رویکرد مبتنی بر مشارکت‌کنندگان داده‌اند. یکی از معروف‌ترین الگوهای ارزشیابی مبتنی بر این رویکرد، رویکرد رابرت ای. استیک به نام الگوی سیمایی است. در این الگوی ارزشیابی بر دو عامل اساسی ارزشیابی یعنی توصیف، و داوری تأکید می‌شود و اینها دو سیما یا دو روی ارزشیابی به حساب می‌آیند، و بدین لحاظ به الگوی ارزشیابی استیک نام سیمایی داده‌اند این دو عمل توصیف و داوری در سه مرحله از یک برنامه آموزشی، یعنی پیش‌آیندها، مبادله‌ها، و بازده‌ها یا پی‌آیندها انجام می‌گیرد. علاوه بر الگوی ارزشیابی استیک، چند الگوی معروف دیگر وابسته به رویکرد طبیعت‌گرایانه و مبتنی بر مشارکت‌کننده وجود دارند که عبارتند از ارزشیابی روشنگر، ارزشیابی دموکراتیک، ارزشیابی پاسخگو، و ارزشیابی طبیعت‌گرایانه.

### دسته‌بندی ارزشیابی‌های آموزشی با توجه به موضوع ارزشیابی

با توجه به تعریف ارزشیابی، آیا هدف‌های مورد نظر تحقق یافته‌اند یا نه؟ بنابراین یکی از راه‌های تقسیم‌بندی انواع ارزشیابی‌ها با توجه به هدف یا موضوع ارزشیابی است و از این لحاظ ارزشیابی‌های مختلف آموزشی را می‌توان در گروه‌های مختلف زیر قرار داد.

۱- **ارزشیابی از دانشجویان و دانش‌آموزان:** عمدتاً ارزشیابی پیشرفت تحصیلی یا ارزشیابی از میزان یادگیری است. استفاده از نتایج آن نه تنها برای قضاوت درباره عملکرد یادگیرندگان، بلکه برای داوری درباره اثربخشی کیفیت آموزش، روش‌ها و مواد آموزشی نیز مفید هستند. عمده‌ترین ملاک قضاوت در ارزشیابی پیشرفت تحصیلی دانش‌آموزان و دانشجویان هدف‌های آموزشی از پیش تعیین شده است. ارزشیابی پیشرفت تحصیلی تعیین می‌کند که دانشجویان تا چه میزانی به هدف‌های آموزشی از پیش تعیین شده رسیده‌اند.

۲- **ارزشیابی از برنامه‌های درسی و مواد آموزشی:** ارزشیابی آموزشی گاهی به منظور تعیین اثربخشی برنامه درسی یا مواد آموزشی انجام می‌شود و شامل ارزشیابی از عواملی چون



روش‌های آموزشی، کتاب‌های درسی، مواد دیداری - شنیداری و تدارکات فیزیکی و سازمانی است. در ارزشیابی از برنامه درسی نیز عمده‌ترین ملاک پیشرفت تحصیلی میزان یادگیری دانشجویان است. ملاک دیگر ارزشیابی برنامه درسی هزینه آنهاست.

۳- ارزشیابی از پروژه‌ها و برنامه‌های آموزشی: منظور از پروژه‌ها و برنامه‌های آموزشی ویژه کوشش‌هایی هستند موقعیتی که به عنوان بخشی از برنامه‌های معمول آموزشگاه‌ها به حساب نمی‌آیند. آنها نوعی ماهیت نوآورانه دارند و طول مدت حضورشان به موفقیتشان وابسته است. مفاهیم پروژه، برنامه، و برنامه درسی از هم متمایزند. پین آنها را به این گونه از هم تمیز داده است. پروژه یک کوشش مجزا و یکباره که با استفاده از یک نوآوری می‌کوشد تا تغییری ایجاد کند. این نوآوری به صورت: الف) آموزش مهارت‌های آزمایشگاهی ب) بهبود نگرش دانش آموزان روستا نسبت به یادگیری مربوط باشد. تفاوت پروژه و برنامه: الف) برنامه از پروژه گسترده‌تر و پیچیده‌تر است. ب) پروژه ممکن است نوآوری جهت رفع مشکلات ابتدایی، اما برنامه رویکرد چندوجهی برای مشکلات کلی آموزش و پرورش است ج) برنامه یک مسئله را پاسخ می‌دهد ولی پروژه به یک منظور خاص مربوط است. برنامه درسی از پروژه و برنامه جامع‌تر است. پین برنامه درسی را به عنوان مجموعه‌ای از برنامه‌ها تعریف کرده است و مواردی چون مطالعات اجتماعی، خواندن علوم و ریاضیات را به عنوان برنامه درسی ذکر کرده که برنامه درسی دارای دامنه و توالی است.

۴- ارزشیابی از آموزشگاه: منظور از ارزشیابی آموزشگاه این است که تعیین شود هدف‌های پیش‌بینی‌شده برای آموزشگاه به چه میزانی تحقق یافته‌اند و نقاط قوت و ضعف کدام‌اند. ارزشیابی از آموزشگاه دربرگیرنده ارزشیابی از کل برنامه آموزشگاه و مستلزم گردآوری اطلاعات لازم درباره همه جنبه‌های آموزشگاه است. یکی از جنبه‌های مهم ارزشیابی از آموزشگاه برنامه آزمون آموزشگاه است. هر چه این برنامه جامع‌تر و کامل‌تر باشد اطلاعات بدست آمده نیز به همان نسبت قابل اعتمادتر و باارزش‌تر خواهند بود.

۵- ارزشیابی از کارکنان آموزشگاه: مقصود از ارزشیابی از کارکنان آموزشگاه، ارزشیابی از همه کسانی است که بطور مستقیم یا غیرمستقیم مسئول بازدهی آموزش مؤسسه هستند. شاید پیچیده‌ترین نوع ارزشیابی همین ارزشیابی باشد. علت پیچیدگی این روش ارزشیابی کم‌اعتباری و بی‌دقتی وسایل و روش‌های اندازه‌گیری و سنجش مورد استفاده در این نوع ارزشیابی است.

## خودآزمائی (۲)

- ۱- چه رابطه‌ای بین اندازه‌گیری و ارزشیابی وجود دارد با ذکر یک مثال توضیح دهید.
- ۲- پژوهشگری می‌خواهد میزان احساس مسوولیت کارکنان یک موسسه را بسنجد، او برای این منظور چه اقداماتی را باید انجام دهد؟
- ۳- معلم در ارزشیابی میزان پیشرفت درسی شاگردان خود ممکن است مرتکب خطای اندازه‌گیری شود، این خطا به چه صورت‌هایی در تصمیمات و اقدامات معلم در ارتباط با فعالیتهای آموزشی احتمالاً قابل مشاهده است؟
- ۴- ارزشیابی نیازمند یک ملاک است، معلم برای ارزشیابی نمره امتحانی دانش‌آموزان از چه ملاک‌هایی می‌تواند استفاده کند.
- ۵- تفاوت ملاک‌های تجربی و قراردادی چیست؟ با ذکر یک مثال توضیح دهید.
- ۶- پژوهشگری میزان محبوبیت گروهی هر آزمودنی را به صورت درصد افرادی که در گروه سنجی به او اظهار تمایل و علاقه کرده‌اند تعریف نموده است. او در تعریف متغیر «محبوبیت گروهی» از چه نوع تعریفی استفاده کرده است؟ چرا؟
- ۷- معلمی نمره امتحانی درس آمار دو دانش‌آموز را در یک آزمون مشترک ۲۰ و ۱۰ محاسبه و نتیجه‌گیری کرده است که معلومات اولی در زمینه آمار دو برابر دومی است، آیا این نتیجه‌گیری درست است؟ چرا؟
- ۸- نوع مقیاس اندازه‌گیری را در هر یک از موارد زیر مشخص کنید:
  - طبقه‌بندی دانشجویان برحسب تأهل
  - تعیین تاریخ تولد افراد
  - محاسبه نمره امتحانی هر دانش‌آموز
  - اندازه‌گیری قد کودکان ۶ ساله
  - تعیین موقعیت و جایگاه هر دانشجو در بین همکلاسان
  - تعیین هوشبهر هر کودک
  - بررسی نظرات دانشجویان نسبت به تدریس استاد در یک مقیاس ۵ درجه‌ای (خیلی خوب، خوب، .....، خیلی ضعیف)
- ۹- به نظر شما کدام یک از این دو نوع ارزشیابی دقیق‌تر و علمی‌تر است؟ چرا؟
  - الف) مقایسه نمره دانش‌آموز با میانگین کلاس
  - ب) مقایسه نمره دانش‌آموزان با ملاک قبولی نمره ۱۰

- ۱۰- نتایج حاصل از اجرای آزمون‌های مربوط به علوم رفتاری غالباً از کدام نوع مقیاس اندازه‌گیری برخوردارند؟
- ۱۱- اکثر اطلاعاتی که در نتیجه‌گیری با ابزارهای فیزیکی به دست می‌آیند از نوع مقیاس ..... به حساب می‌آیند.
- ۱۲- ارزشیابی چه ارتباطی با سایر عناصر نظام آموزش و پرورش دارد؟
- ۱۳- دو مورد از تفاوت‌های فلسفه ارزشیابی در دیدگاه روان‌سنجی و دیدگاه طرح و هیأت کلی را بیان کنید.
- ۱۴- طرفداران نظریه روان‌سنجی استفاده از چه روش و ابزاری را برای ارزشیابی خصوصیات و متغیرهای مورد نظر توصیه می‌کنند؟ طرفداران نظریه طرح کلی در این مورد چه نظری دارند؟
- ۱۵- به طور کلی استفاده از روش‌های مربوط به کدام یک از دو دیدگاه (روان‌سنجی و هیأت کلی) را برای ارزشیابی توصیه می‌کنید؟ چرا؟
- ۱۶- گزاره «ناکامی عبارت است از جلوگیری در دستیابی فرد به هدف مورد نظرش» چه نوع تعریفی در مورد متغیر ناکامی به حساب می‌آید؟

## پاسخ خودآزمایی (۲)

- ۱- اندازه‌گیری مقدمه ارزشیابی است. به بیان دیگر ارزشیابی مستلزم عمل اندازه‌گیری است. به عنوان مثال برای ارزشیابی میزان پیشرفت تحصیلی یک دانش‌آموز نخست باید پیشرفت تحصیلی او را اندازه گرفت و سپس نتیجه اندازه‌گیری را از طریق مقایسه با یک معیار ارزشیابی نمود.
- ۲- برای ارزشیابی از میزان احساس مسئولیت کارکنان یک مؤسسه پژوهشگر باید با توجه به مراحل اندازه‌گیری و ارزشیابی اقدامات زیر را انجام دهد
- الف) تعریف متغیر «احساس مسئولیت» به صورت رفتار قابل مشاهده و سنجش (تعریف عملیاتی)
- ب) تهیه و یا پیش‌بینی یک ابزار و وسیله‌ای برای اندازه‌گیری احساس مسئولیت. این ابزار ممکن است به صورت یک پرسشنامه، مقیاس درجه‌بندی و یا یک آزمون باشد.
- ج) تهیه مقیاس کمی برای ابزار: یعنی ابزار طوری درجه‌بندی (بارم‌بندی) شود که بتواند متغیر احساس مسئولیت را برای هر فرد به صورت کمی در یکی از مقیاس‌های اندازه‌گیری (اسمی، ترتیبی، فاصله‌ای و یا نسبی) مشخص کند.
- د) پیش‌بینی یک معیار: نتیجه اندازه‌گیری باید با یک معیار (تجربی و یا قراردادی) مقایسه شود این معیار ممکن است براساس شواهد تجربی به دست آمده باشد (مثلاً میانگین میزان احساس مسئولیت مربوط به گروه نرم) و یا به صورت قراردادی و ذهنی تعیین شده باشد (مثلاً تعیین حد نصاب قابل قبول)

- ه) مقایسه و داوری: در آخرین مرحله نتیجه مقایسه توسط پژوهشگر مورد قضاوت و داوری قرار می‌گیرد.
- ۳- خطای اندازه‌گیری توسط معلم ممکن است به صورت مختلف در مراحل دوره آموزش انجام گیرد که عبارتند از:
- خطای اندازه‌گیری مربوط به تهیه و تدوین هدف‌های تدریس (هدف‌های کلی، جزئی و رفتاری)
  - خطای اندازه‌گیری مربوط به استفاده از برنامه‌های آموزشی نارسا، و کم ارزش
  - خطای اندازه‌گیری مربوط به استفاده از شیوه‌های نامناسب و سنتی تدریس
  - خطای اندازه‌گیری مربوط به استفاده از مواد و وسایل آموزشی و تجهیزات ناکافی
  - خطای اندازه‌گیری مربوط به نظام ارزشیابی (تهیه سوالات امتحانی نامناسب، اجرای نامطلوب امتحان، تصحیح و نمره‌گذاری اوراق امتحانی به صورت غیرعلمی و سوگیری و تعصب در داوری و قضاوت روی عملکرد دانش‌آموزان)
- ۴- معلم در ارزشیابی از میزان پیشرفت تحصیلی دانش‌آموزان می‌تواند ملاک ارزشیابی را به صورت‌های ذیل در نظر بگیرد:
- معلم پس از اندازه‌گیری میزان پیشرفت، نمره خام دانش‌آموز را با یک ملاک قراردادی (مثلاً نمره قبولی ۱۰) مقایسه می‌کند.
  - معلم نمره خام دانش‌آموز را با نمره سایر دانش‌آموزان همان کلاس (مثلاً میانگین نمره دانش‌آموزان) مقایسه می‌کند.
  - معلم نمره خام دانش‌آموز را با نمره‌های قبلی وی در همان درس مورد مقایسه قرار می‌دهد. یعنی ملاک مقایسه را خود دانش‌آموز در نظر می‌گیرد.
- ۵- پاسخ این پرسش قبلاً توضیح داده شده است.
- ۶- پژوهشگر برای متغیر محبوبیت گروهی از تعریف عملیاتی سنجشی استفاده کرده است. زیرا محبوبیت گروهی یک متغیر خصیصه‌ای (ویژگی) است که در تعریف، نحوه اندازه‌گیری آن بیان شده است.
- ۷- خیر نتیجه‌گیری درست نیست زیرا نمره‌های درسی دارای مقیاس اندازه‌گیری فاصله‌ای هستند و بنابراین صفر مطلق ندارند و عملیات ضرب و تقسیم در مورد آنها انجام‌پذیر نیست.
- ۸- مقیاس اندازه‌گیری ۷ مورد بیان شده به ترتیب عبارتند از:
- اسمی (طبقه‌ای) - فاصله‌ای - فاصله‌ای - نسبی - ترتیبی - فاصله‌ای - طبقه‌ای مرتب شده
- ۹- مورد «الف» دقیق‌تر است زیرا دارای ملاک و معیار تجربی است ولی مورد «ب» دارای ملاک قراردادی است.

- ۱۰- مقیاس‌های فاصله‌ای و ترتیبی
- ۱۱- مقیاس نسبی
- ۱۲- ارزشیابی نقش ناظر را بر سایر عناصر نظام آموزش و پرورش (هدف، برنامه، روش، وسایل و امکانات و تجهیزات آموزشی) دارد.
- ۱۳- دیدگاه روان‌سنجی مبتنی بر چندی (کمی) است ولی دیدگاه طرح کلی مبتنی بر چونی (کیفی) است. روان‌سنجی بر روشی سازمان یافته و تفکیک و تجزیه‌پذیر استوار است ولی طرح کلی به کلیت و وحدت فرد توجه می‌کند و قاطع و سازمان‌یافته نیست.
- ۱۴- طرفداران نظریه روان‌سنجی به استفاده از آزمون‌های مختلف که نتایج آنها به صورت کمی است توجه دارند ولی طرفداران نظریه طرح کلی بیشتر از روش‌هایی مانند مشاهده رفتار، داده‌های توصیفی و گزارش‌های شخصی استفاده می‌کنند.
- ۱۵- با توجه به مزایا و محدودیت‌های دو روش (روان‌سنجی و طرح کلی) معلم باید سعی کند ارزشیابی خود را به ترکیبی از دو روش استوار سازد.
- ۱۶- تعریف از نوع عملیاتی آزمایشی است.

## [سؤال‌های چهارگزینه‌ای فصل دوم]

- ۱- مقیاس اندازه‌گیری مورد استفاده در آزمون‌های چند گزینه‌ای از کدام نوع است؟  
(روان‌سنجی و اندازه‌گیری ۸۵)
- (۱) اسمی (۲) ترتیبی (۳) فاصله‌ای (۴) نسبی
- ۲- کدام یک از انواع ارزشیابی تقریباً در حاشیه قرار دارد و مکمل عمل آموزشی نیست؟  
(روان‌سنجی و اندازه‌گیری ۸۵)
- (۱) تشخیصی (۲) تکوینی (۳) مستمر (۴) مجموعی (پایانی)
- ۳- کدام یک از داده‌های زیر دارای مقیاس فاصله‌ای است؟  
(روان‌سنجی ۸۷)
- (۱) درجه حرارت (۲) رنگ چشم (۳) اندازه قد (۴) میزان موافقت با گویه‌های یک نگرش سنج
- ۴- جمع‌آوری اطلاعات به منظور قضاوت در مورد پدیده‌ها و تصمیم‌گیری ..... نامیده می‌شود.  
(سراسری ۸۷ و ۸۸)
- (۱) سنجش (۲) ارزیابی (۳) اندازه‌گیری (۴) ارزشیابی
- ۵- تفاوت بین ارزشیابی تکوینی و ارزشیابی پایانی اساساً در چیست؟  
(سراسری ۸۸)
- (۱) ارزشیابی تکوینی در طول سال و ارزشیابی پایانی در آخر سال انجام می‌شود.  
(۲) ارزشیابی تکوینی به تدریج ولی ارزشیابی پایانی یکباره انجام می‌شود.  
(۳) ارزشیابی تکوینی برای پاسخگوئی و ارزشیابی پایانی برای ارائه بازخورد انجام می‌شود.  
(۴) ارزشیابی تکوینی برای ارائه بازخورد و ارزشیابی پایانی برای دادن گواهی انجام می‌شود.
- ۶- در کدام فعالیت تعیین کیفیت از ویژگی‌های عمده محسوب می‌شود؟  
(روان‌سنجی ۸۹)
- (۱) ارزشیابی (۲) روان‌سنجی (۳) ارزیابی (۴) اندازه‌گیری
- ۷- مهم‌ترین هدف ارزشیابی تکوینی کدام است؟  
(روان‌سنجی ۸۹)
- (۱) ارائه بازخورد به معلم و دانش‌آموزان (۲) دادن گواهی قبولی یا مردودی  
(۳) انتخاب روش تدریس مناسب (۴) تعیین میزان آموخته‌های قبلی دانش‌آموزان
- ۸- شماره کارت ملی ایرانیان در کدام مقیاس اندازه‌گیری بیان می‌شود؟  
(روان‌سنجی ۸۲)
- (۱) اسمی (۲) ترتیبی (۳) فاصله‌ای (۴) نسبی

- ۹- کدام تعریف بیانگر فرایند «سنجش» است؟ (سنجش و اندازه‌گیری ۸۸)
- (۱) اختصاص کمیت به خصیصه‌ها یا صفات مختلف
  - (۲) تفسیر، ارزشگذاری و داوری در مورد عملکرد اختصاصی فرد
  - (۳) تحلیل چند جانبه عملکرد اختصاصی فرد با استفاده از فنون مختلف
  - (۴) اجرای آزمون پیشرفت تحصیلی به منظور تخمین عملکرد شغلی
- ۱۰- مهمترین تفاوت اندازه‌گیری در علوم پایه و علوم انسانی به متفاوت بودن ماهیت ..... ارتباط پیدا می‌کند. (سراسری-۹۰)
- (۱) آزمودنی‌ها
  - (۲) موضوع اندازه‌گیری
  - (۳) کاربردها
  - (۴) روش‌های اندازه‌گیری
- ۱۱- استفاده از تکنیک‌های مشاهده‌ای برای بهبود کدام ویژگی سنجش مبتنی بر مشاهدات مورد استفاده قرار می‌گیرد؟ (سراسری-۹۰)
- (۱) دقت
  - (۲) مربوط بودن
  - (۳) کیفیت
  - (۴) حساسیت
- ۱۲- مهمترین مرحله تهیه آزمون‌های روانشناختی کدام است؟ (سراسری ۹۳)
- (۱) نوشتن سوالات
  - (۲) تعیین روائی و اعتبار
  - (۳) تعریف صفت مورد اندازه‌گیری
  - (۴) تهیه نرم و ملاک قضاوت
- ۱۳- در کدام یک از مقیاس‌ها مبدأ اندازه‌گیری اعتباری است؟ (سراسری ۹۳)
- (۱) نسبی
  - (۲) ترتیبی
  - (۳) اسمی
  - (۴) فاصله‌ای
- ۱۴- ارزشیابی تکوینی برای چه هدفی انجام می‌گیرد؟ (سراسری ۹۳)
- (۱) به دست آوردن نتیجه عمل آموزشی
  - (۲) میزان دستیابی به هدف
  - (۳) اصلاح جریان آموزش
  - (۴) شناخت فرایند آموزش
- ۱۵- آزمونی که فقط یک ویژگی را اندازه‌گیری می‌کند و نمرات حاصل از اجرای آن، هم نرمال است دارای کدام مقیاس اندازه‌گیری است؟ (سراسری ۹۵)
- (۱) فاصله‌ای
  - (۲) اسمی
  - (۳) نسبی
  - (۴) ترتیبی

### پاسخنامه سؤال‌های چهارگزینه‌ای فصل دوم

- ۱- گزینه (۳). در نمره‌های حاصل از آزمون‌های چند گزینه‌ای صفر قراردادی وجود دارد.
- ۲- گزینه (۱). ارزشیابی تشخیصی به صورت مکمل عمل آموزشی مورد توجه معلمان قرار نمی‌گیرد و امری حاشیه‌ای به شمار می‌رود.
- ۳- گزینه (۱). درجه حرارت دارای مقیاس فاصله‌ای است و صفر آن قراردادی است.
- ۴- گزینه (۴).
- ۵- گزینه (۴).
- ۶- گزینه (۱). در ارزشیابی داده‌های حاصل از اندازه‌گیری با یک معیار مقایسه شده و مورد داوری قرار می‌گیرد بنابراین از ویژگی کیفیت برخوردار است.
- ۷- گزینه (۱).
- ۸- گزینه (۱). در شماره کارت ملی عدد به عنوان نشانه به کار رفته است.
- ۹- گزینه (۳).
- ۱۰- گزینه (۱).
- ۱۱- گزینه (۳). از بین ویژگی‌های سنجش، کیفیت آن از طریق مشاهدات بررسی می‌شود.
- ۱۲- گزینه (۳).
- ۱۳- گزینه (۴). مقیاس فاصله‌ای دارای صفر قراردادی است.
- ۱۴- گزینه (۳). یکی از اهداف ارزشیابی تکوینی اصلاح جریان آموزش است.
- ۱۵- گزینه (۱).



## سؤالات آزمون سال ۱۴۰۰

- ۱- در کدام مورد پایایی نمره اختلافی بین دو آزمون کمتر از میانگین پایایی دو آزمون نیست؟  
 (۱) تعداد سؤالات دو آزمون برابر باشد. (۲) همبستگی بین نمرات دو آزمون صفر باشد.  
 (۳) پایایی حداقل یک آزمون ۱ باشد. (۴) پایایی هر دو آزمون بالاتر از ۰/۵ باشد.
- ۲- سؤالات اولیه یک آزمون پیشرفت تحصیلی ریاضی در اختیار ۱۰ معلم با سابقه قرار گرفته و از آن‌ها خواسته شده مرتبط بودن سؤالات را با دانش ریاضی بر اساس طیف «ضروری است» «مفید است اما ضروری نیست» و «ضروری نیست» درجه‌بندی کنند. در صورتی که ۸ معلم سؤال اول را ضروری درجه‌بندی کرده باشند. نسبت روایی محتوایی کدام است؟  
 (۱) ۰/۲ (۲) ۰/۴ (۳) ۰/۶ (۴) ۰/۸
- ۳- ضریب آلفای کرونباخ آزمونی با سه خرده مقیاس که واریانس هر یک از آن‌ها برابر ۶، ۴ و ۷ و واریانس کل برابر ۳۲ است چند است؟  
 (۱) ۰/۶ (۲) ۰/۵۳ (۳) ۰/۷۰ (۴) ۰/۸۲
- ۴- در کدام گزینه به روش‌های مختلف سنجش ویژگی‌های عاطفی اشاره شده است؟  
 (۱) استفاده از پرسشنامه، تفکیک معنایی، مصاحبه  
 (۲) استفاده از مصاحبه، مقیاس لیکرت، سنجش روانی  
 (۳) سنجش نهان‌یابی، تفکیک معنایی، گروه‌سنجی  
 (۴) پرسشنامه بازپاسخ، مقیاس لیکرت، سنجش فیزیولوژیکی
- ۵- در یک آزمون ۲۰ سؤالی که هر سؤال آن یک نمره دارد، نمره خام شش دانشجوی کلاس برابر (۱۸، ۱۲، ۱۷، ۱۴، ۱۴، ۱۶ و ۱۳) بوده است. ضریب دشواری آزمون کدام است؟  
 (۱) ۰/۶۰ (۲) ۰/۷۵ (۳) ۰/۸۵ (۴) ۰/۹۰
- ۶- از ۱۰ دانشجویی که به سؤال اول یک آزمون در پیش آزمون پاسخ داده‌اند، ۴ نفر پاسخ درست و ۹ نفر از همین دانشجویان در پس آزمون به سؤال مذکور پاسخ درست داده‌اند، ضریب پیش و پس آزمون (ضریب حساسیت به آموزش سؤال کدام است؟  
 (۱) ۰/۳۰ (۲) ۰/۴۰ (۳) ۰/۵۰ (۴) ۰/۶۰
- ۷- کدام روش سنجش برای سنجش هوش‌های چندگانه کاربرد دارد؟  
 (۱) عملکردی (۲) رفتاری فرایند و فرآورده  
 (۳) (۴) ملاک محور و مشاهده‌ای

۸- اگر سؤالات یک آزمون پیشرفت تحصیلی روان‌شناسی عمومی، با اهداف مورد انتظار از فراگیران، کاملاً هماهنگی داشته باشد، کدام روایی این آزمون پیشرفت تحصیلی در حد مناسب است؟

- (۱) ملاکی (۲) سازه (۳) تجربی (۴) محتوایی

۹- اگر همبستگی بین نمرات خرده آزمون‌ها با نمره کل آزمون بالا باشد، نشان‌دهنده کدام یک از انواع روایی است؟

- (۱) صوری (۲) هم‌زمان (۳) پیش‌بین (۴) سازه

۱۰- در دو هدف یک آزمون وماپسته به ملاک، ۵۰ درصد دانش‌آموزان در دو هدف آن به تسلط و ۱۰ درصد در هر دو هدف به عنوان غیر حدتسلط مشخص شده‌اند، روایی پیش‌بین یادگیری در حد تسلط چقدر است؟

- (۱) ۰/۵ (۲) ۰/۹ (۳) ۰/۱۰ (۴) ۰/۶۰

۱۱- کدام نکته در طراحی سؤال زیر رعایت نشده است؟

«اگر طول و عرض زمین مستطیل شکلی به ترتیب برابر ۱۰ و ۸ متر باشد، محیط و مساحت آن به ترتیب کدام است؟»

- (۱) در هر سؤال فقط یک موضوع بگنجانید.  
 (۲) متن سؤال را با بیانی ساده و روشن بنویسید.  
 (۳) تمام مطالبات را در متن اصلی پرسش بگنجانید.  
 (۴) هر سؤال را مستقل از سایر سؤالات بنویسید.

۱۲- اگر دانشجویی بتواند برای اندازه‌گیری یکی از متغیرهای پژوهش خود (مثلاً تاب‌آوری) یک ابزار سنجش بسازد، این توانایی او در چه سطحی از سطوح شناختی طبقه‌بندی تجدیدنظر شده بلوم، طبقه‌بندی می‌شود؟

- (۱) آفریدن (۲) ارزشیابی کردن (۳) به کار بستن (۴) تحلیل کردن

۱۳- در یک آزمون پیشرفت تحصیلی با ۱۵ سؤال چهارگزینه‌ای میانگین کل ۱۰ و انحراف استاندارد ۳ است. در صورتی که مجموع واریانس آزمون ۲/۷ باشد، پایایی آزمون کدام است؟

- (۱) ۰/۴۵ (۲) ۰/۵۰ (۳) ۰/۷۵ (۴) ۰/۸۰

- ۱۴- پایایی یک پرسشنامه همدلی برابر با  $0/64$  است. در صورتی که واریانس نمرات این پرسشنامه ۱۰۰ باشد خطای معیار اندازه‌گیری آن کدام است؟  
 (۱) ۲ (۲) ۴ (۳) ۶ (۴) ۸
- ۱۵- کدام نوع روایی به قابل پذیرش بودن سؤال‌های آزمون هم برای استفاده‌کننده از آزمون و هم برای آزمودنی‌ها اشاره دارد؟  
 (۱) صوری (۲) محتوایی (۳) ملاکی (۴) همگرا
- ۱۶- با توجه به رابطه بین طول آزمون و پایایی مقیاس، در صورتی مقیاس با طول کمتر سریع‌تر به سطح مطلوب پایایی خواهد رسید که پایایی سؤالات برابر ..... باشد؟  
 (۱)  $0/1$  (۲)  $0/3$  (۳)  $0/5$  (۴)  $0/6$
- ۱۷- در یک سؤال چهارگزینه‌ای از ۲۰ نفر گروه قوی ۱۲ نفر به سؤال پاسخ صحیح داده‌اند. در صورتی که ضریب تمیز سؤال برابر با  $0/4 = D$  باشد ضریب دشواری ( $P$ ) سؤال کدام است؟  
 (۱)  $0/2$  (۲)  $0/4$  (۳)  $0/6$  (۴)  $0/8$
- ۱۸- اگر میانگین نمره افرادی که گزینه‌های یک سؤال چهارگزینه‌ای را انتخاب کرده‌اند به ترتیب برابر الف (۱۲، ب (۱۵، ج (۱۰، و د (۸ باشد، منطقی‌ترین گزینه برای کلید کدام است؟  
 (۱) الف (۲) ب (۳) ج (۴) د
- ۱۹- کدام نمره استاندارد براساس تبدیل خطی  $Y = 50 + 10Z$  محاسبه می‌شود؟  
 (۱)  $T$  (۲)  $Z$  (۳)  $CEEB$  (۴)  $Stanine$
- ۲۰- ملاک‌های قبولی در چهار امتحان در زیر ارائه شده است، با توجه به این ملاک‌ها کدام موارد از نوع هنجار- مرجع هستند؟  
 I. ۱۵ نفری که دارای بالاترین نمره هستند.  
 II. افرادی که به ۷۰ درصد از سؤال‌های امتحان پاسخ درست داده‌اند.  
 III. افرادی که در امتحان نمره ۲۴ و بالاتر کسب کنند.  
 IV. افرادی که نمره آن‌ها یک انحراف معیار بالاتر از میانگین است.  
 (۱)  $III - I$  (۲)  $IV - II$  (۳)  $III - II$  (۴)  $IV - I$

## [ پاسخنامه سؤال‌های آزمون سال ۱۴۰۰ ]

۱- گزینه (۲). برای حل این تست به فرمول پایایی نمرات افتراق یا اختلاف نیاز داریم.

$$r_d = \frac{\frac{r_{11} + r_{22}}{2} - r_{12}}{1 - r_{12}}$$

پایایی نمره اختلاف  $r_d$ ، پایایی آزمون اول  $r_{11}$ ، پایایی آزمون دوم  $r_{22}$ ، همبستگی

بین دو آزمون  $r_{12}$

یا فرمول به این صورت داریم:

$$\text{همبستگی بین دو آزمون} - \text{میانگین با متوسط پایایی دو آزمون} = \frac{\text{پایایی نمره افتراق}}{1 - \text{همبستگی بین دو آزمون}}$$

برای مثال اگر دو آزمون مورد نظر ما دارای همبستگی  $0/80$  است همچنین ضریب پایایی آزمون اول  $0/90$  و ضریب پایایی دوم  $0/85$  است. بنا به فرمول ضریب پایایی نمرات اختلاف برابر با  $0/40$  خواهد بود.

$$r_d = \frac{\frac{r_{11} + r_{22}}{2} - r_{12}}{1 - r_{12}} = \frac{\frac{0/90 + 0/85}{2} - 0/80}{1 - 0/80} = \frac{0/88 - 0/80}{0/20} = \frac{0/08}{0/20} = 0/40$$

در مثال بالا، ضریب پایایی نمره اختلاف ( $0/40$ ) بسیار کوچک‌تر از ضریب‌های پایایی هر یک از دو آزمون ( $0/90$ ،  $0/85$ ) است. علت آن است که ضریب همبستگی بین دو آزمون بزرگ است ( $0/80$ ).

نکات مهم پایایی نمرات افتراق یا اختلاف: (۱) هر چه همبستگی بین دو آزمون بیشتر باشد، به همان نسبت نمره اختلاف آن دو آزمون کمتر است. با کاهش میزان ضریب همبستگی بین دو آزمون ضریب پایایی نمرات افزایش می‌یابد. (۲) اگر میانگین پایایی دو آزمون با همبستگی آنها یکسان باشد، آنگاه پایایی نمرات افتراق مساوی صفر خواهند بود. مثلاً اگر متوسط پایایی دو آزمون  $0/80$  و همبستگی بین آنها  $0/80$  باشد، پایایی نمره افتراق به شکل زیر محاسبه می‌شود:

$$\text{پایایی نمره افتراق} = \frac{0/80 - 0/80}{1 - 0/80} = 0$$

۳) همچنین اگر پایایی هر کدام از آزمون‌ها ۱+ باشد، پایایی نمره افتراق یا اختلاف مساوی ۱+ خواهد بود و در این مورد میزان همبستگی بین دو آزمون اهمیتی ندارد.

$$\text{پایایی نمره افتراق} = \frac{1 - 0.70}{1 - 0.70} = \frac{0.3}{0.3} = 1$$

۴) اگر همبستگی بین دو نمرات آزمون صفر باشد، پایایی نمرات اختلافی مساوی یا بیشتر از میانگین پایایی دو اندازه‌گیری اولیه خواهد بود. برای مثال اگر همبستگی بین دو نمرات آزمون صفر و متوسط پایایی برابر ۰/۸۰ باشد، در فرمول داریم: (جواب تست)

$$\text{پایایی نمره افتراق} = \frac{0.80 - 0}{1 - 0} = \frac{0.80}{1} = 0.80$$

۵) پایایی کم یک نمره اختلافی باید روان‌شناس و محقق را نگران سازد. ۶) پایایی نمرات افتراق یا اختلاف به پایایی هر کدام از آزمون‌ها و همبستگی بین آنها وابسته است. مهم‌ترین نمره اختلاف مورد استفاده در آموزش اختلاف بین نمره‌های دانش‌آموزان در یک آزمون واحد یا در دو فرم موازی یک آزمون است که در دو زمان مختلف اجرا شده‌اند. نام دیگر این نمره اختلاف یا نمره افزایش یا نمره تغییر است. (روان‌آزمایی ترجمه دلاور، دلاور و همکاران، سنجش سیف، ص ۵۸۰، سنجش و اندازه‌گیری، دلاور و زهراکار ۱۸۲)

۲- گزینه (۳). روش لاشه برای تعیین ضریب روایی محتوایی: یکی از روش‌های تبدیل قضاوت کیفی داوران درباره روایی محتوایی به کمیت که ویژه آزمون‌های وابسته به ملاک یا ملاکی است به روش لاشه شهرت دارد و به آن ضریب روایی محتوایی (سی‌وی‌آر) گفته می‌شود. برای محاسبه این ضریب روایی محتوایی از تعدادی متخصص موضوع خواسته می‌شود تا هر یک از سؤال‌های آزمون را با محتوایی که سؤال از آن برگرفته شده به سه شکل «ضروری»، «مفید»، یا «غیرضروری» مورد داوری قرار دهند. نظر داوران به کمک فرمول زیر به صورت ضریب روایی محتوایی محاسبه می‌شود.

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}$$

تعداد داورانی که سؤال را ضروری تشخیص می‌دهند =  $n_e$ ، تعداد کل داوران =  $N$

موارد معلوم و خواسته شده سؤال:  $n_e = 8$ ،  $N = 10$ ،  $CVR = ?$

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}} = \frac{8 - \frac{10}{2}}{\frac{10}{2}} = \frac{8 - 5}{5} = \frac{3}{5} = 0.6$$

ضریب روایی محاسبه شده با فرمول بالا برای هر سؤال دارای دامنه‌ای از  $+1$  تا  $-1$  است که طبق آن هر چه رقم بزرگ‌تر باشد روایی سؤال بیشتر است. ضریب صفر (۰) نشان‌دهنده این است که نیمه از داوران سؤال را ضروری تشخیص داده‌اند، و ضریب  $(-1)$  حاکی از آن است که از نظر همه داوران سؤال غیر ضروری است و باید کنار گذاشته شود. ضریب به دست آمده به شکل بالا، مربوط به یک سؤال است. برای تعیین روایی کل آزمون، لازم است میانگین ضریب‌های روایی همه سؤال‌ها محاسبه شود که همان شاخص روایی محتوایی است. لازم به ذکر است که هنگام داوری درباره روایی محتوایی یک آزمون تنها اتکاء به شاخص روایی محتوایی کافی نیست. استفاده از جدول مشخصات آزمون هم لازم است؛ یعنی همان چیزی که اساس روایی محتوایی را تشکیل می‌دهد (سنجش سیف، ص ۵۵۲ - ۵۵۱).

۳- گزینه (۳). برای به دست آوردن ضریب آلفای کرانباخ از فرمول زیر استفاده می‌شود:

$$r_a = \frac{j}{j-1} \left( 1 - \frac{\sum \sigma_j^2}{\sigma^2} \right)$$

ضریب پایایی کل آزمون  $r_a$ ، تعداد بخشهای آزمون  $j$ ، واریانس نمرات هر بخش  $\sigma_j^2$ ،

واریانس نمرات کل آزمون  $\sigma^2$

معلومات و موارد خواسته شده سؤال:  $j = 3$  یا تعداد بخشهای آزمون،  $\sigma^2 = 32$  واریانس

نمرات کل آزمون، و واریانس نمرات هر بخش  $\sigma_1^2 = 6$ ،  $\sigma_2^2 = 4$ ،  $\sigma_3^2 = 7$ ،  $r_a = ?$

$$r_a = \frac{j}{j-1} \left( 1 - \frac{\sum \sigma_j^2}{\sigma^2} \right) = \frac{3}{3-1} \left( 1 - \frac{6+4+7}{32} \right) = \frac{3}{2} \left( 1 - \frac{17}{32} \right) = \frac{3}{2} \left( \frac{15}{32} \right) = \frac{45}{64} = 0.703125 \approx 0.7$$

سومین روش تعیین پایایی آزمون با تأکید بر همسانی درونی روش آلفا نام دارد که به آن ضریب آلفا، یا ضریب آلفای کرانباخ یا حتی آلفا می‌گویند. کرانباخ (۱۹۵۱) واضح این روش تعیین پایایی است. در این روش، اجزا یا قسمت‌های آزمون برای سنجش ضریب پایایی آزمون به کار می‌روند. اگر قسمت‌های آزمون همان سؤال‌های آزمون باشند و سؤال‌ها به صورت درست (۱) و غلط (۰) باشد، ضریب آلفای کرانباخ برابر خواهد بود با  $KR20$ . اما اگر، به جای سؤال‌ها یا ماده‌ها، آزمون از بخش‌ها یا قسمت‌هایی تشکیل شده باشد، مثل زمانی که یک آزمون از تعدادی خرده آزمون تشکیل یافته است، و بخواهیم از آنها در محاسبه ضریب پایایی کل آزمون استفاده کنیم، از روش کرانباخ مطابق با فرمول بالا استفاده می‌کنیم. روش آلفای کرانباخ در شرایطی که نمرات سؤال نه به صورت دو ارزشی صفر و ۱ بلکه به صورت

چندارزشی تعیین می‌شوند  $KR20$  برای آن مناسب نیست نیز قابل استفاده است. (سنجش سیف، ص ۵۷۵)

۴- گزینه (۱). روش‌های مختلف سنجش ویژگی‌های عاطفی عبارتند از: (۱) پرسشنامه یا عقیده‌سنج که دو نوع پرسشنامه بسته‌پاسخ و پرسشنامه بازپاسخ وجود دارد. (۲) روش تفکیک معنایی یا افتراق معنایی (۳) مقیاس لیکرت (۴) استفاده از مصاحبه (۵) گروه‌سنجی (۶) استفاده از یادداشتهای روزانه (سنجش سیف، ص ۳۹۱).

نکات تست: سنجش نهان‌یابی (نامرئی) و سنجش روانی - فیزیولوژیکی از روش‌های سنجش رفتاری می‌باشند که صرفاً رفتارهای آشکار فرد مورد مشاهده و سنجش قرار می‌گیرند.

۵- گزینه (۲).

۶- گزینه (۳). فرمول ضریب تمیز پیش و پس از آموزش یا ضریب حساسیت به آثار آموزش:

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

در این فرمول، تعداد کسانی که پس از آموزش به سؤال جواب درست داده‌اند  $R_A$ ، تعداد کسانی که پیش از آموزش به سؤال جواب داده‌اند  $R_B$  و تعداد کل افراد  $T$  = معلومات و موارد خواسته شده سؤال:  $R_A = 9$ ،  $R_B = 4$  و  $T = 10$

$$S \text{ یا } D = \frac{R_A - R_B}{T} = \frac{9 - 4}{10} = \frac{5}{10} = 0.5$$

با توجه به ضریب تمیز، این سؤال یک سؤال کارآمد است با یک الگوی طبیعی، زیرا بعضی از دانش‌آموزان پیش از آموزش به آن جواب درست داده‌اند، اما تعداد بیشتری از دانش‌آموزان پس از آموزش به آن جواب درست داده‌اند.

ضریب تمیز پیش و پس از آموزش مربوط به سؤال‌های آزمون‌های وابسته به ملاک می‌باشد. مانند ضریب تمیز آزمون‌های وابسته به هنجار، ضریب پیش و پس از آزمون از -۱ تا +۱ در تغییر است. سؤال‌های دارای ضریبهای تمیز مثبت به کیفیت خوب کل آزمون کمک می‌کنند. وقتی ضریب تمیز سؤالی صفر یا نزدیک به صفر است نشان می‌دهد که سؤال قدرت تمیز کل آزمون هیچ کمکی نم‌کند و آن سؤالی است که همه آزمون‌شوندگان در پیش آزمون و پس آزمون به آن جواب درست یا غلط داده‌اند. شرایط آرمانی برای هر سؤال آن است که تفاوت ضریب‌های دشواری قبل و بعد از آموزش آن +۱ باشد، یعنی ضریب دشواری سؤال در پیش آزمون صفر و ضریب دشواری آن در پس آزمون +۱ باشد (سنجش سیف، ص ۴۴۴ - ۴۴۱).

۷- گزینه (۱). مدافعان نظریه هوشهای چندگانه گاردنر کوشیده‌اند تا روش سنجش عملکردی را برای اندازه‌گیری آنها به کار بندند. به باور این افراد، روش‌های سنجش عملکردی می‌تواند امکاناتی فراهم آورد تا پیشرفت دانش‌آموزان به چندین راه مختلف مورد شناسایی قرار بگیرد. گاردنر در رابطه با سنجش عملکردی هوشهای مختلف گفته است: تأکید بر سنجش عملکردی با نظریه هوشهای چندگانه به خوبی حمایت می‌شود. بنا به گفته نیتکو، استفاده از روش‌های سنجش عملکردی برای اندازه‌گیری هوشهای چندگانه مستلزم این است که استانداردهای سطح بالایی برای توفیق یادگیرنده در هدف‌های یادگیری منظور شود و از او خواسته شود تا آن موفقیت را به راه‌های مختلفی نشان دهد.

**نکات تست:** هوارد گاردنر، یکی از نظریه‌پردازان هوش، نظریه‌ای وضع کرده شامل هشت هوش متفاوت و مستقل از یکدیگر با نام‌های هوش زبانی، منطقی - ریاضی، فضایی، بدنی - جنبشی، موسیقایی، میان‌فردی، درون‌فردی، و طبیعت‌گرایانه. در سنجش عملکردی یا آزمون‌های عملکردی فرایندها و فراورده‌های یادگیری دانش‌آموزان و دانشجویان به طور مستقیم سنجش می‌شوند. معلمان می‌توانند از آزمون‌های عملکردی برای سنجش یادگیری‌های شناختی پیچیده، نگرش‌ها و مهارت‌های اجتماعی در موضوعات مختلف تحصیلی مانند علوم، ریاضی، تعلیمات اجتماعی استفاده کنند. برای این منظور، معلمان باید موقعیت‌هایی را تدارک ببینند که در آن رفتار یادگیرندگان را به طور مستقیم مشاهده و سنجش نمایند (سنجش سیف، ص ۲۹۲ - ۲۶۹)

۸- گزینه (۴). روایی محتوایی به این مطلب اشاره می‌کند که نمونه سؤال‌های مورد استفاده در یک آزمون تا چه حد معرف تمامی محتوا یا کل جامعه سؤال‌های ممکن است که می‌توان از محتوا یا موضوع موردنظر تهیه کرد. هر چه آزمون از این لحاظ بهتر باشد، روایی بیشتری دارد. بنابراین اگر معلم درس فیزیک بخواهد برای درس خود یک آزمون پیشرفت تحصیلی روا بسازد، آزمون او باید در برگیرنده نمونه‌ای درست و دقیق از مطالب فیزیک باشد که به دانش‌آموزان آموزش داده است. در ضمن، این آزمون نباید چیزی خارج از محتوا و هدف‌های درس فیزیک این معلم را اندازه‌گیری بکند. بنابراین، برای ساختن آزمون‌های پیشرفت تحصیلی روا، باید بکوشیم تا سؤال‌های آزمون نمونه کاملی از هدف‌ها و محتوای درس باشند. جدول مشخصات بهترین روش برای انجام دادن چنین کاری است. جدول مشخصات وسیله‌ای است که در آن محتوای درس و هدف‌های آموزشی به خوبی ارائه می‌شوند (سنجش و اندازه‌گیری سیف، ص ۵۳۵) با توجه به مطلب فوق، روایی محتوایی بیشتر در آزمون‌های پیشرفت تحصیلی مورد استفاده قرار می‌گیرد. پس اگر سؤالات پیشرفت



تحصیلی یک آزمون با اهداف مورد انتظار از فراگیران طبق جدول مشخصات آزمون هماهنگی داشته باشد، نشان‌دهنده روایی محتوایی است.

۹- گزینه (۴). یکی از روش‌های تعیین روایی سازه، همسانی درونی است. در روش همسانی درونی، ساخت درونی آزمون بررسی می‌شود. ملاک مورد استفاده برای بررسی همسانی درونی نمره کل آزمون است. روش معمول برای این کار این است که همبستگی بین نمرات آزمون فرعی یا خرده آزمون‌ها با نمرات کل آزمون محاسبه می‌شود، و هر یک از این خرده‌آزمون‌ها که با نمره کل آزمون همبستگی اندکی را نشان داد از آزمون حذف می‌شود. از لحاظ نظری نمرات خرده آزمون‌ها باید با نمره کل آزمون همبستگی زیادی داشته باشند، زیرا فرض بر این است که همه آنها بر روی هم حوزه رفتاری واحدی را اندازه‌گیری می‌کنند. در ضمن، خرده‌آزمون‌ها نباید با هم همبستگی زیادی داشته باشند، زیرا در چنین حالتی فرض می‌شود که همه آنها یک چیز را اندازه می‌گیرند و لذا تکراری هستند (سنجش سیف، ص ۵۴۵)

۱۰- گزینه (۴). ۵۰ درصد از دانش‌آموزان در هر دو هدف به حد تسلط رسیده‌اند ولی ۱۰ درصد در هر دو هدف به حد تسلط نرسیده‌اند. بنابراین، ۶۰ درصد یعنی ۶۰ نفر از ۱۰۰ نفر، در دو هدف عملکرد یکسان دارند.

روایی پیش‌بینی آزمون‌های وابسته به ملاک: در آزمون‌های وابسته به ملاک روایی پیش‌بینی به دو منظور زیر به کار می‌رود. (۱) پیش‌بینی عملکرد یادگیرنده در یک هدف آموزشی بر اساس هدف آموزشی دیگر. برای تعیین اینکه عملکرد دانش‌آموزان در یک هدف آموزشی تا چه میزانی عملکرد آنان را در هدف آموزشی دیگر که وابسته به هدف اول است تعیین می‌کند، لازم است ضریب همبستگی (پیرسون) عملکرد دانش‌آموزان را در سؤال‌های مربوط به هدف‌های مورد نظر تعیین نماییم. مانند مهارت در منها کردن (هدف اول) سپس آموختن تقسیم طولانی (هدف دوم). (۲) پیش‌بینی یادگیری در حد تسلط. در این روش، بهتر است روایی پیش‌بینی آزمون‌های وابسته به ملاک به عوض ضریب همبستگی به صورت درصد دانش‌آموزانی که عملکرد حدتسلط و عملکرد غیر حدتسلط آنان در هدف‌های اول و دوم یکسان است تعیین گردد (سنجش سیف، ص ۵۵۶ - ۵۵۳).

۱۱- گزینه (۱). توضیح چند مورد از قواعد تهیه سؤال‌های چندگزینه‌ای: (۱) هر سؤال باید یک موضوع مهم یا یک هدف آموزشی را اندازه بگیرد. در طرح سؤال، نویسنده سؤال باید از خود بپرسد که آیا دانستن و ندانستن پاسخ سؤال در صلاحیت آزمون‌شونده، از لحاظ مطلب مورد آزمایش، تأثیر دارد یا نه؟ اگر جواب مثبت بود آن‌گاه باید به گنجاندن موضوع در سؤال

اقدام کند. بنابراین؛ هر سؤال باید یک هدف آموزشی را بسنجد و شامل بخش مهمی از محتوای درس باشد. (۲) بیشتر از یک هدف یا یک مطلب در هر سؤال قرار ندهید. هر سؤال باید تنها به یک مطلب یا یک هدف آموزشی مربوط باشد. اگر بیش از یک مطلب یا یک هدف در سؤال مطرح شود، نه تنها سؤال را پیچیده می‌کند بلکه اگر آزمون شونده در پاسخگویی به سؤال موفق نباشد معلوم نیست که کدام یک از مطالب را ندانسته و علت عدم موفقیت او چه بوده است. سنجش سیف، ص ۲۰۱). در هر پرسش، بیش از یک مطلب نگنجانید. زیرا اگر دانش‌آموزی نتواند پاسخ صحیح را پیدا کند، معلوم نمی‌شود که آیا همه مطالب مطرح شده را نمی‌دانسته یا یکی از آنها را، و اگر یکی از مطالب را نیاموخته است، آن مطلب کدام بوده است. اشکال دیگر این‌گونه پرسش‌ها این است که هرگاه آزمودنی پاسخ قسمتی از پرسش را بداند، بدون آگاهی از مطالب دیگر، احتمالاً پاسخ درست را پیدا خواهد کرد (اصول روان‌سنجی پاشا شریفی، ص ۱۴۷).

۱۲- گزینه (۱). سطح آفریدن (خلق کردن). قرار دادن عناصر یا اجزاء در یک الگو یا ساختار نو. الف: تولید کردن (فرضیه ساختن). مانند: ساختن یک فرضیه برای توجیه داده‌های مشاهده شده. ب: طرح‌ریزی کردن. مانند: طراحی یک برنامه ارزشیابی برای تعیین اثربخشی برنامه‌های آموزش هنر در یک مدرسه. ج: پدیدآوردن (ساختن). مانند: ساختن یک وسیله سنجش برای ارزشیابی از درس علوم اجتماعی پایه سوم راهنمایی. چند نفر از صاحب‌نظران آموزشی (آندرسون، کراتول، و همکاران، ۲۰۰۱) در طبقه‌بندی حوزه شناختی معروف به طبقه‌بندی بلوم تجدیدنظر کرده و طبقه‌بندی تازه‌ای با نام یک طبقه‌بندی برای یادگیری، آموزش، و سنجش به وجود آورده‌اند. در طبقه‌بندی تازه حوزه شناختی یک بعد دانش و یک بُعد فرایند شناختی وجود دارد. بُعد دانش: شامل دانش امور واقعی، دانش مفهومی، دانش روندی و دانش فراشناختی است. بُعد فرایند شناختی در برگرفته به یادآوردن، فهمیدن، به‌کار بستن، تحلیل کردن، ارزشیابی کردن و آفریدن است. طبقه‌های هر دو بُعد این طبقه‌بندی به صورت سلسله‌مراتبی، یعنی از عینی به انتزاعی و از ساده به پیچیده تنظیم یافته‌اند (سنجش سیف، ص ۱۴۲). با توجه به مطلب بالا، ساخت یک وسیله سنجش مانند سنجش تاب‌آوری در سطح آفریدن است.

۱۳- گزینه (۱). چون سؤالات بصورت چهار گزینه‌ای هستند و همچنین به علت داده‌های صورت سؤال، از فرمول ضریب آلفای کرانباخ استفاده می‌کنیم.  
فرمول ضریب آلفای کرانباخ

$$r_a = \frac{j}{j-1} \left( 1 - \frac{\sum \sigma_j^2}{\sigma^2} \right)$$

ضریب پایایی کل آزمون  $r_a$ ، تعداد بخشهای آزمون  $j$ ، واریانس نمرات هر بخش  $\sigma_j^2$ ، واریانس نمرات کل آزمون  $\sigma^2$

معلومات و موارد خواسته شده سؤال:  $j = 15$  یا تعداد سؤالهای آزمون،  $3 =$  انحراف معیار

یا  $\sigma^2 = 9$  واریانس،  $10 =$  میانگین،  $\sum \sigma_j^2 = 27$  یا مجموع واریانس سؤالات،  $r_a = ?$

$$r_a = \frac{j}{j-1} \left( 1 - \frac{\sum \sigma_j^2}{\sigma^2} \right) = \frac{15}{15-1} \left( 1 - \frac{27}{9} \right) = \frac{15}{14} \left( \frac{9-27}{9} \right) = \frac{15}{14} \left( \frac{-18}{9} \right) = \frac{15}{14} \left( \frac{-2}{1} \right) = \frac{15}{14} \left( \frac{6}{3} \right) = \frac{94.5}{126} = 0.75$$

(سنجش سیف، ص ۵۷۴)

۱۴- گزینه (۳). خطای معیار اندازه‌گیری، شاخصی است که متوسط اندازه خطا را در اندازه‌گیری با یک آزمون نشان می‌دهد. هر قدر خطای اندازه‌گیری آزمون بیشتر باشد، اطمینان ما نسبت به دقت خصیصه مورد اندازه‌گیری کمتر است. بین خطای معیار اندازه‌گیری و ضریب پایایی آزمون رابطه معکوس وجود دارد. هر چه ضریب پایایی به یک نزدیکتر باشد، اندازه خطای معیار اندازه‌گیری کمتر می‌شود و بالعکس.

خطای معیار اندازه‌گیری آزمون را از فرمول زیر می‌توان محاسبه کرد.

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

در این فرمول،  $SEM =$  خطای معیار اندازه‌گیری،  $SDt =$  انحراف معیار نمره‌های خام یا مشاهده شده،  $r_{tt} =$  ضریب پایایی آزمون است. معلومات سؤال و موارد خواسته شده:

$r_{tt} = 0.64$  ضریب پایایی  $s^2 = 100$  واریانس  $SEM = ?$  خطای معیار اندازه‌گیری

در صورت سؤال به ما واریانس برابر با ۱۰۰ داده است در حالی که در فرمول خطای معیار اندازه‌گیری باید از انحراف معیار استفاده کنیم. پس اگر از واریانس جذر بگیریم انحراف معیار

یا استاندارد به دست می‌آید.  $s = \sqrt{100} = 10$ .

$$SEM = SDt \sqrt{1 - r_{tt}} \gg SEM = 10 \sqrt{1 - 0.64} \gg SEM = 10 \sqrt{0.36} \gg SEM = 10 \times 0.6 \gg SEM = 6$$

(اصول روانسنجی پاشا شریفی، ص ۲۵۱) و (سنجش و اندازه‌گیری سیف، ص ۵۸۵)

۱۵- گزینه (۱). روایی صوری یک از مشتقات روایی محتوایی است. روایی صوری به این مطلب اشاره می‌کند که سؤالهای آزمون تا چه حد در ظاهر شبیه به موضوعی هستند که برای اندازه‌گیری آن تهیه شده‌اند. از آنجاکه روایی صوری نوع خاصی از روایی محتوایی است،

همان روشی که برای تعیین روایی محتوایی به کار می‌رود یعنی استفاده از نظر متخصصان، در اینجا نیز برای تعیین روایی صوری قابل استفاده است. هرچند روایی صوری یک ویژگی مطلوب آزمون است با این حال در بعضی آزمون‌ها این ویژگی نه تنها ضروری نیست، بلکه بهتر است وجود نداشته باشد. برای مثال، اگر کسی بخواهد بیماری روانی را در افراد تشخیص دهد، بهتر است آزمونی را به کار ببرد که روایی صوری اندکی دارد. ثن‌دایک و هگن می‌گویند که گاه صورت ظاهر یک آزمون (روایی صوری)، از لحاظ قابلیت پذیرش و معقول بودن آزمون برای آزمون شونده‌گان، دارای اهمیت است. می‌توان گفت تا آن اندازه که ظاهر آزمون بر انگیزش آزمون شونده اثر می‌گذارد، روایی صوری می‌تواند مهم باشد. (سنجش سیف، ص ۵۳۶).

۱۶- گزینه (۴). تعداد سؤالات آزمون یا طول آزمون با سؤالات مشابه بر پایایی یا اعتبار آن تأثیر می‌گذارد. هر چه تعداد سؤالات آزمون بیشتر باشد، ضریب پایایی آزمون نیز بیشتر خواهد شد. تا حدودی دلیل این تأثیر این است که افزایش تعداد سؤال‌های آزمون با افزایش پراکندگی موجود در نمرات آزمودنی‌ها همراه است. پس در صورتی مقیاس با طول کمتر سریع‌تر به سطح مطلوب پایایی خواهد رسید که پایایی سؤالات بیشتر باشد. که در میان گزینه‌ها پایایی ۰/۶ از همه بیشتر است.

۱۷- گزینه (۲). برای به دست آوردن ضریب دشواری ابتدا باید از طریق فرمول ضریب تمیز، مقدار انتخاب‌های صحیح گروه ضعیف را به دست آریم. قدرت تمیز یا تفکیک سؤال عبارت است از میزان قدرت سؤال در تفکیک آزمون‌شوندگان دارای نمره بالا، از آزمون‌شوندگان دارای نمره پایین. به عبارت دیگر؛ ضریب تمیز که با  $D$  نشان داده می‌شود قدرت سؤال را در تمایزگذاری یا تشخیص بین گروه قوی و گروه ضعیف آزمون‌شوندگان مشخص می‌کند. یعنی معلوم می‌سازد که سؤال تا چه اندازه می‌تواند گروه قوی را از گروه ضعیف جدا سازد. برای محاسبه ضریب تمیز یک سؤال از فرمول زیر استفاده می‌شود:

$$D = \frac{\text{انتخاب‌های درست گروه پایین} - \text{انتخاب‌های درست گروه بالا}}{\text{تعداد افراد یک گروه (بالا یا پایین)}} = \text{تمیز ضریب}$$

معلومات و موارد خواسته شده سؤال:  $D = 0,4$ ،  $12 =$  انتخاب‌های درست گروه بالا یا قوی،  $20 =$  تعداد افراد گروه بالا یا قوی، انتخاب‌های درست گروه پایین یا ضعیف  $= ?$  ضریب دشواری  $= ?$

$$0,4 = \frac{\text{انتخاب‌های درست گروه پایین} - 12}{20} = 0,4 \times 20 = 12 - X \gg 8 = 12 - X \gg X = 12 - 8 = 4$$

پس انتخاب‌های درست گروه ضعیف یا پایین برابر ۴ شد. و حال از فرمول ضریب دشواری استفاده می‌کنیم.

$$P = \frac{\text{انتخاب‌های درست گروه پایین} + \text{انتخاب‌های درست گروه بالا}}{\text{تعداد افراد گروه بالا} + \text{تعداد افراد گروه پایین}} \times ۱۰۰$$

$$= \frac{۱۲ + ۴}{۲۰ + ۲۰} \times ۱۰۰ = \frac{۱۶}{۴۰} \times ۱۰۰ = ۴۰$$

ضریب دشواری به دست آمده برابر ۴۰ است که اگر بصورت نسبت ۱۰۰ در نظر بگیریم ۰/۴ خواهد شد (سنجش سیف، ص ۴۳۵ - ۴۳۱).

۱۸- گزینه (۲). چون میانگین افرادی که گزینه ب را انتخاب کرده‌اند بیشتر است، پس منطقی‌ترین گزینه برای کلید همان گزینه ب می‌باشد. البته این در شرایطی است که سؤال طبق قواعد سؤال‌های چندگزینه‌ای طرح شده باشد.

۱۹- گزینه (۱). آن دسته از نمرات معیار که دارای توزیعی با میانگین ۵۰ و انحراف معیار ۱۰ است توزیع  $T$  نام دارد. می‌توان فرمول محاسبه نمرات  $T$  را به صورت زیر نوشت:  
 $T = ۱۰Z + ۵۰$  هر نمره  $T$  به تنهایی و به طور مستقیم قابل تفسیر و استفاده است. برای مثال، نمره  $T = ۵۵$  حاکی از این است که آن نمره به اندازه نصف انحراف معیار بالای میانگین است، و نمره  $T = ۳۵$  نشان می‌دهد که آن نمره به اندازه یک و نیم انحراف معیار پایین میانگین قرار دارد. (سنجش و اندازه‌گیری سیف، ص ۵۱۸)

نکات تست: نمرهای  $CEEB$  که معمولاً در آزمون‌های ورودی دانشگاه‌ها به کار می‌روند، نمره‌هایی هستند که میانگین آنها ۵۰۰ و انحراف معیار آنها ۱۰۰ است و رابطه آن برابر است  $CEEB = ۱۰۰Z + ۵۰۰$  با (اصول روان‌سنجی پاشا شریفی، ص ۲۹۴).

نمرات نه بخشی یا نه‌گانه ( $Stanine\ score$ ) به صورت نمرات معیار یک‌رقمی نشان داده می‌شود. علت نام‌گذاری این مقیاس نمرات به ۹ بخشی آن است که توزیع این نمرات به ۹ بخش محدود می‌شود. این مقیاس دارای میانگین ۵ و انحراف معیار ۲ است. یعنی

$$\text{نمرات نه بخشی} = ۲z + ۵$$

نمره  $Z$ : آن دسته از نمرات معیار یا استاندارد که دارای توزیعی با میانگین صفر و انحراف معیار واحد (۱) هستند به نمرات  $Z$  شهرت دارند. برای تبدیل نمرات خام به نمرات  $Z$  از فرمول زیر استفاده می‌شود:

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

۲۰- گزینه (۱). گزینه یک با توجه به عملکرد سایر افراد در نظر گرفته شده است یعنی ۱۵ نفری که دارای بالاترین نمره هستند. گزینه چهار هم بر اساس منحنی توزیع طبیعی و با توجه به عملکرد سایر آزمودنی‌ها صورت پذیرفته است. گزینه سه بر اساس یک ملاک از پیش تعیین شده صورت می‌گیرد و عملکرد یادگیرنده با توجه به آن ملاک در اینجا افرادی که به ۷۰ درصد سؤالات پاسخ درست بدهند، مورد قضاوت قرار می‌گیرد و نه با توجه به نمره سایر دانش‌آموزان. گزینه سه مربوط به ارزشیابی ملاک مرجع است زیرا عملکرد آزمودنی با یک ملاک مشخص یعنی کسانی که نمره ۲۴ و بالاتر کسب نمایند و بدون توجه به عملکرد سایر افراد گروه مقایسه و درباره‌ی قضاوت می‌شود.

ارزشیابی آموزشی بر اساس ملاک مورد استفاده به دو دسته زیر تقسیم می‌شوند:

**الف) ارزشیابی وابسته به ملاک یا ملاک مرجع:** در این نوع ارزشیابی معیار یا ملاک ارزشیابی از پیش تعیین می‌شود و عملکرد یادگیرنده در آزمون با توجه به آن ملاک مورد قضاوت قرار می‌گیرد. در این نوع ارزشیابی که بیشتر ویژه پیشرفت تحصیلی است نوعی ملاک مطلق که همان هدف‌های آموزشی از پیش تعیین شده هستند مورد استفاده قرار می‌گیرند. در استفاده از ملاک مطلق در ارزشیابی پیشرفت تحصیلی، میزان یادگیری یا پیشرفت یادگیرنده با توجه به آنچه آموزش داده شده است ارزشیابی می‌شود. یعنی کسب موفقیت در آزمون‌هایی که بر مبنای ملاک مطلق هستند مستلزم یادگیری هدف‌های آموزشی در سطحی است که معلم تعیین می‌کند.

**ب) ارزشیابی وابسته به هنجار یا هنجار - مرجع:** آن نوع ارزشیابی است که به جای یک ملاک مطلق از پیش تعیین شده بر نوعی ملاک نسبی وابسته است. در این نوع ارزشیابی، اگر بخواهیم پیشرفت تحصیلی گروهی دانش‌آموز را ارزشیابی کنیم، عملکرد دانش‌آموزان را با یکدیگر مقایسه می‌کنیم، نه با یک ملاک مشخص از پیش تعیین شده. علت نام‌گذاری این نوع ارزشیابی به ارزشیابی هنجاری این است که در آن ملاک یا معیار ارزشیابی نوعی هنجار گروهی است. برای مثال، نمره ۱۶ یک دانش‌آموز اگر از ۹۰٪ نمرات سایر دانش‌آموزان بالاتر باشد نمره خوبی ارزشیابی می‌شود، اما اگر همین نمره در حد ۱۰٪ آزمون‌شوندگان باشد نمره ضعیفی به حساب می‌آید.

مقایسه ارزشیابی وابسته به ملاک با ارزشیابی وابسته به هنجار: (۱) تفاوت در نحوه قضاوت در مورد نتایج آزمون‌ها، (۲) ارزشیابی هنجاری و آزمون‌های مبتنی بر ملاک نسبی در سنجش هوش و استعداد و نیز در سنجش شخصیت و سایر جنبه‌های عاطفی به کار می‌روند. در مقابل، ارزشیابی ملاکی و آزمون‌های مبتنی بر ملاک مطلق به طور عمده در سنجش پیشرفت تحصیلی که به منظور تعیین میزان موفقیت یادگیرندگان در رسیدن به هدف‌های

آموزشی اجرا می‌شوند مورد استفاده قرار می‌گیرند. به عنوان مثال، آزمون‌های ورودی ویژه داوطلبان ورود به آموزش عالی (کنکور سراسری) موردی از ارزشیابی هنجاری است. در این آزمون‌ها، عملکرد هر داوطلب با توجه به عملکرد سایر داوطلبان مورد قضاوت قرار می‌گیرد. در مقابل، امتحانات نهایی مقاطع تحصیلی مختلف دبستان، راهنمایی و دبیرستان بر نوعی ملاک مطلق که محتوای درسی و هدف‌های آموزشی هستند متکی‌اند، و قبولی و یا عدم قبولی شرکت‌کنندگان در این امتحانات، صرفاً با توجه به این ملاک مطلق و مستقل تعیین می‌شود. ۳) نمرات آزمون‌های هنجاری معمولاً به صورت توزیع طبیعی پراکنده می‌شوند. در این توزیع اکثر نمرات حول و حوش میانگین قرار دارند و تعداد کمی از آنها بسیار بالاتر از میانگین و به همان نسبت هم تعداد کمی از نمرات بسیار پایین‌تر از میانگین قرار دارند. ولی توزیع نمرات آزمون‌های ملاکی یک منحنی بهنجار نمی‌سازد. بلکه این نمرات دارای توزیعی است که معرف منحنی‌های دارای کجی منفی است. توزیع کجی منفی نشان‌دهنده‌ی بالا بودن سطح اکثریت نمرات دانش‌آموزان است که نشان می‌دهد تعداد زیادی از آنان به اکثریت هدف‌های آموزشی مورد نظر رسیده‌اند. (سنجش و اندازه‌گیری سیف، ص ۹۹ - ۹۶)

## منابع و مأخذ

- ۱- اردبیلی، یوسف، اصول علمی تهیه، اجرا و استاندارد کردن تست، دانشسرایعالی ۱۳۴۶
- ۲- استودولا، مترجم شریفی، مبانی اندازه‌گیری و آزمون‌های تربیتی ۱۳۵۳
- ۳- ثرن‌دایک و دیگران، مترجم هومن، روانسنجی کاربردی، دانشگاه تهران ۱۳۶۹
- ۴- ثرن‌دایک و دیگران، مترجم مهندسی و دیگران، سنجش و اندازه‌گیری در تعلیم و تربیت، اندیشه ۱۳۴۳
- ۵- سیف، علی اکبر، اندازه‌گیری و ارزشیابی پیشرفت تحصیلی، آگاه ۱۳۷۱
- ۶- سیف، علی اکبر، روش‌های اندازه‌گیری و ارزشیابی آموزشی، آگاه
- ۷- شریفی، حسن پاشا، اصول روان‌سنجی و روان آزمائی، رشد ۱۳۷۲
- ۸- شریفی، شریفی، اصول روان‌سنجی و روان آزمائی رشد ۱۳۹۱، وزارت آموزش و پرورش ۱۳۷۰
- ۹- شعبانی، حسن، مهارت‌های آموزشی و پرورشی، سمت ۱۳۸۳
- ۱۰- گنجی - حمزه، آزمون‌های روانی، آستان قدس ۱۳۶۷
- ۱۱- گنجی، و ثابت، مبانی نظری آزمون‌های روانی (روان سنجی) ساوالان ۱۳۷۹
- ۱۲- گنجی، حمزه، اندازه‌گیری و ارزشیابی در آموزش و پرورش، بعثت ۱۳۷۳
- ۱۳- مگنسون، دیوید، مترجم براهنی، مبانی نظری آزمون‌های روانی دانشگاه تهران ۱۳۵۱
- ۱۴- مهندسی خسرو، اندازه‌گیری و سنجش در روانشناسی و تعلیم و تربیت، اندیشه ۱۳۴۳
- ۱۵- نادری، عزت اله و دیگران، سنجش و اندازه‌گیری و بنیادهای تحلیلی ابزارهای آن در روانشناسی و علوم تربیتی، معین، ۱۳۷۱
- ۱۶- نوروزی، ابوالقاسم، مقدمه‌ای بر سنجش و اندازه‌گیری در علوم تربیتی، آوای نور ۱۳۷۴
- ۱۷- هومن، حیدرعلی، اندازه‌گیری‌های روانی و تربیتی و فن تهیه تست، سلسله ۱۳۶۶