

بسمه تعالی

ارزشیابی آموزشی و چگونگی تحلیل سوالات آزمون

برای تهیه یک آزمون خوب تحلیل پاسخهای آزمون‌شوندگان و بررسی تجربی آنها ضروری است. در تحلیل آزمون‌های پیشرفت تحصیلی دانشجویان می‌باید اطلاعات ذیل در اختیار اساتید قرار گیرد:

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

الف) ارزشیابی وابسته به ملاک:

در این نوع ارزشیابی معیار یا ملاک ارزشیابی از پیش تعیین می‌شود و عملکرد فراگیر 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 این شرایط سنجش برای بررسی دستیابی یا عدم دستیابی به اهداف آموزشی با ابزاری مانند آزمون صورت می‌گیرد. یعنی کسب موفقیت در آزمون‌هایی که مبتنی بر ملاک است مستلزم یادگیری هدف‌های آموزشی در سطحی است که برنامه آموزشی بر آن اساس طراحی شده است.



ب) ارزشیابی وابسته به هنجار:

این نوع ارزشیابی به جای یک ملاک از پیش تعیین شده و مطلق بر اساس نوعی ملاک نسبی طراحی شده است. به این صورت که عملکرد فراگیران را با یکدیگر مقایسه می‌کنیم. در این روش می‌توان تعیین کرد که پیشرفت فراگیر نسبت به سایر فراگیران چه وضعی دارد اما نمی‌توان تعیین کرد که پیشرفت یک یادگیرنده نسبت به هدف‌های آموزشی چگونه است. در پایان این نوع آزمون‌ها جایگاه فراگیر را بر حسب رتبه درصدی یا نوعی هنجار سنی یا کلاسی و در مقایسه با گروه هنجار تعیین می‌شود. مانند آزمون‌های دستیاری.

آزمون‌های وابسته به ملاک دارای یک نمره مرزی برای حد تسلط هستند به زبانی دیگر هدف ما این است که فراگیران ما پس از اتمام دوره آموزشی به سطحی از تسلط در اهداف آموزشی رسیده باشند، و وظیفه آزمون‌های وابسته به ملاک نیز دقیقاً تفکیک بین گروه فراگیران مسلط و غیر مسلط است. در همین راستا تحلیل این آزمون‌ها نیز باید بر همین اساس اجرا گردد.

چگونه سؤال های آزمون را تحلیل کنیم؟

هدف از تحلیل سؤالهای آزمون بررسی تک تک سؤالها، تعیین میزان دقت و نارساییهای آنهاست. در تحلیل سؤالهای آزمون، نقاط قوت و ضعف آزمون و کیفیت سؤالهای آن تعیین می شود. لازم است مدرسین پس از اجرای هر آزمون سؤالهای آن را تحلیل کنند و با استفاده از نتایج حاصل به تجدید نظر در آزمون و بهبود کیفیت سؤالها برای استفاده های بعدی اقدام نمایند. برای تهیه یک آزمون خوب تحلیل پاسخهای آزمون شوندگان و بررسی تجربی آنها ضروری است.

مراحل تحلیل سؤال :

چگونگی تعیین ضریب دشواری آزمون :

«تعریف: تعداد کل فراگیران که به یک سوال جواب صحیح می دهند ضریب دشواری آن سوال است»

اگر تعداد کل فراگیران در یک آزمون ۲۰ نفر یا کمتر باشد برگه های آزمون کلاً به دو دسته بالا و پائین تقسیم می شود. اگر تعداد فراگیران بین ۲۰ تا ۴۰ باشد ۱۰ برگه بالا و ۱۰ برگه پائین انتخاب می گردد. در شرایطی که تعداد برگه های آزمون بیش از ۴۰ نفر است بهترین رقم برای گروه بالا و پائین ۲۷٪ کل برگه هاست. (بین ۲۵ تا ۳۳ درصد به عنوان تعداد برگه های گروه بالا و پائین مناسب است) اطلاعات مورد نیاز برای تحلیل سؤالهای یک آزمون پاسخی هستند که آزمون شوندگان به هر سؤال داده اند بنابراین باید تعیین شود که در هر سؤال چند نفر گزینه درست را انتخاب کرده اند، هر یک از گزینه های انحرافی چند نفر را به خود جلب کرده است و چند نفر آن را بی جواب گذاشته اند.

برای تعیین گروه های بالا و پائین، برگه های آزمون را به ترتیب نمره ای که گرفته اند از کوچک به بزرگ مرتب کنید. بعد از بالاترین نمره شروع کنید و تعداد برگه هایی را که می خواهید در گروه بالا قرار دهید انتخاب کنید و به همین ترتیب از پایین ترین نمره شروع کنید و تعداد برگه هایی را که می خواهید در گروه پایین قرار دهید نیز برگزینید. متخصصان آزمون سازی پیشنهاد کرده اند که اگر تعداد کل دانش آموزان که در آزمون شرکت کرده اند بیشتر از ۲۰ و تا ۴۰ نفر هستند ۱۰ برگه بالا و ۱۰ برگه پایین را انتخاب کنید و کاری به بقیه برگه ها نداشته باشید اگر تعداد دانش آموزان ۲۰ نفر یا کمتر است برگه های آزمون را کاملاً به دو دسته بالا و پائین تقسیم بندی کنید. در شرایطی که تعداد کل آزمون شوندگان بیش از ۴۰ نفر است بهترین رقم برای گروه بالا و پائین ۲۷٪ کل برگه هاست این رقم برای شرایطی که نمرات دارای توزیع بهنجار باشد پیشنهاد شده است. اما برای آزمونهای کلاسی هر یک از ارقام ۲۵ تا ۳۳ درصد به عنوان تعداد برگه های گروه بالا یا پائین مناسب است بعد از تهیه گروه های بالا و پائین لازم است برای هر سؤال اطلاعات زیر را تهیه کنید:

۱- تعداد افراد گروه بالا که هر یک از گزینه های سؤال را انتخاب کرده یا آن را بی جواب گذاشته اند.

۲- تعداد افراد گروه پایین که هر یک از گزینه های سؤال را انتخاب کرده یا آن را بی جواب گذاشته اند.

روش محاسبه ضریب دشواری سؤال

درصد کل آزمون شوندگانی که به یک سؤال جواب درست می دهند ضریب دشواری آن سؤال است که با حرف P نشان داده می شود. اگر در تحلیل یک سؤال کلیه افراد یا کلیه برگه های امتحانی دخالت داشته باشند برای محاسبه ضریب دشواری سؤال آن کافی است که تعداد کل افرادی را که به آن سؤال جواب درست داده اند بر تعداد کل آزمون شوندگان

تقسیم کنیم و نتیجه را در ۱۰۰ ضرب نماییم رقم حاصل ضریب دشواری سؤال است مطابق فرمول $P = 100 * R/T$

در این فرمول R معرف تعداد کسانی است که به سؤال جواب درست داده و T نشان دهنده تعداد کل آزمون شوندگان است. به عنوان مثال اگر تعداد پاسخ دهندگان به یک سؤال ۸۰ نفر باشند و از این تعداد ۶۵ نفر به آن سؤال جواب درست

داده باشند ضریب دشواری این سؤال برابر خواهد بود با $81 = 80/85 * 100$ در مواردی که تعداد آزمون شوندگان (تعداد برگه‌های آزمون) زیاد است و اطلاعات ما به نحوی به پاسخدهی افراد گروه بالا و گروه پایین محدود می‌شود لازم است از فرمول زیر استفاده کنیم.

$$P = \frac{\text{انتخابهای درست} + \text{انتخابهای درست گروه بالا}}{\text{تعداد افراد گروه بالا} + \text{تعداد افراد گروه پایین}} \times 100$$

براساس اطلاعات مربوط به سؤال می‌توان نتیجه گرفت که این سؤال نسبتاً دشوار است زیرا حتی نصف گروه انتخابی (بالا و پایین) نتوانسته‌اند به آن جواب درست بدهد براساس توضیحات بالا، هر اندازه ضریب دشواری یک سؤال بزرگ تر (به 100 نزدیکتر) باشد آن سؤال آسان تر است و هر اندازه که این ضریب کوچک تر (به صفر نزدیکتر) باشد سؤال دشوارتر است. بنابراین به جای ضریب دشواری می‌توان از ضریب آسانی یا سهولت نام برد اما معمول این است که به آن ضریب دشواری بگویند.

چنان که ملاحظه شد فرمول بالا ضریب دشواری را برحسب یک عدد صحیح دورقمی نشان می‌دهد. برخی از متخصصان اندازه‌گیری و ارزشیابی ترجیح می‌دهند که این ضریب مانند ضریب تمیز که در قسمت بعد توضیح داده خواهد شد نیز برحسب اعداد اعشاری باشد که تفسیر آن با تفسیر ضریب تمیز همانند صورت گیرد. در این حالت فرمول محاسبه ضریب دشواری به صورت زیر است

نسبت افراد گروه پایین که به سؤال جواب درست داده‌اند + نسبت افراد گروه بالا که به 2 سؤال جواب درست داده‌اند.

چگونگی تفسیر ضریب دشواری :

به طور کلی ضریب دشواری بین 0/3 تا 0/7 حداکثر اطلاعات را درباره تفاوت آزمودنی‌ها به دست می‌دهد. برای آزمون‌های چندگزینه‌ای سطح بهینه دشواری کمی کمتر از فاصله بین 0/1 و سطح موفقیت با حدس زدن است بنابراین در آزمون‌های چهارگزینه‌ای سطح بهینه دشواری 0/6 خواهد بود. یکی از راههای قضاوت درباره مفید بودن سؤالهای یک آزمون آن است که ببینیم سؤال تا چه اندازه به پراکندگی یا واریانس نمرات کمک می‌کند. واریانس نمرات گروهی پاسخ دهنده به یک آزمون از دو قسمت تشکیل می‌یابد:

۱- واریانس سؤالها

۲- همبستگی‌های بین سؤالها

هر سؤالی که دارای واریانس بزرگتری است و با سایر سؤالهای آزمون همبستگی بیشتری دارد به واریانس کل آزمون کمک بیشتری می‌کند در مقابل سؤالی که با سایر سؤالها همبستگی ندارد و دارای واریانس کوچکی است کمک زیادی به واریانس کل آزمون نمی‌کند اگر در تحلیل آزمون به انتخاب سؤالهایی بپردازیم که همبستگی زیادی با یکدیگر دارند و در

ضمن دارای واریانسهای بزرگتری هستند در مجموع آزمون بهتری تدارک دیده‌ایم. $P * (1 - P) =$ واریانس سؤال

در فرمول بالا، P برابر است با درصد کسانی که به سؤال جواب درست داده‌اند یعنی همان ضریب دشواری، یک سؤال زمانی دارای حداقل واریانس است که ضریب دشواری آن 0 یا 1 باشد یعنی هیچ یک از آزمون شوندگان به آن سؤال جواب درست نداده باشد یا اینکه همه آزمون شوندگان به آن جواب درست داده باشند

$$p * (1 - p) = \text{واریانس سؤال} = 0 * (1 - 0) = 0 \quad \text{و} \quad 1 * (1 - 1) = 0 = \text{واریانس سؤال}$$

اما زمانی که ضریب دشواری سؤال $P=0.5$ باشد واریانس سؤال حداکثر خواهد بود. $0.25 = 0.5 * (1 - 0.5) =$ واریانس بهتر است سؤال‌هایی برای آزمون انتخاب شوند که ضریب دشواری آنها از ۱ کمتر، و از صفر بیشتر و به ۰.۵ نزدیک باشد. البته انتخاب سؤال‌هایی با ضرایب دشواری مناسب به نوع سؤال‌های آزمون مربوط است. در آزمون‌های صحیح - غلط ضریب دشواری $p = 0.5$ زمانی بدست می‌آید که همه آزمودنی‌ها را با حدس زدن به سؤال جواب بدهند.

برای آزمون‌های چند گزینه‌ای سطح بهینه دشواری اندکی کمتر از وسط فاصله بین ۱ و سطح موفقیت با حدس زدن است. منظور از سطح موفقیت با حدس زدن عبارت است از نسبت پاسخ‌های درستی که با حدس زدن به دست می‌آید. بنابراین برای سؤال‌های چهارگزینه‌ای سطح موفقیت با حدس زدن حدود ۰.۲۵ است. لذا سطح بهینه دشواری در وسط ۰.۲۵ و ۱ یعنی در حدود ۰.۶ خواهد بود. برای تعیین رقم دقیق مصرف ضریب دشواری بهینه سؤال ابتدا سطح موفقیت ۱۰٪ را از سطح عملکرد و با حدس زدن (مثلاً ۰.۲۵٪) کم می‌کنیم و نتیجه را به ۲ تقسیم می‌نماییم و نقطه میانی را مشخص می‌کنیم. ضریب دشواری بهینه = عملکرد با حدس زدن + نقطه میانی

راه ساده‌تر برای بدست آوردن همان نتیجه این است که عدد ۱ را با عملکرد مبتنی بر حدس زدن جمع کنیم و نتیجه را بر ۲ تقسیم نماییم. ضریب‌های دشواری بین ۰.۳ تا ۰.۷ حداکثر اطلاع را درباره تفاوت بین آزمودنیها بدست می‌دهد.

چگونگی تعیین ضریب تمیز سؤال

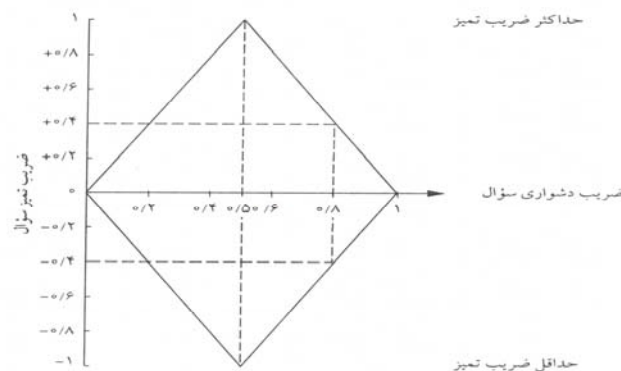
« تعریف: قدرت سؤال را در تمایزگذاری یا تشخیص بین گروه قوی و گروه ضعیف آزمون شوندگان مشخص می‌کند. ضریب تمیز با d نشان می‌دهند. قدرت سؤال را در تمایزگذاری یا تشخیص بین گروه قوی و گروه ضعیف آزمون شوندگان مشخص می‌کند و یعنی معلوم می‌نمایند که سؤال تا چه اندازه می‌تواند گروه قوی را از گروه ضعیف جدا سازد. برای محاسبه ضریب تمیز یک سؤال از فرمول زیر استفاده می‌شود.

انتخاب‌های درست گروه پایین - انتخاب‌های درست گروه بالا

$d =$ ضریب تمیز سؤال =

تعداد افراد یک گروه (بالا یا پایین)

هرقدر ضریب تمیز بزرگ‌تر باشد، قوه تمیز آن سؤال بیشتر و هر قدر این ضریب کوچک‌تر باشد قوه تمیز آن کمتر است. مثلاً اگر قوه تمیز سوالی ۰/۹۰ باشد آن سؤال آزمون شوندگان قوی و ضعیف را خیلی خوب از هم جدا خواهد کرد، اما اگر ضریب تمیز سوالی ۰/۱۰ باشد آن سؤال از عهده جداسازی دانش آموزان قوی و ضعیف به خوبی بر نخواهد آمد. ضریب تمیز صفر حاکی از این است که آن سؤال به هیچ وجه نتوانسته بین گروه قوی و ضعیف تمایز قائل شود.



شکل ۱۳-۳ رابطه ضریب تمیز با ضریب دشواری

روش محاسبه ضریب تمیز در آزمون های وابسته به ملاک :

a: تعداد آزمون شوندگان که سوال را درست پاسخ داده‌اند اما نمره کمتر از حد تسلط گرفته‌اند.
b: تعداد آزمون شوندگان که سوال را درست جواب داده‌اند و در آزمون نیز به حد تسلط رسیده‌اند.
c: تعداد آزمون شوندگان که سوال را غلط جواب داده‌اند و در کل به حد تسلط نرسیده‌اند.
d: تعداد آزمون شوندگان که سوال را غلط جواب داده‌اند و در آزمون به حد تسلط رسیده‌اند.
تفسیر ضریب تمیز بدست آمده از این روش همانند شیوه محاسبه قبلی است. در آزمون های وابسته به ملاک ضریب تمیز صفر (زمانی که کلیه فراگیران به سوال پاسخ صحیح داده‌اند) نشان دهنده کاستی در سوال نیست چرا که ممکن است بر اساس اهداف آموزش کلیه فراگیران باید به این سوال پاسخ صحیح بدهند.
بنابراین در آزمون های وابسته به ملاک تفسیر ضریب دشواری و تمیز باید توسط مدرس و با توجه به اهداف آموزشی صورت گیرد.

چگونگی تفسیر ضریب تمیز :

هر قدر ضریب تمیز سوال بزرگتر باشد قوه تمیز آن بیشتر و هر قدر این ضریب کوچکتر باشد قوه تمیز آن کمتر است. مثلاً اگر قوه تمیز سؤال ۹۰٪ باشد آن سؤال آزمون شوندگان قوی و ضعیف و خیلی خوب از هم جدا خواهد کرد. اما اگر ضریب تمیز سؤال ۰.۱۰ باشد آن سؤال از عهده جداسازی دانشجویان قوی و ضعیف به خوبی بر نخواهد آمد. ضریب تمیز صفر حاکی از این است که آن سؤال به هیچ وجه نتوانسته بین گروه قوی و ضعیف تمایز قائل شود گاه اتفاق می افتد که ضریب تمیز سئوالی منفی است این نوع ضریب تمیز نشان می دهد که در آن سؤال گروه قوی ضعیفتر از گروه ضعیف عمل کرده است . و گروه پایین در کلاس آن را بهتر پاسخ داده‌اند. بدیهی است این سوال اشکال اساسی دارد و باید به کلی کنار گذاشته شود و یا در آن تجدید نظر اساسی به عمل آید. اگر سؤال دارای اشکال فنی نباشد دلیل چنین ضریبی می تواند این باشد که افراد گروه بالا یا آن سؤال را به طور کامل یاد نگرفته‌اند یا آن را به غلط آموخته‌اند.

تحلیل سوال های آزمون های وابسته به ملاک :

نتیجه آرمانی آزمون های وابسته به ملاک این است که در آنها اکثر آزمون شوندگان اکثر سوال ها را درست جواب دهند. در این حالت ضریب دشواری بزرگ و ضریب تمیز بسیار کوچک خواهد بود.

تحلیل گزینه های انحرافی :

برای دستیابی به اطلاعات لازم در مورد سوال های آزمون ، علاوه بر تعیین ضریب دشواری و تمیز برای هر سوال می بایست نحوه پراکندگی پاسخ ها مربوط به گزینه های انحرافی نیز بررسی گردد.
گزینه ای به عنوان گزینه انحرافی مناسب است که تعداد افرادی را که از گروه ضعیف به خود جلب می کند بیش از گروه قوی باشد .

تجدید نظر و اصلاح آزمون :

زمانی که ضریب دشواری و ضریب تمیز سوال های یک آزمون تعیین شد باید به اصلاح سوالاتی که نیاز به تغییر دارند اقدام نمود . سوال هایی که ضریب دشواری پایین یا بالا دارند و یا ضریب تمیز آنها خیلی کم است باید مورد تجدید نظر قرار گیرند. این موارد شامل گزینه های انحرافی معیوب نیز می شود که باید اصلاح یا عوض شوند.

منحنی ویژگی سوال :

یکی از روش هایی دقیق برای تحلیل سوال آزمون مخصوصاً در آزمون های وابسته به ملاک است. در این روش ویژگی های سوال را به صورت نمودار نشان می دهند منحنی ویژگی سوال که برای هر یک از سوال های آزمون رسم می شود باز نمایی

از رابطه بین احتمال پاسخ درست دادن به ۸ سوال و موفقیت آزمون شونده در صفت مورد اندازه گیری توسط آزمون است. برای تهیه این منحنی نسبت یا درصد آزمون شوندگانی که آن سوال را درست جواب داده اند در برابر نمره کل آزمون آنها رسم می شود. به سخن دیگر محور افقی نمره کل آزمون و بر روی محور عمودی نسبت آزمون شوندگان که به سوال پاسخ درست داده اند مشخص می شود.

از روی منحنی ویژگی سوال می توان ضرایب دشواری و تمیز سوال را تعیین نمود. ضریب دشواری عبارت است از نمره معیاری که در آن ۵۰ درصد آزمون شوندگان سوال را درست پاسخ داده اند و ضریب تمیز عبارتند از شیب منحنی ویژگی سوال.

در این نمودار افزایش تدریجی شیب منحنی نشانگر آن است که با افزایش نمرات آزمون درصد کسانی که به سوال پاسخ صحیح داده اند افزایش می یابد.

این روش تحلیل سوال این خصوصیت را دارد که به دقت و تفکیک معین می کند که سوال در چه محدوده ای، از قابلیت تفکیک مناسبی برخوردار است. زمانی که استاد، برگزار کننده آزمونی با معیار مشخصی است می بایست قابلیت تفکیک و تمیز آزمون در محدود معیار مشخص شده، بیشترین میزان باشد تا بتواند فراگیران بالاتر از معیار و کمتر را، از هم تمیز دهد.

منحنی ویژگی یک سوال خوب آن نوع منحنی است که دارای شیب مثبت و ضریب دشواری متوسط باشد. اگر قرار است از یک آزمون برای انتخاب داوطلبانی که در بالا یا پائین یک نمره معین (معیار) قرار دارند استفاده کنیم. سطح دشواری باید برابر با همان نمره باشد و منحنی ویژگی سوال نیز در نمره معین نشان دهنده قدرت تمیز بالا باشد. اگر بخواهیم از میان آزمون شوندگان دارای نمرات بالا دست به انتخاب بزنیم شیب منحنی باید در قسمت انتهایی آن بیشتر باشد.

پایایی آزمون :

آزمونی که در پایان دوره تحصیلی میزان یادگیری فراگیران را می سنجد باید پایا باشد. به این معنی که اگر این آزمون را در فاصله کوتاهی چندین بار به گروه واحدی از افراد بدهیم نتایج حاصل نزدیک به هم باشد. اگر آزمون در هر بار اجرا نتایج متفاوتی بدهد آن آزمون پایا نخواهد بود و در واقع چیزی را به درستی اندازه نخواهد گرفت و در این حالت آزمون اطلاعات مفیدی را به ما نخواهد داد. روش های متعدد محاسبه پایایی وجود دارد اما با توجه به امکانات بهترین روش برای اندازه گیری پایایی یک آزمون روش های ذیل است :

۱- آزمونی روش گودرریچاردسون : در این روش آزمون یک بار اجرا می شود اما در این روش متوسط ضریب پایایی آزمون بر اساس ضریب همبستگی در نیمه های آن از طریق همه روش های ممکن دو نیمه کردن بدست می آید. عدد بدست آمده پایایی و همسانی درونی محتوای آزمون را نشان می دهد .

۲- روش دو نیمه کردن آزمون (اسپلیت هاف) : آزمون یک بار با گروه واحدی از آزمون شوندگان اجرا می شود و سپس آن را به دو نیمه تقسیم می کنیم. بهترین روش دو نیمه کردن آزمون بصورت جدا نمودن سوالات فرد و زوج است. برای محاسبه ضریب پایایی در این روش باید همبستگی بین نیمه های آزمون را بدست آوریم.

۳- روش ضریب آلفا : این روش زمانی از روش گودرریچاردسون متفاوت خواهد بود که آزمون از تعدادی خرده آزمون تشکیل شده باشد.