



راهنمای طراحی سوالات: آزمون چند گزینه‌ای (MCQ)



آزمون چندگزینه‌ای (MCQ) multiple choice question



آزمون‌های با سوالات چندگزینه‌ای از رایج‌ترین آزمون‌های آموزشی در ارزشیابی دانشجویان هستند. اکثر آزمون‌های استاندارد که برای گزینش دانشجویان در مقاطع مختلف برگزار می‌شود (مثلاً کارشناسی، کارشناسی ارشد و دکترا) نیز از نوع چندگزینه‌ای هستند. در این آزمون، سوال به همراه چند گزینه به آزمون‌شونده داده می‌شود و او باید پاسخ درست را از بین آنها انتخاب نماید. سوالات چندگزینه‌ای بیشتر اهداف آموزشی حیطه شناختی را مورد ارزیابی قرار می‌دهد. اگر این سوالات خوب طرح شوند می‌توانند سطوح دانش، فهم و کاربرد را ارزشیابی نمایند. معروف‌ترین آزمون چندگزینه‌ای، آزمون چهارگزینه‌ای است. هرچند تعداد گزینه‌ها می‌تواند کمتر و یا بیشتر از ۴ گزینه نیز باشد. برای دانش‌آموزان مقاطع پائین مثلاً دوم یا سوم ابتدایی، می‌توان از سوال‌های دارای گزینه‌های کمتر استفاده کرد. سوالات ۴ یا ۵ گزینه‌ای برای دانش‌آموزان مقاطع بالاتر و دانشجویان مناسب است.

مثال: بالاترین سطح حیطه شناختی کدام مورد است؟

الف) فهم (ب) کاربرد (ج) ترکیب (د) ارزشیابی (و) تحلیل

قسمت‌های مختلف آزمون‌های چندگزینه‌ای عبارتند از:

۱- قسمت اصلی یا تنه سوال: شامل متن اصلی سوال است.

۲- گزینه درست (کلید): پاسخ درست سوال است که آزمون‌شونده در بین گزینه‌ها باید آن را انتخاب نماید.

۳- گزینه‌های انحرافی: سایر گزینه‌ها (بجز پاسخ درست) را گزینه‌های انحرافی می‌گویند که برای منحرف کردن آزمون‌شوندگانی که پاسخ سوال را نمی‌دانند، طراحی می‌گردد.

محاسن آزمون‌های چندگزینه‌ای

۱- تصحیح پاسخ آزمون‌ها ساده است.

۲- نمره تحت تاثیر نمره‌گذار یا مصحح قرار نمی‌گیرد

۳- با استفاده از این آزمون در یک جلسه و یک زمان محدود، تعداد زیادی از هدف‌های آموزشی را می‌توان اندازه‌گیری کرد.

۳- عامل حدس و گمان در این آزمون نسبت به آزمون‌های دیگر عینی، کمتر است.

۴- دانش‌آموزان و معلمان این نوع آزمون‌ها را به سایر آزمون‌های عینی ترجیح می‌دهند.

معایب آزمون‌های چندگزینه‌ای

- ۱- ساختن این آزمون مشکل است.
- ۲- سطوح پایین یادگیری را می‌سنجد.
- ۳- آزمودنی آزادی عمل ندارد.
- ۴- شانس و اقبال و حدس کورکورانه در آن دخالت دارد.
- ۵- خواندن پرسش‌ها و گزینه‌ها نسبت به آزمون صحیح-غلط، وقت‌گیرتر است.

قواعد تهیه سوال‌های چندگزینه‌ای

- ۱- هر سؤال باید موضوع یا هدف آموزشی را در برگیرد.
- ۲- بیشتر از یک مسأله یا یک مطلب در هر سؤال قرار ندهید.
- ۳- سؤالات را کاملاً روشن و در حد درک آزمون شوندگان بنویسید.
- ۴- از به کار بردن اشاره‌های دستوری نامربوط خودداری کنید.
سوال ضعیف: اسناد املاک مردم در کجا ثبت می‌شوند؟
(الف) اداره کل ثبت اسناد و املاک
(ب) اداره کل ثبت احوال
(ج) وزارت امور اقتصاد و دارایی
(د) وزارت علوم و آموزش عالی
- ۵- سؤال را طوری بنویسید که پاسخ درست تنها پاسخ درست یا درست‌ترین پاسخ باشد.
سوال ضعیف: یک پسر بچه کوررنگ، کوررنگی خود را از چه کسی به ارث می‌برد؟
(الف) پدر (ب) مادر* (ج) پدر مادر (د) پدر پدر (ه) نیاکان دور
سوال بهتر: یک پسر بچه کوررنگ، به احتمال بسیار زیاد کوررنگی خود را از چه کسی به ارث می‌برد؟
(الف) پدر (ب) مادر (ج) پدر مادر (د) پدر پدر (ه) نیاکان دور
- ۶- در سؤالی که تنه جمله ناتمام است، بهتر است جای خالی در آخر جمله باشد.
سوال ضعیف: طبق نظر دیبالکی..... عامل اصلی بیماری قلبی است.
(الف) کلسترول* (ب) چاقی (ج) سیگار کشیدن (د) استرس
سوال بهتر: طبق نظر دیبالکی عامل اصلی بیماری قلبی است.
(الف) کلسترول* (ب) چاقی (ج) سیگار کشیدن (د) استرس
- ۷- بهتر است در تنه سوال از جمله پرسشی استفاده شود تا جمله ناتمام
- ۸- از تکرار مطالب در گزینه‌ها خودداری کنید.
سوال ضعیف: در آزمونهای عینی
(الف) به روش مشخص کردن نتایج یادگیری مربوط است.
(ب) به روش انتخاب محتوای آزمون مربوط است.
(ج) به روش تصحیح جوابها مربوط است.*
(د) به روش ارائه مطالب مربوط است.

سوال بهتر: در آزمونهای عینی، اصطلاح عینی به روش مربوط است.

الف) مشخص کردن نتایج یادگیری

ب) انتخاب محتوای آزمون

ج) تصحیح جوابها *

د) ارائه مطالب

۹- گزینه ها از نظر دستوری و جمله بندی مکمل ساقه سوال باشند.

سوال ضعیف: نزدیک به ۷۰ درصد جمعیت مردم یک مملکت افرادی هستند که هوشبهر آنان

الف) بین ۱۰۰ تا ۱۵۰ است.

ب) بین ۵۰ تا ۱۰۰ است

ج) بین ۸۵ تا ۱۱۵ است. *

د) هیچکدام درست نیست.

۱۰- همه گزینه‌های یک سؤال باید متجانس و به موضوع واحدی مربوط باشد.

سوال ضعیف: کدام یک از موارد زیر درست است؟

الف) سوالهای انشایی بهترین نوع سوال برای ارزشیابی هدفهای طبقه آفریدن هستند. *

ب) تهیه جدول مشخصات برای آزمون های انشایی مفید نیست.

ج) سوالهای عینی را تنها در کلاسهای دانشگاه می توان به اجرا درآورد.

د) هدفهای رفتاری از دیرباز مورد استفاده معلمان قرار داشته اند.

سوال بهتر: در رابطه با استفاده از سوالهای انشایی و ارزشیابی هدفهای مختلف آموزشی، کدام مورد زیر درست است؟

الف) بهترین نوع سوال برای ارزشیابی هدفهای طبقه آفریدن هستند. *

ب) برای ارزشیابی هدفهای هیچ یک از طبقات حوزه شناختی مفید نیستند.

ج) بین سوالهای انشایی و عینی از لحاظ ارزشیابی هدفهای گوناگون چندان تفاوتی وجود ندارد.

د) بهتر است از سوالهای انشایی به ندرت استفاده شود.

۱۱- از نوشتن سؤالاتی که هم متن سؤال منفی و هم پاسخ منفی دارد (منفی مضاعف)، خودداری کنید.

سوال ضعیف: در رابطه با استفاده از سوالهای انشایی در ارزشیابی هدفهای آموزشی نمی توان گفت که...

الف) سوالهای انشایی برای ارزشیابی هدفهای طبقه ترکیب بهترین نوع نیستند. *

ب) سوالهای انشایی برای ارزشیابی همه طبقات حوزه شناختی مفید نیستند.

ج) سوالهای انشایی و عینی تفاوت زیادی با یکدیگر ندارند.

د) استفاده از سوالهای عینی برای هدفهای دانش همیشه صلاح نیست.

سوال بهتر: در رابطه با استفاده از سوالهای انشایی و ارزشیابی هدفهای مختلف آموزشی، کدام مورد زیر درست است؟

الف) بهترین نوع سوال برای ارزشیابی هدفهای طبقه آفریدن هستند. *

ب) برای ارزشیابی هدفهای هیچ یک از طبقات حوزه شناختی مفید نیستند.

- (ج) بین سوالهای انشایی و عینی از لحاظ ارزشیابی هدفهای گوناگون چندان تفاوتی وجود ندارد.
- (د) بهتر است از سوالهای انشایی به ندرت استفاده شود.
- ۱۲- حتی الامکان از جملات منفی استفاده نکنید و در صورت ضرورت زیر کلمه منفی خط بکشید.
- ۱۳- از طرح سؤالهای گمراه کننده بپرهیزید.
- سوال گمراه کننده: با توجه به اینکه نیروی جاذبه زمین شش برابر نیروی جاذبه کره ماه است، اگر جرم جسمی در روی زمین ۶۰ کیلوگرم باشد، جرم این جسم در کره ماه چند کیلوگرم است؟
- الف) ۰/۶ (ب) ۶ (ج) ۱۰ (د) ۶۰ *
- سوال بهتر: اگر جرم جسمی در روی زمین ۶۰ کیلوگرم باشد، جرم این جسم در کره ماه چند کیلوگرم است؟
- الف) ۰/۶ (ب) ۶ (ج) ۱۰ (د) ۶۰ *
- ۱۴- دو گزینه متضاد که یکی از آنها درست است، بکار نبرید.
- سوال ضعیف: وقتی که شباهت بین محرک غیرشرطی و محرک شرطی افزایش می یابد،
- الف) تعمیم بیشتر اتفاق می افتد، اما تمیز دشوارتر می شود.*
- ب) تمیز بیشتر اتفاق می افتد، اما تعمیم دشوارتر می شود.
- ج) تعمیم و تمیز کمتر اتفاق می افتد.
- د) اتفاق تازه ای رخ نمی دهد.
- سوال بهتر: وقتی که شباهت بین محرک غیرشرطی و محرک شرطی افزایش می یابد،
- الف) تعمیم بیشتر اتفاق می افتد، اما تمیز دشوارتر می شود.*
- ب) تمیز بیشتر اتفاق می افتد، اما تعمیم دشوارتر می شود.
- ج) تعمیم و تمیز کمتر اتفاق می افتد.
- د) تعمیم و تمیز بیشتر اتفاق می افتد.
- ۱۵- برای هر سؤال بین ۳ تا ۵ گزینه در نظر بگیرید.
- ۱۶- تا حد ممکن از عبارات «همه موارد» یا «هیچکدام» استفاده نشود.
- ۱۷- گزینه‌های انحرافی، با پاسخ صحیح متجانس باشند تا توجه فراگیران بی اطلاع را به خود جلب کنند.
- سوال ضعیف: امیرکبیر صدراعظم کدام یک از شاهان زیر بود؟
- الف) شاه عباس (ب) سلطان محمود (ج) ناصرالدین شاه * (د) مظفرالدین شاه
- سوال بهتر: امیرکبیر صدراعظم کدام یک از شاهان زیر بود؟
- الف) احمدشاه (ب) محمدعلی شاه (ج) ناصرالدین شاه * (د) مظفرالدین شاه
- ۱۸- طول گزینه‌های صحیح و غلط تقریباً هم اندازه انتخاب شود.
- ۱۹- محل گزینه درست به صورت تصادفی انتخاب شود.
- ۲۰- تنها در صورت لزوم از آزمونهای چندگزینه‌ای استفاده کنید.

تحلیل سوال های چندگزینه ای



سوال های چندگزینه ای یکی از متداول ترین روش های ارزشیابی فراگیران است. هرچند استفاده از این نوع آزمون بالاخص برای بررسی سطح دانش و فهم گروه بزرگی از فراگیران مفید است اما عدم آشنایی با اصول و قواعد طراحی سوالات تاثیر این نوع آزمون را کاهش می دهد. بنابراین آنالیز سوالات اهمیت خاصی دارد. در علوم پزشکی برای ارتقای کیفیت طراحی سوالات معمولا از دو منظر تاکسونومی و اصول ساختاری استفاده می شود.

۱- تاکسونومی

در حیطه شناختی از دانشجو انتظار می رود که مطالب آموزش داده شده را بازشناسی کند، یا مشکل یا مساله را تشخیص داده و راه حل آن را بیان نماید. از این دید، سه سطح از سطوح شش گانه بلوم قابلیت کاربرد در سوالات چند گزینه ای دارند و می توان آنها را به سه دسته تاکسونومی ۱ (دانش)، ۲ (درک و فهم) و ۳ (کاربرد) تقسیم کرد:

تاکسونومی ۱: سؤال های سطح دانش. این سطح فقط بر قدرت حافظه فراگیر متکی می باشد. مانند یادآوری مفاهیم، اصول، حقایق و ... بنابراین برای شناخت یادگیری واقعی دانشجویان مناسب نمی باشد. مثال: علائم شایع بیماری ام اس چیست؟

تاکسونومی ۲: سؤال های مبتنی بر درک و فهم. پاسخ به این سوال ها نیازمند درک معنا و مفهوم یک مطلب و بیان آن با عبارات و جملات خود شخص است. مثلا: وقتی در سؤال یک Case مطرح شده و دانشجو باید تشخیص خود را بنویسد.

تاکسونومی ۳: سؤال های مبتنی بر کاربرد دانش. پاسخ به این سؤال ها به جز حافظه به سطوح بالاتر یادگیری مثل کاربرد، تجزیه و تحلیل و حتی ترکیب و ارزشیابی نیازمند است. طرح سوال های چندگزینه ای از نوع تاکسونومی ۳ تا حدی مشکل است. به عنوان مثال در سؤال یک Case مطرح می شود و درمان مورد پرسش قرار می گیرد. در این حالت فراگیر باید ابتدا بهترین تشخیص را پیدا کرده و سپس درمان مناسب را بیان کند.

سوالات منتخب از نظر سطح شناختی بلوم، الزاماً باید در تاکسونومی ۱، ۲ و ۳ طرح شده باشند و به عنوان یک قاعده کلی می‌بایست حداکثر نصف سوال‌های امتحان جزو سوال‌های تاکسونومی ۱ باشد.

۲- فهرست واریسی میلمن (Millman checklist)

روش‌های مختلفی برای طراحی سوالات چهارگزینه‌ای وجود دارد. یکی از بهترین شیوه‌ها، روش میلمن است که در قالب فهرست واریسی تهیه می‌شود. فهرست واریسی میلمن جهت بررسی رعایت اصول ساختاری در سوالات چندگزینه‌ای طراحی شده است و به طراح سوالات کمک می‌کند که سوالات مناسبی تهیه نماید. این چک‌لیست دارای ۱۲ سوال بوده و بر اساس طیف دو گزینه‌ای (بلی / خیر) سوالات را بررسی می‌کند. برخی از سوالات این شیوه در جدول ذیل بیان شده است.

چک لیست میلمن - آنالیز کیفی سوالات چندگزینه‌ای

خیر ۰	بله ۱	فهرست واریسی
		۱- آیا سؤال به روشنی موضوع مورد نظر را بیان می‌کند؟
		۲- موضوع اصلی سؤال را در تنه سؤال نوشته‌اید؟
		۳- مطالب تکراری گزینه‌ها را در تنه سؤال قرار داده‌اید؟
		۴- سؤال یک موضوع مهم را اندازه می‌گیرد؟
		۵- گزینه‌های سؤال با هم تجانس دارند؟
		۶- در سؤال فقط یک گزینه درست یا یک گزینه که از همه بهتر است وجود دارد؟
		۷- در سؤالی که برای سنجش فرآیندهای پیچیده ذهنی نوشته‌اید موقعیت‌های جدید را مطرح کرده‌اید؟
		۸- گزینه‌های انحرافی آزمون شوندگان بی‌اطلاع از موضوع را به خود جلب می‌کند؟
		۹- گزینه‌های سؤال از نظر دستوری و جمله بندی با تنه سوال مطابقت دارد؟
		۱۰- از کاربرد اشاره‌های نامربوط خودداری کرده‌اید؟
		۱۱- در سؤال منفی زیر عبارات منفی خط کشیده‌اید؟
		۱۲- از کاربرد منفی مضاعف در نوشتن سؤال‌ها خودداری کرده‌اید؟
		۱۳- از کاربرد عباراتی چون: همه آنها، همه آنچه در بالا گفته شد، خودداری کرده‌اید؟
		۱۴- سؤال مستقل از سوال‌های دیگر است؟
		۱۵- آیا از طرح سؤال به شکل گمراه کننده (که شکل و جواب مختلفی دارند) اجتناب کرده‌اید؟
		۱۶- از کاربرد دو گزینه متضاد که یکی از آنها درست است خودداری کرده‌اید؟
		۱۷- از کاربرد دو گزینه هم معنی خودداری کرده‌اید؟
		۱۸- محل گزینه درست را در میان گزینه‌های انحرافی به طور تصادفی انتخاب کرده‌اید؟
		۱۹- آیا طول گزینه درست با سایر گزینه‌ها متفاوت است؟
		۲۰- سؤال در حد درک و فهم آزمون شوندگان هست؟
		۲۱- گزینه، هیچیک از موارد فوق، را مورد استفاده قرار نداده‌اید؟
		۲۲- جای خالی را در انتهای تنه سؤال قرار داده‌اید؟
		جمع امتیازات
		درصد امتیازات

ضریب تمیز و دشواری

ضریب تمیز (Discriminative Index): بیانگر آنست که سوال مزبور تا چه اندازه توانسته است دانشجویان قوی را از ضعیف متمایز نماید (دامنه تغییرات : +1 تا -1) به عبارت دیگر در حالت مطلوب ضریب تمیز سوالی مثبت می باشد که دانشجویان با نمره بالاتر، بیشتر به این سوال پاسخ داده اند اما اگر ضریب تمیز صفر یا منفی باشد یعنی دانشجویان با نمره بالا و پایین به یک اندازه به سوال پاسخ داده اند ($DI=0$)، و یا دانشجویان ضعیف بیشتر به سوال پاسخ داده اند ($DI<0$)، که در این صورت سوال مطلوب نخواهد بود. هر چه تعداد داوطلبان قوی که به سوال پاسخ صحیح داده اند بیشتر از داوطلبان ضعیفی باشد که به سوال پاسخ درست داده اند این ضریب عددی بزرگتر خواهد شد. به هر سوال باید بیشتر داوطلبانی پاسخ صحیح بدهند که در کل آزمون نمره بیشتری آورده اند.

ضریب دشواری (Index difficulty): ضریب دشواری هر سوال تعداد افرادی است که به سوال پاسخ صحیح داده اند و با حرف p نشان داده می شود. به عنوان مثال هرگاه ۶۷٪ شرکت کنندگان در آزمون به سوالی پاسخ صحیح داده باشند ضریب دشواری سوال ۶۷٪ خواهد بود. هرچه سوال ساده تر باشد (افراد بیشتری بدان پاسخ داده باشند) ضریب دشواری بالاتری خواهد داشت: نکته ای که قابل ذکر است این است که ضریب دشواری سوال، خصوصیتی وابسته به آزمودنی است. سوالی که در یک آزمون ضریب دشواری ۵۵٪ دارد ممکن است در آزمونی دیگر (که شرکت کنندگان دیگری دارد) ضریب دشواری ۴۰٪ را داشته باشد.

رابطه بین ضریب دشواری و ضریب تمیز

بین ضریب دشواری و ضریب تمیز رابطه معکوس وجود دارد. هر چه ضریب دشواری به ۱ نزدیک تر شود، قدرت تمیز کوچک تر می شود. به عبارت دیگر، هر چه سوال آسان تر باشد، قدرت تمیز دو گروه کمتر است. در سوال هایی با ضریب دشواری ۰/۵، ضریب تمیز به حداکثر قدرت خود می رسد. بنابراین، سوال های خوب، سوال هایی هستند که دارای ضریب دشواری ۰/۵ و ضریب تمیز بالا باشند.

منابع:

- ۱- جلیلی، محمد. خباز مافی نژاد، محبوبه. گندم کار، محبوبه. مرتاض هجری، سارا. ۱۳۹۶. اصول و روش های ارزیابی فراگیران در علوم پزشکی. فرهنگستان علوم پزشکی جمهوری اسلامی ایران.
- ۲- حسن زاده، رمضان. مداح، محمدتقی. ۱۳۹۵. سنجش و اندازه گیری در علوم رفتاری. تهران: نشر روان.
- ۳- سیف، علی اکبر. ۱۳۹۷. اندازه گیری، سنجش و ارزشیابی آموزشی. تهران: دوران.