

بین‌رشته‌هایی از المپیاد جهانی زیست‌شناسی



هانس مورلیس

رئیس هیئت داوران بین‌المللی المپیاد جهانی زیست‌شناسی

محمد کرام‌الدینی

عضو هیئت داوران بین‌المللی المپیاد جهانی زیست‌شناسی از سال ۱۹۹۸ تاکنون

کمک به فراسازی آزمون‌های مناسب‌تر در زمینه‌ی زیست‌شناسی و ارزشیابی از پیشرفت تحصیلی دانش‌آموزان کشورهای عضو کمابیش راه‌گشا و مفید خواهد بود.

یکی از مهم‌ترین وظایف دبیران زیست‌شناسی طرح سؤال و تهیه‌ی آزمون است. فرض بر این است که نکاتی که در این جا آورده می‌شود، برای تهیه‌ی آزمون‌های استاندارد و مناسب به کار همکاران خواهند آمد.

بیست سال تجربه

در تابستان امسال، المپیاد جهانی زیست‌شناسی نوزدهمین سال تولد خود را در بمبئی هندوستان به پایان رساند و پای به بیستمین سال حیات خود گذاشت، در حالی که در این مدت، در حیات پر فراز و نشیب خود بیش از پنجاه کشور را به عضویت خود درآورده

سال‌ها داوری در هیأت داوران بین‌المللی المپیاد جهانی زیست‌شناسی باید برای حلاجی نقطه‌های قوت و ضعف آن کافی باشد. علت پدید آمدن این نوشته نیز همین است، یعنی تصمیم گرفته‌ایم برخی از درس‌ها و نکته‌هایی را که در این مدت از المپیاد جهانی زیست‌شناسی آموخته‌ایم در این جا به طور خلاصه مرور کنیم، تا همان طور که در گذشته قول داده‌ایم، بتوانیم تجربه‌های المپیاد جهانی زیست‌شناسی را برای بهبود آموزش زیست‌شناسی کشورهای عضو به کار بگیریم.

می‌دانید که المپیاد جهانی زیست‌شناسی چیزی چندان فراتر از یک آزمون نیست، منتهی آزمونی که در سطح بین‌المللی انجام می‌شود و لذا می‌کوشد خود را به عالی‌ترین سطح برساند، تا علاوه بر اعتبار بخشیدن به کار خود، الگویی باشد برای کشورهای عضو. بنابراین، تحلیل نقاط قوت و ضعف آن به منظور انتقال تجربه و

است .

حیات المپیاد جهانی زیست شناسی را از آن جهت پرفراز و نشیب دانستم که سطح علمی- آموزشی آن از سالی به سال دیگر بسیار متفاوت بوده است . ممکن است کیفیت آزمون در یک سال خوب و مناسب و در نتیجه در فراز باشد ، اما در سال بعد به شدت افت کند و به نشیب در غلطد .

بالا یا پایین بودن کیفیت آزمون المپیاد جهانی زیست شناسی را گروه آزمون سازی کشور میزبان تعیین می کند . یعنی اگر کمیته ی برگزارکننده ی کشور میزبان گروهی قوی ، کارآزموده و خلاصه خیره را مأمور تهیه ی آزمون کند که برخی اعضای آن در موضوع های مختلف علم زیست شناسی تبحر کافی دارند و برخی دیگر متخصصان آموزش و ارزشیابی هستند ، طبیعی است که نتیجه ی مطلوب به دست می آید و به عکس .

اول : حتی در سطح بین المللی هم تهیه ی یک آزمون استاندارد ممکن نبوده است

اگر در پی آزمونی استاندارد و بی نقص هستیم تا از آن الگو بگیریم ، نباید به یافتن آن امید چندان داشته باشیم . آدمی هنوز در هیچ یک از آزمون های ملی یا بین المللی ، حتی آزمون های المپیاد جهانی زیست شناسی به مجموعه ای بی نقص دست نیافته است . تنها چیزی که در جست و جوی چنین آزمونی نصیب مان می شود ، امکان مقایسه ی آزمون ها با هم و ارائه ی استدلال هایی برای مناسب تر بودن یا نامناسب تر بودن آن هاست .

در مجموع ، در یک برآورد کلی می توان گفت که تاکنون بسیاری از سؤال های آزمون نظری المپیاد جهانی زیست شناسی برخلاف قواعد و اصول مصوب ، نه تنها بر ارزشیابی از نیروی استدلال دانش آموزان تکیه نداشته اند ، بلکه بیش تر دانش محور بوده اند و دانسته ها را ارزشیابی می کرده اند .

البته ، برخی از داوران از سؤال های دانش محور راضی اند ، از این وضعیت دفاع می کنند و دلیل می آورند که لازم است دانش شرکت کنندگان را از مفاهیم مهم ارزشیابی کنیم . اما اینان غافل اند که اساس این آزمون جهانی بر ارزشیابی از نیروی استدلال و تفکر خلاقانه است و اگرچه ما نیز معتقدیم که دانستن مفاهیم مهم لازم است ، اما باور داریم که باید آن ها را در زمینه های استدلالی ارزشیابی کرد . به علاوه ، مطابق با اصول و قواعد این المپیاد حداکثر فقط ۲۰ درصد سؤال ها را می توان از انواع سؤال های دانش محور انتخاب کرد . نمونه ای از پرسش های دانش محور :

● **دایناسورها چه زمانی از بین رفتند؟**

الف . حدود ۲ میلیون سال پیش

ب . حدود ۶۵ میلیون سال پیش

ج . حدود ۲۰۰ میلیون سال پیش

د . حدود ۴۰۰ میلیون سال پیش

در این پرسش دانش آموز باید فقط با یادآوری اعداد و ارقام ، گزینه ی درست را مشخص کند و اثری از ارزشیابی از نوآوری و استدلال در آن موجود نیست .

دوم : بهتر است پرسش های آسان را حذف کنیم

در آزمون های رقابتی که در آن ها تعدادی دانش آموز برای به دست آوردن رتبه های برتر به رقابت می پردازند ، پرسش های آسانی که همه ی دانش آموزان پاسخ درست بدان ها می دهند ، بی ارزش به شمار می آیند . چون همه ی دانش آموزان به چنین پرسش هایی پاسخ درست می دهند ، در نتیجه نیروی تشخیصی این گونه پرسش ها صفر است ، یعنی نمی توانند در تعیین رتبه ی دانش آموزان نقشی داشته باشند . مثلاً بود و نبود این پرسش یکسان است ، چون همه به آن پاسخ داده اند :

● **انرژی فتوسنتز از کجا تأمین می شود؟**

○ نور

○ باد

○ هوا

○ آب

سوم : پرسش های چندگزینه ای نامناسب اند

همان گونه که از آزمون های رقابتی سنتی و در عین حال بین المللی انتظار می رود ، سؤال های بخش نظری المپیاد جهانی زیست شناسی سؤال هایی بسته هستند ، یعنی فقط یک پاسخ درست دارند . در این آزمون از دانش آموزان خواسته می شود پاسخ درست را با عدد ، حرف یا کد مشخص کنند . چاره ای هم جز این نیست ، چون دانش آموزان به زبان کشور خود به سؤال ها پاسخ می دهند و در صورتی که بخواهند به سؤال ها پاسخ تشریحی بدهند ، تصحیح برگه های پاسخ نامه از سوی کشور میزبان ناممکن می شود . می توان این نوع سؤال ها را چند گروه قرار داد :

● **چندگزینه ای**

● **جورکردنی**

● **تعیین توالی و ترتیب رویدادها**

● **پرکردن جاهای خالی با عدد یا کد**

● **تعیین و تشخیص عبارت ها یا جمله های درست از نادرست**

اکنون سؤال این است که کدام نوع سؤال ها مناسب ترند . به عبارت دیگر اگر بخواهیم آزمونی مناسب درست کنیم ، هر یک از

این انواع سؤال باید چه نسبتی را در آن دخیل باشد؟ یعنی چند درصد از سؤال‌های آزمون‌های خوب و تشخیص دهنده باید چندگزینه‌ای باشد، چند درصد جورکردنی و غیره؟ مرور گزارش‌های کشورهای برگزارکننده‌ی این المپیاد اطلاعات ارزشمندی می‌دهند.

یکی از تجربه‌های جالب در گزارش برگزارکنندگان نهمین المپیاد که در سال ۱۹۹۸ در آلمان برگزار شد، منعکس شده است. این گزارش ترکیب پرسش‌های نظری و نمره‌های دانش‌آموزانی را که در این دوره شرکت کرده بودند، بسیار خوب ارزیابی کرده، اما نکته‌ی جالبی که در این گزارش وجود دارد، این است که پرسش‌های جورکردنی، تعیین توالی و پرکردنی را مناسب‌تر از پرسش‌های چندگزینه‌ای دانسته است. در این گزارش چنین آمده است: «پایایی^۱ پرسش‌های ما به عنوان ابزارهای ارزشیابی برای ارزیابی توانایی دانش‌آموزان مناسب بود، اما توانایی پرسش‌های جورکردنی، تعیین توالی، پرکردنی و داوری بهتر از پرسش‌های چندگزینه‌ای بود. اگر به جای پرسش‌های چندگزینه‌ای، از پرسش‌هایی از نوع تعیین توالی یا داوری و پرکردنی استفاده می‌کردیم، می‌توانستیم تعداد پرسش‌ها را کاهش و روایی^۲ و نیروی تشخیصی^۳ آزمون را افزایش دهیم.»

پایایی، روایی و نیروی تشخیصی آزمون‌ها چه هستند؟ متخصصان ارزشیابی^۴ این مفاهیم را این‌گونه معرفی کرده‌اند: پایایی هر وسیله‌ی اندازه‌گیری به دقت آن اشاره می‌کند. هر آزمون در صورتی پایایی دارد که اگر در فاصله‌ی زمانی کوتاهی آن را چندبار به گروه واحدی از افراد بدهیم، نتایج حاصل به هم نزدیک باشد. روایی به هدف آزمون اشاره دارد. یعنی، آزمون‌ی دارای روایی است که برای اندازه‌گیری آنچه مورد نظر است، مناسب باشد. مثلاً آزمون زیست‌شناسی در صورتی رواست که فقط دانش، مهارت و نگرش زیست‌شناختی دانش‌آموزان را اندازه‌گیری کند، نه چیز دیگر را و نیروی تشخیصی آزمون قدرت آزمون است در تمایزگذاری و تشخیص دانش‌آموزان قوی از دانش‌آموزان ضعیف.

در سال ۱۹۹۹، نخستین تیم دانش‌آموزی ایرانی به دهمین المپیاد جهانی زیست‌شناسی عازم سوئد شد. مسؤولان برگزارکننده‌ی این المپیاد اصول پیشنهادی همکاران آلمانی خود را که یک سال پیش از آن اعلام کرده بودند، به کار بسته و به مجموعه‌ای بی‌نظیر و عالی از پرسش‌های نظری رسیده بودند. اینان به علاوه، سؤال‌ها را با کمک گروهی از دانشگاهیان به طور آزمایشی اجرا و آن‌ها را با کمک گروهی از معلمان محلی زیست‌شناسی بررسی کرده بودند.^۵

متأسفانه در سال‌های بعد، این تجربه‌ی خوب نادیده گرفته شد و بار دیگر تعداد پرسش‌های دانشی رو به فزونی گذاشت. به طوری که حتی در سال‌های ۲۰۰۴ (استرالیا) و ۲۰۰۵ (چین) فقط با پرسش‌های چندگزینه‌ای سروکار داشتیم که مطابق با اصول المپیاد جهانی زیست‌شناسی نبودند و پایایی چندانی برای آزمون ایجاد نمی‌کردند.

برای اطمینان از این‌که آیا میان روایی کلی آزمون و تعداد پرسش‌های جورکردنی، تعیین توالی، پرکردنی ارتباط معنی‌داری وجود دارد یا خیر، بد نمی‌بود اگر به تجزیه و تحلیل آماری نتایج آزمون‌های سال‌های ۱۹۹۰ به بعد نگاهی می‌انداختیم. اما متأسفانه چنین کاری امکان‌پذیر نیست، چون در مورد بسیاری از آن‌ها چنین اطلاعاتی در دست نداریم. تنها چیزی که داریم، یک انحراف معیار کلی است که بیانگر نیروی تشخیصی آزمون است. می‌دانیم که هر اندازه انحراف معیار بیش‌تر باشد، نمره‌های شرکت‌کنندگان متفاوت‌تر خواهد بود. در این رابطه می‌توانیم نیروی تشخیصی همه‌ی آزمون‌ها را به این روش به دست آوریم و مقایسه کنیم:

$$\text{انحراف معیار} = \frac{\text{نیروی تشخیصی آزمون}}{\text{میانگین نمره‌ها}}$$

وارسی کرده‌ایم که آیا برای همه‌ی آزمون‌های المپیاد جهانی زیست‌شناسی رابطه‌ای میان این نیروی تشخیصی و نسبت پرسش‌های جورکردنی، پرکردنی و داوری وجود دارد، یا خیر و به این نتیجه رسیده‌ایم که این رابطه وجود دارد و نتیجه‌گیری آلمانی‌ها درست بوده است. کاهش نسبت پرسش‌های چندگزینه‌ای و در عوض افزایش پرسش‌های انواع دیگر، مانند پرسش‌های جورکردنی، پرکردنی و مانند آن‌ها اعتبار آزمون را افزایش می‌دهد.

آزمون المپیادهای سال‌های ۲۰۰۴ (استرالیا) و ۲۰۰۵ (چین) نامناسب‌ترین نوع آزمون‌های سال‌های اخیر بودند. گفتیم که پرسش‌های این آزمون‌ها فقط از نوع چندگزینه‌ای بود. مسؤل طرح سؤال‌ها در پاسخ به سؤال ما در مورد نوع سؤال‌ها چنین پاسخ داد: «چون با این روش تصحیح برگه‌های پاسخنامه سریع‌تر انجام می‌شود و آسان‌تر است. تصحیح برگه‌هایی که انواع دیگر پرسش‌ها را دارند کار بسیار بزرگی است و تیم بزرگی می‌خواهد که ما در اختیار نداریم». اما می‌دانیم که مدتی است که امکاناتی فراهم شده است که دانش‌آموزان بتوانند اعداد را نیز با علامت زدن در خانه‌های مربوط بیان کنند. مثلاً عدد ۴۷ را این‌گونه نشان دهند:

| | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|
| | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ | ۹ | ۰ |
| دهگان | ○ | ○ | ○ | ● | ○ | ○ | ○ | ○ | ○ | ○ |
| یکان | ○ | ○ | ○ | ○ | ○ | ○ | ● | ○ | ○ | ○ |

چهارم: حضور متخصص ارزشیابی در تیم آزمون ساز لازم است

گاه این نکته‌ی مهم نادیده گرفته می‌شود که حضور متخصص ارزشیابی در تیم آزمون ساز لازم است و آزمون خوب و مناسب محصول مشارکت متخصصان موضوع و متخصصان ارزشیابی است. اکنون پیش از ادامه‌ی بحث بهتر است به این سؤال توجه کنید:

● سیانوباکتری‌ها (جلبک‌های آبی سبز) گروهی از باکتری‌ها هستند که فتوسنتز انجام می‌دهند. درباره‌ی سیانوباکتری‌ها کدام گزاره یا گزاره‌های زیر درست است؟

۱. باکتری‌هایی گرم منفی هستند
۲. در فتوسنتز اکسیژن تولید می‌کنند
۳. همه می‌توانند نیتروژن تثبیت کنند
۴. برخی از آن‌ها با قارچ‌ها همزیست می‌شوند
۵. رنگ آبی سبز آن‌ها فقط از کلروفیل است

الف) همه درست اند

ب) ۱، ۲، ۳ و ۴

ج) ۱، ۲ و ۳

د) ۱، ۲ و ۴

ه) ۱ و ۲

قرار است در این سؤال مهارت انتخاب و داوری ارزشیابی شود. دانش‌آموز باید گزاره‌ها را یکی پس از دیگری بررسی کند و گزینه‌های درست را از گزاره‌های نادرست تشخیص دهد و سپس با مشخص کردن گزینه‌ی مورد نظر آن را معرفی کند. اما این سؤال به شکل سؤال چندگزینه‌ای ساخته شده است و ۵ گزینه دارد. اما متخصصان آزمون‌سازی می‌دانند که می‌توان ۲^۵ یا ۳۲ گزینه برای این سؤال نوشت. به علاوه، اشتباه فاحش سؤال این است که گزاره‌های ۱ و ۲ در همه‌ی گزینه‌ها به عنوان گزاره‌های درست مطرح شده‌اند، بنابراین وجود گزینه‌ی ه که در آن گزاره‌های ۱ و ۲ موجود است، سودی ندارد. متخصصان ارزشیابی این اشتباه‌ها را به آسانی می‌شناسند و آن‌ها را رفع می‌کنند.

پنجم: دشواری سؤال‌ها نیروی تشخیصی آزمون را تعیین نمی‌کند

ممکن است برخی تصور کنند که هر قدر پرسش‌های آزمون دشوارتر باشد، آن آزمون مناسب‌تر است. اما این طور نیست. برای تهیه‌ی آزمون خوب، لازم نیست این اشتباه رایج را تکرار کنیم و بکوشیم همه‌ی سؤال‌ها یا بیش‌تر آن‌ها حتماً دشوار باشند. دشواری

پرسش‌ها موجب افزایش پایایی آزمون یا سبب افزایش نیروی تشخیصی آزمون‌های رقابتی نمی‌شود. آزمون‌هایی که میانگین نمره‌های دانش‌آموزان (P) در آن کم‌تر از ۵۰ درصد باشد، آزمون مناسبی نیست. می‌گویید که اگر میانگین نمره‌ها در آزمون‌هایی بین ۵۰ تا ۷۰ درصد باشد، آن آزمون از نیروی تشخیصی مناسبی برخوردار است. به عبارت دیگر، مقدار P میانگین کلی نمره‌ی دانش‌آموزان و بیانگر دشواری پرسش‌های آزمون است. اگر $P < 40\%$ باشد، آزمون دشوار و اگر $P > 80\%$ باشد، آزمون آسان بوده است. آزمون‌های دشوار و نیز آزمون‌های آسان چندان مناسب نیستند، چون نیروی تشخیص آن‌ها برای تمییز دادن دانش‌آموزان قوی از ضعیف چندان زیاد نیست. مقدار P برای هر سؤال درجه‌ی دشواری آن را روشن می‌کند.

ششم: پرسش‌های دانش‌محور ارزش چندانی ندارند
آزمون خوب و مناسب آزمون‌هایی است که در آن تعداد سؤال‌های ارزشیابی‌کننده‌ی دانش حداقل باشد. معمولاً پرسش‌هایی که با کلمات استفهامی چیست، چه موقع و کجا ختم می‌شوند، پرسش‌های خوبی نیستند. به عکس پرسش‌هایی که با چرا آغاز می‌شوند، مناسب‌ترند.

منظور این نیست که در آزمون‌مان پرسش‌های دانشی نگنجانیم؛ بلکه منظور این است که در مجموع بهتر است همواره پرسش‌های دانشی را در زمینه‌ای استدلالی مطرح کنیم. یکی از آسان‌ترین راه‌های تبدیل پرسش‌های نامناسب به پرسش‌های مناسب آن است که در آن‌ها جدول، طرح، نمودار یا شکل بگنجانیم تا بتوانیم میزان نوآوری و نیروی استدلال دانش‌آموزان را اندازه بگیریم و ارزشیابی^۷ کنیم. چنین پرسش‌هایی دانش‌آموز را وامی‌دارند که از دانش خود در زمینه‌ای نوآورانه و استدلالی استفاده کند.

هفتم: اعداد و ارقام را در قالب پرسش‌های چندگزینه‌ای نپرسیم

اگر می‌خواهیم اعداد و ارقام و درصدهایی را از دانش‌آموزان نپرسیم، بهتر است از پرسش‌های عملکردی استفاده کنیم، نه چندگزینه‌ای. چون دانش‌آموزان ممکن است حتی بدون دانستن پاسخ، به طور تصادفی گزینه‌ی درست را انتخاب کنند، در صورتی که اگر پاسخ نوشتنی باشد، یعنی از دانش‌آموزان خواسته شود عدد را بنویسند، این شانس حذف می‌شود. مثلاً:

● ۷۰٪ از مردم پکن می‌توانند مزه‌ی ماده‌ی فنیل تیوکاربامید را حس کنند. توانایی تشخیص مزه‌ی این ماده (الل T) بر ناتوانی تشخیص آن (الل t) غالب است. چند درصد از زاده‌های افراد توانا،

در تشخیص مزه‌ی این ماده ناتوان خواهند بود؟

- الف. ۲۵٪
- ب. ۱۵٪
- ج. ۱۳٪
- د. ۲۰٪
- ه. ۷/۵٪

به علاوه، در پرسش‌های استاندارد گزینه‌های عددی را به ترتیبی می‌نویسند، مثلاً به ترتیب بزرگی و کوچکی. این نکته در این پرسش رعایت نشده است.

هشتم: از درازنویسی پرهیزیم

هنگام نوشتن سؤال باید از درازنویسی و پیچیده‌نویسی پرهیز کنیم، چون این کار از روایی آزمون می‌کاهد. یعنی در این صورت، ارزشیابی از مهارت‌ها و دانش زیست‌شناختی جای خود را به ارزشیابی از درک و فهم زبان می‌دهد. مثال:

● تفاوت‌های مهم آنزیمی، میان سلول‌های کبدی، کلیه، ماهیچه‌ای و مغزی به علت تفاوت در کاربرد سوخت‌های متابولیک است. کدام مورد زیر نماینده‌ی این تفاوت‌های بیوشیمیایی نیست؟
الف. کبد گلوکز ۶- فسفات دارد، در حالی که سلول‌های ماهیچه‌ای و مغزی ندارند. بنابراین ماهیچه و مغز به عکس کبد، گلوکز به خون آزاد نمی‌کنند.

ب. در کبد مقدار اندکی ترانسفراز مورد نیاز برای فعال کردن استواستات و تبدیل آن به استیل کوآنزیم A وجود دارد. در نتیجه، استواستات و ۳- هیدروکسی بوتیرات از کبد خارج می‌شوند و در ماهیچه‌های قلبی و اسکلتی و مغز مورد استفاده قرار می‌گیرند.

ج. هنگام گرسنگی درازمدت اسیدهای چرب انباشته شده در سلول‌های چربی به اجسام کتون‌ی تبدیل می‌شوند و سپس برای تکمیل اکسایش به مغز و ماهیچه‌ها برده می‌شوند.

د. لاکتات دهیدروژناز در ماهیچه‌ی قلبی وجود ندارد و در نتیجه قلب برای به دست آوردن انرژی جهت انقباض به اکسایش بی‌هوازی روی می‌آورد.

نهم: گزینه‌ها و گزاره‌ها را با دقت انتخاب کنیم

یکی از بهترین رویکردها هنگام طرح پرسش‌های آزمون این است که توصیفی از برخی اطلاعات زیستی (پدیده‌ها، داده‌های پژوهشی، نمودار، جدول و مانند آن‌ها) ارائه دهیم و آن‌گاه بر پایه‌ی چنین توصیف‌هایی گزینه‌ها یا گزاره‌هایی تعیین کنیم به طوری که دانش‌آموز یکی یکی آن‌ها را بررسی کند و به پاسخ برسد. مثال:

● حلزونی روی تخته‌ای راه می‌رود. اگر سنگریزه‌ای به روی تخته پرتاب کنیم، حلزون فوراً به درون صدفش فرو می‌رود. تکرار این کار موجب آهسته‌تر شدن پاسخ حلزون به پرتاب سنگریزه می‌شود، به طوری که سرانجام حلزون به پرتاب سنگریزه پاسخ نمی‌دهد. کدام اصطلاحات زیر ناپدید شدن پاسخ حلزون را توجیه می‌کند؟ (کلمات و اعدادی که روی آن‌ها خط کشیده شده است، هنگام داوری و بازیابی حذف شده و در آزمون نیامده است).

- ۱. سازگاری
- ۲. شرطی شدن
- ۳. خوگیری
- ۴. نقش‌پذیری
- ۵. شناخت
- ۶. رفتار یادگیری
- ۷. الگو شدن
- ۸. کوشش و خطا
- الف. ۱، ۲، ۳، ۵
- ب. ۳، ۵، ۸، ۲، ۴
- ج. ۳، ۶، ۷، ۸
- د. ۲، ۳، ۴، ۵
- ه. ۳، ۴، ۵، ۶

در این پرسش ۶ گزاره (و در پرسش اصلی، پیش از بازیابی ۸ گزاره) وجود دارد. با این گزاره‌ها می‌توانیم $2^6 = 64$ نوع گزینه بنویسیم. روشن است که قرار دادن این همه گزینه در زیر سؤال ناممکن است. بهتر است که دانش‌آموزان گزاره‌ها را یکی یکی بررسی کنند و به پاسخ برسند. مثلاً این گونه:

● ناپدید شدن رفتار گریز حلزون از سنگریزه با کدام مورد یا موارد زیر قابل توجیه است؟

- ۱. سازگاری درست نادرست
- ۲. شرطی شدن درست نادرست
- ۳. خوگیری درست نادرست
- ۴. نقش‌پذیری درست نادرست
- ۵. شناخت درست نادرست
- ۶. رفتار یادگیری درست نادرست

اگر می‌خواهیم درک و فهم دانش‌آموزان را از موضوعی مانند تفاوت‌های آنزیمی ارزشیابی کنیم، سؤال‌هایمان باید روشن، شفاف و دقیق نوشته شده باشد. اگر از دانش‌آموزان بخواهیم منظور ما را از میان انبوهی از واژه‌ها بیرون بکشند، کار درستی نکرده‌ایم. بنابراین پرسش‌ها باید حتی الامکان کوتاه و خلاصه نوشته شده باشند. به علاوه، کاربرد واژه‌ی منفی «نیست» که در سؤال آورده شده است، چندان درست نیست. سؤال باید حتی الامکان فعل مثبت داشته باشد. پایایی سؤال‌های دارای فعل منفی اندک است.

می‌توانیم سؤال را به این شکل‌ها نیز دریاوریم:

| در محل درست علامت بزنید | | نادرست | درست |
|-------------------------|---------------|--------|------|
| ۱ | سازگاری | | |
| ۲ | شرطی شدن | | |
| ۳ | خوگیری | | |
| ۴ | نقش‌پذیری | | |
| ۵ | شناخت | | |
| ۶ | رفتار یادگیری | | |

یا:

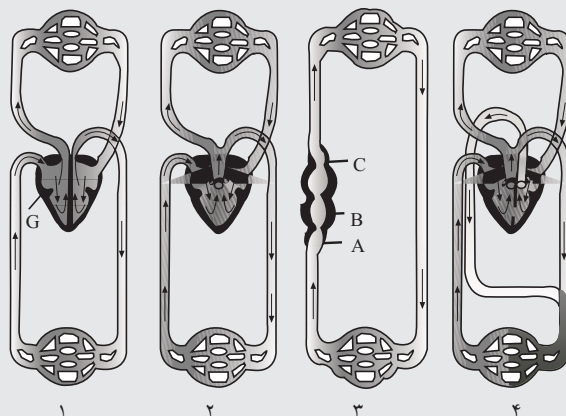
گزینه‌های درست را با علامت + و گزینه‌های نادرست را با علامت - مشخص کنید:

| درست یا نادرست | | نادرست | درست |
|----------------|---------------|--------|------|
| ۱ | سازگاری | | |
| ۲ | شرطی شدن | | |
| ۳ | خوگیری | | |
| ۴ | نقش‌پذیری | | |
| ۵ | شناخت | | |
| ۶ | رفتار یادگیری | | |

دهم: پرسش‌های جورکردنی مناسب‌ترند

یک نمونه از پرسش‌های جورکردنی:

● در شکل زیر چهار دستگاه گردش خون، متعلق به چهار نوع مهره‌دار نشان داده شده است. تعیین کنید از چپ به راست این دستگاه‌ها متعلق به کدام مهره‌داران اند.



- الف. پستاندار، خزنده، دوزیست، ماهی
 ب. ماهی، دوزیست، خزنده و پستاندار
 ج. پستاندار، دوزیست، خزنده و ماهی
 د. پستاندار، دوزیست، ماهی و خزنده

در این مثال از لحاظ نظری می‌توان $4! = 24$ گزینه ساخت که در این جا فقط چهارتای آن‌ها آمده است. پس طراحی این سؤال

چندان ماهرانه نبوده است. اگر این سؤال را به سؤالی جورکردنی تبدیل کنیم، سؤال طرح مناسب‌تری خواهد داشت. آسان‌ترین راه برای این کار استفاده از جدول است:

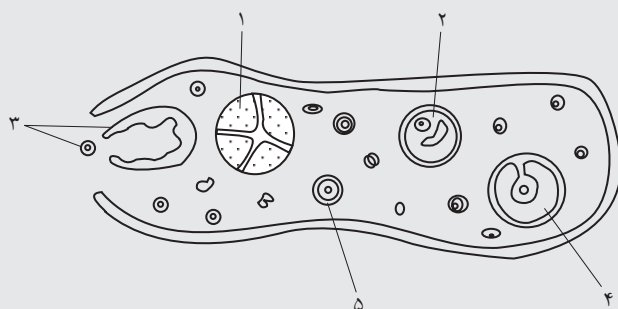
| ۴ | ۳ | ۲ | ۱ | |
|---|---|---|---|----------|
| | | | | ماهی |
| | | | | دوزیست |
| | | | | خزنده |
| | | | | پستاندار |

توجه داشته باشید که در نخستین ستون نام گروه‌ها به ترتیب تکاملی آمده است. در نظر گرفتن ترتیب در طراحی سؤال‌ها اهمیت دارد.

یازدهم: پرسش‌های تعیین توالی، پرسش‌های خوبی هستند

پرسش‌هایی که ترتیب وقایع زیستی را از دانش‌آموزان سؤال می‌کنند، سؤال‌های خوبی اند که معمولاً حیف است آن‌ها را در قالب پرسش‌های چندگزینه‌ای مطرح کنیم، چون چنین سؤال‌هایی باید به تعداد وقایع مورد سؤال گزینه داشته باشند. مثال:

● در این شکل وقایعی را که در تخمدان نوعی پستاندار روی می‌دهد، مشاهده می‌کنید. مراحل مختلف این وقایع با اعداد نشان داده شده‌اند. ترتیب درست را انتخاب کنید.



- الف. ۱، ۲، ۳، ۴، ۵
 ب. ۱، ۲، ۳، ۴، ۵
 ج. ۳، ۱، ۴، ۲، ۵
 د. ۱، ۳، ۴، ۲، ۵
 هـ. ۵، ۳، ۱، ۴، ۲

در این سؤال از ۱۲۰ گزینه‌ی ممکن، فقط ۵ گزینه مطرح شده است. به علاوه، یکی از اشکالات سؤال این است که گزینه‌های ب، ج و د با عدد ۵ شروع شده‌اند. بنابراین دانش‌آموز به راحتی راهنمایی می‌شود که گزینه‌های الف و هـ نادرست‌اند. اگر این سؤال



را به سؤالی تعیین توالی تبدیل کنیم، بهتر است:

● ترتیب رویدادهای شکل را بنویسید: اول ، دوم ، سوم ، چهارم ، پنجم

می توان این سؤال را به چند سؤال چندگزینه ای نیز تبدیل کرد:

● کدام رویداد اول اتفاق می افتد؟

- الف . ۱
- ب . ۲
- ج . ۳
- د . ۴
- هـ . ۵

● کدام رویداد دوم اتفاق می افتد؟

- الف . ۱
- ب . ۲
- ج . ۳
- د . ۴
- هـ . ۵

و نیز به سؤالی جدول دار:

| | اول | دوم | سوم | چهارم | پنجم |
|----------|-----|-----|-----|-------|------|
| رویداد ۱ | | | | | |
| رویداد ۲ | | | | | |
| رویداد ۳ | | | | | |
| رویداد ۴ | | | | | |
| رویداد ۵ | | | | | |

دوازدهم: ارزیابی آماری آزمون به تعیین کیفیت آن کمک می کند
ارزیابی آماری هر آزمون اطلاعاتی مفید درباره ی کیفیت و روایی
آن فراهم می کند. برخی از این ارزیابی ها چنین اند:

Rit

Rit که با این فرمول محاسبه می شود، کیفیت سؤال را در آزمون
نشان می دهد.

$$Rit = \frac{Yc - Yw}{Sy} \sqrt{pq}$$

Yc = میانگین نمره های دانش آموزانی که به آن پرسش پاسخ

درست داده اند.

Yw = میانگین نمره های دانش آموزانی که به آن پرسش پاسخ

نادرست داده اند.

Sy = انحراف معیار نمره های دانش آموزان از آن پرسش.

p = نسبت دانش آموزانی که به آن پرسش پاسخ درست

داده اند.

q = نسبت دانش آموزانی که به آن پرسش پاسخ نادرست

داده اند.

Rit برای سنجش کیفیت هر سؤال در آزمون مناسب است و اطلاعاتی به ما می دهد مبنی بر آن که تا چه حد دانش آموزان قوی به سؤالی خاص بهتر از دانش آموزان ضعیف پاسخ داده اند. منظور از دانش آموز قوی، دانش آموزی است که نمره ای بیش تر از میانگین نمره ها به دست آورده باشد و دانش آموز ضعیف دانش آموزی است که نمره ای کم تر از میانگین کسب کرده باشد. مقدار Rit بین +۱ و -۱ متغیر است. اگر $Rit = +1$ باشد سؤال بسیار مناسب است، یعنی همه ی دانش آموزان قوی بدان پاسخ درست و همه ی دانش آموزان ضعیف بدان پاسخ نادرست داده اند. اما این حالت بسیار به ندرت اتفاق می افتد. برای سؤال های خوب $Rit > 0.3$ است.

اگر $Rit = 0$ باشد، نیروی تشخیصی سؤال صفر و چنان که گفته شد، ارزش آن در آزمون های رقابتی نیز صفر است. سؤال های دارای $Rit < 0$ نامناسب اند. چون تعداد دانش آموزان قوی ای که به آن پاسخ نادرست داده اند از تعداد دانش آموزان ضعیفی که بدان پاسخ نادرست داده اند بیش تر است.

زیرنویس

1. Reliability
2. Validity
3. Discriminating power

۴. مثلاً: سیف، دکتر علی اکبر؛ روش های اندازه گیری و ارزشیابی آموزشی؛ نشر دوران؛ چاپ سوم؛ ۱۳۷۸.
۵. این نوآوری هم اکنون به عنوان توصیه در راهنمای المپیاد جهانی زیست شناسی وجود دارد.
۶. توضیح: موارد ۱ تا ۵ این سؤال را گزاره و موارد الف تا هـ آن را گزینه می نامند.
۷. اندازه گیری و ارزشیابی دو مفهوم مجزا و مستقل اند. هرگاه کمیتی را به چیزی نسبت دهیم، مثلاً طول شیئی را تعیین کنیم، آن را اندازه گرفته ایم؛ اما هرگاه اندازه های متناظر چند چیز را به منظور تعیین میزان دسترسی به هدف ها با هم مقایسه و دآوری کنیم، مثلاً طول های چند شیئی را با هم مقایسه و نتیجه گیری کنیم، آن گاه آن ها را ارزشیابی کرده ایم.