



آزمونهای

پیشرفت تحصیلی

دکتر علوم پزشکی
پزشکی



نویسندگان:

دکتر بهزاد ذوالفقاری

ثریا درخشانفر

ندا ادیبی

افسانه کرباسی

مزگان تنساز

پروانه نیرومند

از سری انتشارات
مدیریت مطالعات و
توسعه آموزش
پزشکی - ۲



آزمونهای پیشرفت تحصیلی در علوم پزشکی

نویسندگان

دکتر بهزاد ذوالفقاری

ندا ادیبی ثریا درخشان فر

مژگان تن ساز افسانه کرباسی

پروانه نیرومند

مرکز مطالعات و توسعه آموزش پزشکی

۱۳۷۹

نوالفقاری، بهزاد، ۱۳۳۹ -

آزمونهای پیشرفت تحصیلی در علوم پزشکی / بهزاد نوالفقاری. - اصفهان: دانشگاه علوم پزشکی و خدمات بهداشتی درمانی اصفهان، مدیریت مطالعات و توسعه آموزش پزشکی، ۱۳۷۸. ۳۰۸ ص.

ISBN 964-92438-1-X

فهرست نویسی بر اساس اطلاعات فیها.

۱ پزشکی - آزمونها و تمرینها. ۲. پزشکی - کتابهای درسی - راهنمای آموزشی. الف. دانشگاه علوم پزشکی و خدمات بهداشتی درمانی اصفهان، مدیریت مطالعات و توسعه آموزش پزشکی. ب. عنوان.

۶۱۰/۷۶

RC ۵۸/ز ۹۱۲

م ۷۸-۲۰۰۶۳

کتابخانه ملی ایران

مروری بر آزمونهای پیشرفت تحصیلی در علوم پزشکی

نویسندگان:	دکتر بهزاد نوالفقاری؛ ندا انبیبی؛ ثریا درخشان‌فر؛ مژگان تن‌ساز؛ افسانه کریاسی؛ پروانه نیرومند
ویراستاران علمی:	دکتر قربانعلی اسداللهی؛ دکتر بهزاد نوالفقاری
ویراستار ادبی:	دکتر محمدرضا یوسفی کوهپایی
ناشر:	مدیریت مطالعات و توسعه آموزش پزشکی دانشگاه علوم پزشکی اصفهان
نوبت چاپ:	اول - بهار ۱۳۷۹
تعداد:	۲۰۰۰ نسخه
چاپ:	چاپ نشاط اصفهان

ISBN 964-92438-1-X

شابک ۹۶۳-۹۲۳۳۸-۱-X

پیشگفتار نویسندگان

با سپاس بیکران بر ذات لایزال او که بدون توفیق او هیچکس بر کاری استوار نگردد. و این کتاب نیز که حاصل همکاری و همراهی جمعی است جز به منت او، حاصل نگردید.

با گسترش دانشگاهها مؤسسات آموزش عالی، توجه به کیفیت اجرای برنامه‌های آموزشی به عنوان بخشی از روند مستمر ارتقاء کیفیت به عنوان ضرورتی اجتناب‌ناپذیر مطرح گردیده است و در این راستا آزمونهای دانشگاهی که به عنوان ابزاری برای ارتقاء قبول یا رد دانشجویان به کار می‌رود، به عنوان شاخصی کلیدی و حساسترین نقطه هر برنامه آموزشی از جایگاهی ویژه برخوردار است چرا که دانشجویان همانگونه می‌آموزند که مورد آزمون قرار می‌گیرند یا به سخن دیگر، آزمونها هستند که شیوه، سطح و عمق یادگیری دانشجویان و در نتیجه میزان بهره‌وری هر نظام آموزشی را تعیین می‌کنند.

از این رو کتاب حاضر در پی ارائه راهکارهایی اجرایی است تا به کمک آن طراحان آزمونها با مبانی، تعاریف و اهداف و انواع آزمونها (بخش اول)؛ شیوه‌های طراحی آزمونهای مختلف با ذکر معایب، مزایا و محدودیتهای هر روش همراه با مثالهای تفضیلی (بخش دوم)، اصول قضاوت در زمینه یک آزمون خوب و شیوه تجزیه و تحلیل سوالات آن و اصول اجرای آزمونها، همچنین یک الگوی سازمان یافته آزمونهای استاندارد (بخش سوم) آشنا گردند؛ به نحوی که بتوانند با انتخاب روشهای مناسب و بهینه در وضعیتهای خاص زمانی و مکانی کیفیت آموزش را در بخشهای مختلف نظام آموزشی ارتقاء بخشند.

از آنجا که اصول هر نظام آموزشی و مبانی سنجش و ارزشیابی، اصولی واحد است، این کتاب می‌تواند به کسانی که به نحوی در چرخه آموزشی به عنوان معلم عمل می‌کنند، کمک نماید اما هدف اصلی آن کمک به بهبود نظام ارزشیابی دانشجویان در دانشگاههای علوم پزشکی کشور می‌باشد، که امید است با توجه و امعان نظر کلیه خوانندگان محترم بتوانیم از دیدگاههای ارزشمند همگان بهره‌مند گردیم، چرا که بر وجود نقایص و کاستیها پیشاپیش اعتراف داریم.

در پایان نویسندگان وظیفه خود می‌دانند مراتب تشکر خود را از افراد زیر ابراز

دارند:

مسئولان محترم دانشگاه علوم پزشکی اصفهان به ویژه معاونت محترم آموزشی (آقای دکتر بهزاد شمس)، مدیریت محترم مرکز مطالعات و توسعه آموزش پزشکی (خانم دکتر طاهره چنگیز)، مدیر محترم گروه آموزش پزشکی (آقای دکتر علیرضا یوسفی) که بدون ارشاد و حمایت ایشان، انجام این کار میسر نبود؛ آقای دکتر قربانعلی اسداللهی که علیرغم مشغله فراوان، با بازبینی کتاب و با ارائه دیدگاههای ارزشمند خود به غنای علمی این مجموعه افزودند، آقای دکتر علی دباغ که زحمت بازبینی کتاب را به عهده گرفتند، آقای دکتر محمدرضا یوسفی که زحمت ویرایش ادبی کتاب را به عهده داشتند، آقایان دکتر مسیح صبوری، دکتر علیرضا قنادی، دکتر ناصر توکلی، دکتر عباس جعفریان که در طول تدوین کتاب با سعه صدر و ارائه پیشنهادات ارزنده ما را قرین منت خود قرار دادند. آقایان دکتر پیمان ادیبی و دکتر همایون بفعتی که در مراحل اجرایی به ما یاری رساندند، آقای رضا جوشن که طراحی روی جلد را به عهده داشتند، سرکار خانم آزیتا رستمی که تایپ کتاب را انجام دادند.

فهرست مطالب

	بخش اول / کلیات
۲.....	فصل اول / آشنایی با مفاهیم.....
۵.....	مقدمه.....
۶.....	ارزیابی.....
۱۲.....	اندازمگیری.....
۱۵.....	۱- اندازمگیری ملاک مرجع.....
۱۵.....	۲- اندازمگیری هنجار مرجع.....
۱۷.....	ارزشیابی.....
۱۹.....	۱- ارزشیابی مجموعی.....
۱۹.....	۲- ارزشیابی تکوینی.....
۲۰.....	۳- ارزشیابی تشخیصی.....
۲۰.....	آزمون.....
۲۳.....	۱- تعیین نقطه آغاز تعلیم در دوره بعدی.....
۲۳.....	۲- آزمون برای ایجاد انگیزه.....
۲۳.....	۳- آزمون برای تمرین و تکرار.....
۲۴.....	۴- آزمون تعیین میزان پیشرفت تحصیلی.....
۲۵.....	۵- آزمون برای ارزشیابی میزان موفقیت در تدریس.....
۲۵.....	انتخاب نوع پرسش مناسب.....
۲۷.....	مراحل تدوین سؤالات.....
۳۰.....	مزایای استفاده از جدول طرح آزمون.....
۳۵.....	فصل دوم / طبقه‌بندی آزمونها.....
۳۷.....	۱- آزمون معلم ساخته.....
۳۸.....	۲- آزمون استاندارد شده.....
۴۱.....	۱- آزمون عینی.....
۴۲.....	۲- آزمون نهنی.....
۴۲.....	۱- آزمون گزیده پاسخ.....
۴۳.....	۲- آزمون باز پاسخ.....
۴۳.....	۱- سؤالات گسترده یا بی ساختار.....
۴۳.....	۲- سؤالات سازمان یافته.....

۲۵	۲- سؤالات بسیار سازمان یافته
۲۵	۱- آزمون ملاک نسبی
۲۵	۲- آزمون ملاک مطلق
۲۹	۱- آزمونهای پیشرفت تحصیلی
۵۰	۲- آزمونهای رغبت
۵۰	۳- آزمونهای شخصیت
۵۱	۴- آزمونهای هوش و استعداد
۵۱	۵- پیش آزمون
۵۱	۶- آزمون مهارت
۵۱	۷- آزمون تشخیصی
۵۲	۱- آزمون سرعت
۵۲	۲- آزمون قدرت

بخش دوم / انواع آزمون

۵۷	فصل سوم / آزمونهای کتبی مرسوم
۵۹	سؤالات چند گزینه‌ای
۵۹	تعریف سؤال چندگزینه‌ای
۶۰	انواع سؤالات چندگزینه‌ای
۶۳	الف) قواعد مربوط به نوشتن ریشه سؤال
۶۸	ب) قواعد مربوطه به نوشتن گزینه‌ها
۷۳	ج) نکات دیگر
۷۶	نکات اجرایی آزمون چندگزینه‌ای
۷۷	موارد استفاده از سؤالات چند گزینه‌ای
۷۹	انواع سؤالات ص - غ
۷۹	طراحی سؤالات ص - غ
۸۲	مزایا
۸۲	معایب
۸۷	طراحی سؤالات جورکردنی
۸۹	مزایا
۹۱	کاربرد سؤالات کوتاه پاسخ
۹۲	نحوه ساخت
۹۳	مزایا
۹۳	معایب
۹۵	پرسش تشریحی چیست؟
۹۶	۱- سؤالات تشریحی با جواب محدود
۹۶	۲- سؤالات تشریحی گسترده پاسخ یا جواب نامحدود

۹۷	کاربرد آزمون تشریحی
۹۸	طراحی پرسشهای تشریحی
۱۰۱	معایب آزمونهای تشریحی
۱۰۲	نمره‌گذاری آزمون تشریحی
۱۰۳	طراحی آزمونهای تشریحی
۱۰۵	راهنمایی روی جلد
۱۰۹	فصل چهارم / آزمون انارۀ بیمار
۱۱۱	آزمون انارۀ بیمار
۱۱۳	معاینه فیزیکی
۱۱۳	مطالعات آزمایشگاهی
۱۱۷	طراحی
۱۱۸	انواع آزمونهای انارۀ بیمار
۱۱۹	انواع گزینه‌های به کار برده شده
۱۱۹	روش نمره‌گذاری
۱۲۰	تعیین حدنصاب قبولی
۱۲۱	مزایا
۱۲۱	معایب
۱۲۳	فصل پنجم / آزمونهای شفاهی
۱۲۵	آزمونهای شفاهی
۱۲۶	انواع آزمونهای شفاهی
۱۲۷	قواعد طراحی آزمون شفاهی
۱۳۰	مزایای آزمون شفاهی
۱۳۰	معایب آزمونهای شفاهی
۱۳۱	چند نکته در برگزاری آزمونهای شفاهی
۱۳۵	فصل ششم / امتحانات میان دوره
۱۳۷	امتحانات میان دوره
۱۳۸	الف) شیوۀ مشاهده‌ای
۱۳۸	ب) شیوۀ کتبی
۱۳۳	فصل هفتم / خودآزمایی
۱۳۵	خودآزمایی
۱۳۵	انواع آزمونهای خودآزمایی و کاربردهای آن
۱۳۹	خودآزمایی و برآورد دانشجویان
۱۵۳	جنبه‌های کاربردی خودآزمایی
۱۶۱	فصل هشتم / سنجش حیطه‌های غیرشناختی
۱۶۳	سنجش عملکرد

۱۶۵	سطوح سنجش عملکرد.....
۱۶۸	مراحل طراحی آزمونهای عملکرد.....
۱۷۱	ابزارها و روشهای مشاهدهای.....
۱۷۲	۱- فهرستهای بازبینی (وارسی).....
۱۷۵	۲- مقیاسهای درجه‌بندی.....
۱۷۵	انواع مقیاسهای درجه‌بندی.....
۱۷۷	منابع خطا در مقیاسهای درجه‌بندی.....
۱۷۸	راههای بهبود داده‌های ناشی از مقیاسهای درجه‌بندی.....
۱۸۰	آزمونهای سنجش نگرش.....
۱۸۱	انواع مقیاسهای نگرش گزارش شخصی.....
۱۸۱	ساخت مقیاس نگرش به روش لیکرت.....
۱۸۲	روش تفکیک معنایی آنگود.....
۱۸۳	رهنمودهای کلی برای ساخت مقیاسهای نگرش.....
۱۸۵	مراحل طراحی یک پرسشنامه ایده‌آل.....
بخش سوم / فرآوری آزمون	
۱۹۱	فصل نهم / پایایی و روایی.....
۱۹۲	پایایی آزمون.....
۱۹۳	ضریب همبستگی.....
۱۹۳	روش محاسبه ضریب همبستگی.....
۱۹۷	روشهای تعیین پایایی.....
۲۰۰	پایایی در آزمونهای ملاک مرجع.....
۲۰۱	عوامل مؤثر بر پایایی.....
۲۰۱	تفسیر ضریب پایایی.....
۲۰۲	خطای معیار اندازه‌گیری.....
۲۰۳	برآورد نمره واقعی با استفاده از ضریب پایایی.....
۲۰۳	روایی.....
۲۰۵	انواع روایی.....
۲۰۹	روش محاسبه روایی.....
۲۱۱	عوامل مؤثر بر روایی آزمون.....
۲۱۳	ارتباط روایی و پایایی.....
۲۱۵	فصل دهم / تجزیه و تحلیل سؤالات آزمون.....
۲۱۷	تجزیه و تحلیل سؤالات آزمون.....
۲۱۸	خطای اندازه‌گیری.....
۲۱۸	تعریف خطای اندازه‌گیری.....
۲۱۹	تجزیه و تحلیل سؤال.....

۲۱۹	روشهای تجزیه و تحلیل سؤالات
۲۲۰	ارائه بازخورد به دانشجویان
۲۲۱	رواسازی محتوای سؤالات
۲۲۳	کمی سازی نظر داوران
۲۲۴	توصیف نمرات آزمون
۲۲۶	روش دیگر برای محاسبه انحراف معیار
۲۲۹	شاخص نسبت صحیح
۲۲۹	استفاده از مقدار P برای تجزیه و تحلیل سؤال
۲۳۱	سؤالات با کلید اشتباه
۲۳۲	مقایسه مقادیر P برای گروههای بالا و پایین
۲۳۳	انجام مقایسه چند طبقه‌ای
۲۳۶	شاخصهای تمیز سؤال
۲۳۹	ضریب همبستگی دو رشته‌ای نقطه‌ای
۲۴۲	همبستگی دورشته‌ای تخمینی
۲۴۳	ضریب فی
۲۴۵	عوامل سؤال
۲۴۸	تجزیه و تحلیل آزمونهای ملاک مرجع
۲۴۹	۱- ضریب حساسیت
۲۴۹	۲- ضریب تمیز تسلط - فقدان تسلط
۲۵۰	تفسیر داده‌های حاصل از تجزیه و تحلیل سؤالات
۲۵۵	فصل یازدهم / اجرا و نمره‌گذاری
۲۵۷	برگزاری آزمون
۲۵۷	توزیع سؤالات امتحانی
۲۵۷	مدیریت
۲۵۹	تعیین مدت زمان آزمون
۲۵۹	جمع‌آوری برگه‌ها
۲۶۰	تقلب
۲۶۲	اضطراب آزمون
۲۶۳	نمره‌گذاری
۲۶۶	چگونگی نمره‌گذاری و تصحیح آزمون
۲۶۷	نمره منفی
۲۶۷	نمره منفی در سؤالات عینی
۲۶۹	اثر شانس بر میزان موفقیت
۲۷۱	تبدیل نمرات آزمون چندگزینه‌ای
۲۷۲	عوامل مؤثر در نمرات آزمون

۲۷۳	عینی بودن نمره آزمون
۲۷۹	فصل دوازدهم / آزمونهای استاندارد
۲۸۱	مراحل استانداردسازی آزمون
۲۸۲	تعیین اهداف آزمون
۲۸۲	انجام پیش‌آزمون
۲۸۳	انجام آزمونهای مقدماتی
۲۸۴	USMLE
۲۸۶	تأثیر آزمونهای USMLE بر روند آموزشی دانشکده‌های پزشکی
۲۸۹	نتیجه کلی

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

پناه می‌بریم به خدا از علمی که سود نداشته باشد و آن علمی است که با عمل خالصانه همراه نیست و بدان که اندکی از علم نیازمند عمل فراوان است، زیرا علمی که انسان در یک ساعت می‌آموزد، او را به عمل کردن بر طبق آن در سراسر عمر ملتزم می‌کند. پیامبر اکرم (ص)، ترجمه الحیات، ج ۱، ص ۱۴۲

امروزه مراکز توسعه آموزش پزشکی در جهان بعنوان نهادهای مسؤول پایش و ارتقای کیفیت آموزش پزشکی در جهت بهبود خدمات فارغ‌التحصیلان علوم پزشکی به جامعه شناخته می‌شوند. بر اساس توافق بین‌المللی، این مراکز از طریق برنامه‌ریزی، بهبود فرآیند یاددهی و یادگیری، ارزشیابی، آموزش مداوم و تحقیق در آموزش برای دستیابی به مقاصد فوق تلاش می‌نمایند.

خوشبختانه دانشگاه علوم پزشکی اصفهان از سال ۱۳۶۸ با ایجاد واحدی بنام «مدیریت برنامه‌ریزی، نظارت و ارزشیابی آموزش پزشکی» این موضوع را در دستور کار خود قرار داد. سپس فعالیت واحد مذکور از سال ۱۳۷۱ در قالب مرکز توسعه آموزش پزشکی شناخته شد. این مرکز خدمات خود را در پنج گروه برنامه‌ریزی درسی، ارزشیابی، آموزش اساتید، آموزش مداوم و پژوهش در آموزش ارائه می‌دهد. در بیشتر فعالیت‌های آموزشی مرکز (اعم از کارگاه‌های مقدماتی روش تدریس، ارزیابی دانشجو، ارزشیابی معلم و برنامه‌ها، روش‌های نوین یاددهی و یادگیری) که برای اعضای هیأت علمی، مدیران آموزشی و مسؤولین هر قسمت طراحی و اجرا می‌شود، فقدان منابع علمی فارسی که از یک طرف در برگیرنده مطالب علمی جدید و از طرف دیگر منعکس‌کننده تجارب کشوری باشد، همواره محسوس بود. به همین جهت، این مدیریت از مدتها پیش در پی آن بوده است که برای

گسترش هر چه مؤثرتر حیطه فعالیت‌های خود، مجموعه انتشاراتی را با کمک افراد علاقه‌مند و با تجربه در زمینه‌های موردنیاز فراهم نماید.

در این راستا نگارش سری کتابهای آموزش پزشکی در برنامه کار این مرکز قرار گرفت. اولین کتاب از این سری با عنوان «مقدمه‌ای بر آموزش بر مبنای مسأله (PBL)» همزمان با سومین همایش کشوری آموزش پزشکی در سال ۱۳۷۷ منتشر شد. دومین کتاب تحت عنوان «تحلیل کاربردی داده‌ها با کمک نرم‌افزار SPSS» در پاییز سال ۱۳۷۸ منتشر گردید و آنچه پیش رو دارید، سومین کتاب از مجموعه فوق است که با مشارکت جمعی از دانشجویان فعال در مرکز پژوهش‌های دانشجویی و هدایت و سرپرستی آقای دکتر ذوالفقاری که از همکاران شایسته و پرتلاش این مدیریت هستند با صرف وقت و دقت فراوان به رشته تحریر درآمده است. همچنین کتاب با مساعدت و همکاری جناب آقای دکتر قربانعلی اسداللهی (دانشیار محترم گروه روانپزشکی این دانشگاه و همکار مدیریت مطالعات و توسعه آموزش پزشکی) و جناب آقای دکتر ذوالفقاری بطور کامل و دقیق بازخوانی و ویرایش گردیده است. با وجود این، نویسندگان محترم و ناشر خود را همچنان نیازمند راهنمایی و انتقادهای ارباب نظر می‌دانند و چشم به راه ارشاد خوانندگان محترم می‌مانند.

در پایان بر خود لازم می‌دانیم سپاس این مدیریت را از حمایت‌های مادی و معنوی مرکز مطالعات و توسعه آموزش پزشکی وزارت بهداشت، درمان و آموزش پزشکی و همکاری مؤثر مسئولین دانشگاه علوم پزشکی اصفهان و مرکز پژوهش‌های دانشجویان ابراز نماییم و از خداوند متعال توفیق روزافزون این عزیزان را در خدمت به میهن اسلامی طلب کنیم.

مدیریت مطالعات و توسعه آموزش پزشکی

دانشگاه علوم پزشکی و خدمات بهداشتی - درمانی استان اصفهان

فهرست

۱	بخش اول / کلیات
۳	فصل اول / آشنایی با مفاهیم
۲۵	فصل دوم / طبقه‌بندی آزمون‌ها
۵۵	بخش دوم / انواع آزمون
۵۷	فصل سوم / آزمونهای کتبی مرسوم
۱۰۹	فصل چهارم / آزمون اداره بیمار
۱۲۳	فصل پنجم / آزمون‌های شفاهی
۱۳۵	فصل ششم / امتحانات میان دوره
۱۴۳	فصل هفتم / خودآزمایی
۱۶۱	فصل هشتم / سنجش حیطه‌های غیرشناختی
۱۸۹	بخش سوم / فراوری آزمون
۱۹۱	فصل نهم / پایایی و روایی
۲۱۵	فصل دهم / تجزیه و تحلیل سوالات آزمون
۲۵۵	فصل یازدهم / اجرا و نمره‌گذاری
۲۷۹	فصل دوازدهم / آزمونهای استاندارد

بخش اول

کلیات

فصل اول

آشنایی با مفاهیم

- مقدمه
- ارزیابی
- اندازه‌گیری
- ارزشیابی
- آزمون

مقدمه

افراد آموزش دیده، محصول نظامهای آموزشی هستند و هر نظام آموزشی مجموعه‌ای از افراد، اشیا و امکانات است که برای مدت زمانی با صرف هزینه، همراه با یک برنامه مدون در مواجهه با یکدیگر قرار می‌گیرند. در این نظام، گروهی آموزش می‌دهند و گروهی فرا می‌گیرند؛ گروهی با صرف هزینه و وقت، برنامه‌ای را تدوین می‌کنند و گروهی دیگر محصول و ثمره این دسته خواهند بود.

پس در هر فرآیند آموزشی، افرادی به عنوان ماده اولیه وارد نظام آموزشی می‌شوند و فعل و انفعالاتی روی آنها انجام می‌گیرد تا به عنوان ماده تولیدی کارگاه آموزش، به صورت افرادی فرهیخته وارد جامعه شوند. آنچه نظامهای آموزشی را از هم متمایز می‌سازد فعل و انفعالات درون مجموعه آموزشی است و ال‌اهداف تمام نظامهای آموزشی آموزش و پرورش انسانها است.

از سوی دیگر هر مجموعه برای ارتقای کیفی خود ابتدا باید بداند محصولاتی که تولید نموده است تا چه حد مطلوب، منطقی و مطابق با اهداف و استانداردهای اولیه موردنظر است. به بیان دیگر باید بداند بازده (راندمان) دستگاه آموزش تا چه اندازه بوده است و زمان و هزینه مصرف شده در نظام آموزشی تا چه حد بهره‌وری داشته است. برای نیل به چنین هدفی است که فرآیند ارزیابی^۱ در یک نظام آموزشی بکار گرفته می‌شود.

البته مسلم است که ارزیابی به تنهایی سودمند نخواهد بود، یعنی اگر در ارزیابی دو فرد آموزش دیده، به یک نفر نمره هیجده و به دیگری دوازده بدهیم این دو عدد به تنهایی مفهومی نخواهد داشت؛ مگر وقتی که ما ارزش نسبی این دو عدد را نسبت به عددی مثل بیست یا صفر بسنجیم و بعد تفسیر نماییم که بیست یعنی کسب کامل مهارت موردنظر، پس فرد اول در کسب هدف ما موفق‌تر بوده است، چنین تجزیه و تحلیلی از فرآیند ارزیابی، مفهوم جدیدی را به نام «ارزشیابی^۲» مطرح می‌کند؛ که به معنای ارزش دادن به آن چیزی است که فرد یا افراد کسب نموده‌اند.

اما ابزار^۱ ارزیابی آموزش چیست؟ بهترین وسیله ارزیابی، «آزمون^۲» است. به عبارت بهتر ارزیابی یا سنجش فرد از طریق فرآیندی به نام امتحان یا آزمون صورت می‌پذیرد که به وسیله آن میزان کارایی و قابلیت فرد آموزش دیده سنجیده می‌شود.

طراح، مدرس و مسئول هر نظام آموزشی ابتدا باید بدانند که ارزیابی، ارزشیابی و اندازه‌گیری سه فرآیند جداگانه بوده، وظیفه هر یک از آنها، اجرای یکی از این فرآیندهاست، یعنی طراح یک برنامه آموزشی باید متوجه باشد وظیفه او فراهم کردن برنامه‌ای است که:

- ۱- هدف مشخص و مهمی را دنبال کند؛

- ۲- هدف مذکور در دسترس بوده، نحوه دستیابی به آن در متن برنامه گنجانده شده باشد؛

- ۳- هدف آن، از نظر کمی و کیفی قابل سنجش باشد.

در حقیقت، یک طراح، فراهم آورنده مقدمات لازم برای اجرای ارزیابی و ارزشیابی است. وظیفه مدرس، آموزش و اجرای فرآیند ارزیابی و اندازه‌گیری با یک روال منطقی و عینی است به طوری که میزان دستیابی به اهداف را به طور کامل بسنجد. وظیفه مسئولان آن است که به این ارزیابیها، ارزشهای کیفی اختصاص داده، بر مبنای آن به اصلاح نظام آموزشی و ارتقای کیفی آموزش بپردازند؛ زیرا در حقیقت آنها تأمین‌کننده بازخورد^۳ آموزشی هستند.

ممکن است هر سه مسئولیت به عهده یک نفر باشد و در بیشتر اوقات مدرس در روند ارزشیابی دخالت می‌کند؛ چون او نیز برای کسب بازخورد از روند تدریس و انتقال آن به دانشجویان از نتایج ارزیابی استفاده می‌نماید.

در مطالب آینده سعی خواهیم کرد نقاط اشتراک و افتراق فرآیندهای ارزیابی، ارزشیابی و اندازه‌گیری را بیان نموده، مشخص کنیم که هر یک در چه زمینه‌ای کاربرد دارد.

ارزیابی

ارزیابی یکی از مهمترین مراحل نظام آموزشی است. ارزیابی به معنی کشف میزان یادگیری دانشجو است و به این منظور، از برگزاری آزمون یا مشاهده دانشجوین حین انجام کار استفاده می‌شود.

ارزیابی باید دارای هدف بوده، آنچه مورد سنجش قرار می‌گیرد هم برای استاد و هم

1- Tool

2- Examination

3- Feedback

برای دانشجو کاملاً مشخص باشد. هدف از انجام ارزیابی، سنجش وسعت اندوخته‌های دانشجو از نظر دانش^۱، مهارت^۲ و نگرش^۳ است که این دانسته‌ها توسط هدفهای آموزشی یک دوره خاص معین شده است. به طور کلی با ارزیابی می‌توان به اهداف زیر دست یافت:

۱- تعیین سطح آگاهی و تواناییهای دانشجو در ابتدای آموزش برای مشخص نمودن نقطه شروع درس.

۲- تشخیص مشکلات یادگیری هر دانشجو و در صورت لزوم برگزاری دوره‌های درسی تقویتی. آموزش همراه با ارزیابی مستمر، رغبت دانشجو را برای یادگیری افزایش داده، به سعی و تلاش او در زمینه یادگیری نظام می‌بخشد. علاوه بر این، مشخص گردیده است که دانشجویان بر همان مبنایی که از آنها امتحان گرفته می‌شود، می‌آموزند. لذا ارزیابی می‌تواند در بهبود فرآیند یادگیری مؤثر باشد.

۳- تشخیص اشکالات تدریس و مواد درسی و ایجاد اصلاحات لازم در آن. زیرا علل عدم موفقیت را همیشه نمی‌توان به دانشجو نسبت داد و گاهی شیوه تدریس عامل شکست و ناکامی است.

۴- تعیین فاصله بین اهداف موردنظر و آنچه تحقق یافته است. چون گاهی اهداف آموزشی که در ابتدای تدریس در نظر گرفته شده به درستی انتخاب گردیده است ولی موانعی وجود دارد که دستیابی به آنها را ناممکن می‌سازد. با انجام آزمون می‌توان از این حقیقت آگاه شد و در جهت رفع آن تلاش کرد.

۵- بررسی میزان پیشرفت دانشجو، و تصمیم‌گیری راجع به طبقه‌بندی و اعطای مدرک. در هر ارزیابی باید به چند نکته توجه نمود از جمله:

۱- هدف مورد نظر چه بوده و آنچه دانشجو بدان دست یافته چیست؟

۲- چه افرادی از نتیجه ارزیابی استفاده می‌کنند؟

۳- با توجه به اهداف مورد انتظار^۴ و حقیقی^۵ آیا روند ارزیابی مناسب است؟

ارزیابی، دانشجویان را تشویق می‌کند که بهتر تلاش کنند. همچنین استاد را مطمئن می‌کند که دانشجو کار محول شده را به خوبی انجام می‌دهد یا خیر. البته هیچ ارزیابی

1- Knowledge

2- Skill

3- Attitude

۴- Expected Goal. هدف موردانتظار هدفی است که استاد در ابتدای یک دوره درسی در نظر دارد. یعنی هدفی که پیش از برگزاری دوره درسی وجود دارد و دوره درسی برای کسب این هدف طراحی شده است.

۵- Real Goal. هدف حقیقی هدفی است که در جریان برگزاری یک دوره آموزشی بدست آمده و در حقیقت بازده دوره و چیزی است که دانشجو پس از طی دوره آموزشی کسب نموده است.

نمی‌تواند به همه اهداف ذکر شده دست یابد. به عنوان مثال:

آزمون پایان هر دوره برای مشخص نمودن قابلیت دانشجو در انجام وظایف محول شده مناسب است، اما بازخوردی برای او ندارد و نمی‌توان از طریق این آزمون برای دانشجو مشخص نمود که چه مطالبی را خوب نیاموخته بنابراین در جهت تصحیح روند یادگیری کاربرد کمتری دارد؛ زیرا دیگر دانشجو دوره آموزشی را گذرانده است. اما در جهت اصلاح روند تدریس در دوره‌های بعد قابل استفاده است.

از آنجا که آزمون براساس طرح درس^۱ طراحی می‌شود باید راجع به سه مطلب زیر تصمیم گرفت:

۱- چه کسی ارزیابی می‌شود؟

۲- چه وقت ارزیابی انجام می‌گیرد؟

۳- چه نوع سؤالاتی بکار می‌رود؟

برای انجام یک ارزیابی خوب از روشها و ابزارهای مختلف که متناسب با هدف طراحی شده می‌باشد، استفاده می‌گردد:

۱- برای ارزیابی دانش از آزمونهای کتبی و شفاهی استفاده می‌شود.

۲- برای ارزیابی نگرش و مهارت از پرسشنامه‌های تهیه شده بر مبنای مقیاسهای درجه‌بندی شده^۲، و فهرستهای بازبینی^۳ استفاده می‌شود. در راستای ارزیابی حیطه‌های مذکور، رویکردهای زیر وجود دارد:

الف- از دانشجو خواهیم تا علایق و تواناییهای خود را بیان کند (روش گزارش شخصی^۴).

ب- از دانشجو خواهیم تا کار معینی را با استفاده از توانایی خود انجام دهد (روش آزمایشی^۵).

ج- از گروهی خواهیم تا وقتی دانشجو مشغول فراگیریهای روزمره و فعالیتهای دوره درسی است بر او نظارت کرده، به اعمال او امتیاز دهند که ارزیابی همتا^۶ (یا هم‌آزمایی) نامیده می‌شود. البته گاهی از خود دانشجو نیز می‌توان در این امر استفاده نمود. خودآزمایی^۷ جزو این رویکرد قرار می‌گیرد.

کلیه رویکردهای فوق، در فصل‌های بعد به تفصیل توضیح داده خواهد شد. البته باید

1- Lesson Plan

3- Check List

5- Experimental Method

7- Self Assessment

2- Rating Scale

4- Self Report Method

6- Peer Assessment

به خاطر داشت در صورتی ارزیابی استاندارد خواهد بود که پرسشها را مدرس درس یا فردی متخصص در آن موضوع خاص طرح نماید که با طرح درس و اهداف آموزشی، آشنایی کامل دارد.

فواصل زمانی ارزیابی دانشجو بستگی به نوع دوره آموزش و تصمیم مسئولان آن دارد. به طور کلی بر اساس دفعات برگزاری آزمون، دوره‌های آموزشی را به سه دسته تقسیم می‌کنند:

الف- در بعضی دوره‌های آموزشی دانشجویان فقط در یک آزمون به نام پایان نیمسال^۱ شرکت می‌کنند.

ب- در برخی دوره‌ها دانشجویان همواره تحت نظارت و ارزیابی هستند^۲.

ج- در بین این دو روش، دوره‌هایی وجود دارد که در آن هفته به هفته، ماه به ماه یا نیمسال به نیمسال از دانشجو امتحان به عمل می‌آید. این نوع ارزیابی را مداوم^۳ گویند. مزایای روش اخیر عبارت است از:

۱- به دلیل انجام ارزیابیهای متعدد، خطا در هر یک اهمیت کمتری یافته، اشتباهات اندازه‌گیری هر امتحان در نتیجه کلی تأثیر کمتری می‌گذارد. این شیوه ارزیابی مداوم به نظر پایاتر^۴ می‌رسد^۵.

۲- ترس و اضطراب دانشجویان از امتحان پایان نیمسال کاهش می‌یابد.

۳- چون دانشجویان در طول دوره، پیوسته امتحان می‌دهند، از همان ابتدا سعی بیشتری می‌کنند. در صورتی که در سایر موارد فقط یک تلاش شدید و ناگهانی در پایان دوره برای آنها کفایت می‌کند.

۴- اگر دانشجویی در یک امتحان نتایج خوبی کسب نکرده باشد فرصت جبران آن را دارد. این نوع ارزیابی، راهنمایی بیشتر و خط مشی واضح‌تری، هم به استاد و هم به دانشجو می‌دهد.

۵- دانشجویان در خلال دوره خواهند فهمید که به چه معیارهایی باید دست یابند.

البته اشکال این نوع امتحان آن است که وقت و نیروی بیشتری نیاز دارد. اما با توجه به فواید آن، اشکال فوق قابل چشم‌پوشی است.

1- Final

2- Incourse Assesment

3- Frequent or Continuous

4- Reliable

۵- Reliability پایایی یا قابلیت اعتبار آزمون، نشان‌دهنده وجود نتایج مشابه در اجراهای مکرر یک آزمون، برای هر فرد شرکت‌کننده می‌باشد؛ که به‌طور کامل‌تری در فصل مربوط شرح داده خواهد شد.

ذکر این نکته ضروری است که ارزیابی مداوم را می‌توان به صورت یک سلسله امتحانات کتبی یا نظارت بر فعالیتهای دانشجو در بخش یا آزمایشگاه انجام داده، از نتایج و نمرات آن برای سه منظور استفاده نمود:

اول- اطمینان از موفقیت دانشجو در گذراندن دوره؛

دوم- ارائه نوعی بازخورد به دانشجو: با استفاده از نتایج، نقصهای یادگیری برای دانشجو مشخص شده، سطح یادگیری او ارتقا می‌یابد؛

سوم- دستیابی استاد به یک داوری دقیق و باثبات در مورد آنچه دانشجو آموخته است. از انواع دیگر ارزیابی می‌توان به خودآزمایی و ارزیابی همتا اشاره کرد. خودآزمایی نوعی ارزیابی دانشجو از خود است؛ یعنی دانشجو با استفاده از معیارهایی که از طرف نظام آموزشی در اختیار او قرار می‌گیرد خود را مورد سنجش قرار می‌دهد. این شیوه بخصوص در برنامه‌های آموزش ضمن خدمت، کاربرد دارد که در فصول بعدی به تفصیل بیان می‌شود.

ارزیابی همتا جایگزینی برای خودآزمایی است که در این نوع ارزیابی، دانشجویان یکدیگر را می‌آزمایند. این روش برای تشخیص تعداد دانشجویانی که در انتهای دوره، رد یا قبول شده‌اند اصلاً مناسب نیست، اما برای کمک به یادگیری آنها شیوه خوبی است. در اغلب موارد، دانشجویان از یکی از دوستان خود می‌خواهند آنها را بیازمایند. شما می‌توانید این عمل را رهبری و تشویق کنید. مثلاً: ممکن است دستورالعملهای کتبی برای انجام کار عملی به دانشجویان بدهید، سپس یکی از آنها، آن کار را انجام دهد و سایرین او را نظاره و نقد نمایند. البته لازم است به طور پیوسته جای مجری با سایرین عوض گردد.

مثال: از یکی از دانشجویان بخواهید انقباضات ایزومتریک و ایزوتونیک را به طور عملی آزمایش کند و بقیه به دقت نگاه کرده قضاوت نمایند. دستورالعملهای کتبی را می‌توانید از تجارب شخصی یا از منابع موجود استخراج کنید.

با استفاده از ارزیابی همتا - چون دانشجویان از قضاوتهای یکدیگر می‌آموزند - سعی می‌کنند تا روش خود را بهبود بخشیده، عملکرد بهتری داشته باشند و به عبارت دیگر، سعی می‌کنند در برابر همکلاسهای خود نقصانی نداشته باشند.

ویژگیهای ارزیابی خوب

۱- کارآمد بودن در اجرا و تصحیح. آزمونی که ترتیب می‌دهید باید در مدت زمان مناسبی اجرا شده، نمره‌گذاری آن نیز به آسانی صورت گیرد. تفسیر نتایج آزمون هیچ ابهامی نداشته، به

طور دقیق سطوح یادگیری را بسنجد و روایی^۱ و پایایی بالایی داشته باشد.

کارآمد بودن آزمون را به هنگام انتخاب سؤالات در نظر بگیرید. به عنوان مثال اگر تعداد دانشجویان زیاد باشد و قصد دارید سؤالات را برای بار دیگر به کار برید نوع چند-گزینه‌ای^۲ کارآمدتر است؛ ولی اگر تعداد دانشجویان کم باشد برای بار دوم قابل استفاده نباشد آزمونهای تشریحی نیز کارآمد است؛ زیرا تهیه سؤالات تشریحی نسبت به چند گزینه‌ای آسانتر و سریعتر صورت می‌پذیرد ولی تصحیح سؤالات چند گزینه‌ای خیلی سریعتر از تشریحی است. از نظر طول مدت اجرا، آزمونهای تشریحی و چند گزینه‌ای یکسان بوده، هیچکدام بر دیگری برتری ندارد.

۲- قابل پیش‌بینی بودن برای دانشجو. هدف آموزش و ارزیابی باید یکسان باشد. هدفهایی را که در ابتدای تدریس در نظر می‌گیرید باید در طول آموزش، ارزیابی، نمره‌گذاری و ارزشیابی همواره مدنظر قرار گیرد. در این صورت دانشجو می‌تواند سؤالات امتحانی را پیش‌بینی کند. اگر بعد از اجرای آزمون، دانشجویان بگویند: «انتظار چنین سؤالاتی را نداشتیم» شیوه ارزیابی سؤال برانگیز بوده است.

۳- جامع و مختصر بودن. ارزیابی خوب باید جامع باشد؛ یعنی ارتباط مستقیم با اهداف آموزشی داشته، همه آنها را در بر گرفته، مفید، دقیق، روشن و مرتبط با مسائل مهم باشد. همچنین ارزیابی باید فاقد مشکلات اجرایی، متناسب با هزینه، وقت و امکانات موجود باشد.

۴- بازخورد برای استاد. اطلاعات کسب شده از ارزیابی باید در جهت بهبود تدریس و رفع اشکالات احتمالی به استاد کمک کند.

۵- کمک به یادگیری دانشجو. اولاً ارزیابی، دانشجو را وادار به فعالیت بیشتر می‌کند و در نتیجه وی بهتر می‌آموزد. ثانیاً ارزیابی به دانشجو نشان می‌دهد که کدام قسمت درس نیاز به مطالعه بیشتری دارد تا یادگیری آن کاملتر شود.

اگر شما هم ضمن دوره‌های آموزشی، پیوسته امتحاناتی از دانشجو گرفته، به او بگویید که به طور دقیق چه چیزی را نفهمیده است، وی در مورد کیفیت کاری خود نوعی بازخورد به دست می‌آورد و می‌تواند تواناییهای خود را بهبود بخشد.

به عنوان مثال: جدول زیر نتایج ارزیابی پنج دانشجو را نشان می‌دهد، که یک آزمون چهار قسمتی را در میان دوره گذرانده‌اند:

۱- Validity. روایی یعنی یک آزمون همان چیزی را بسنجد که برای آن منظور طراحی شده است.

نام دانشجو	قسمت ۱	قسمت ۲	قسمت ۳	قسمت ۴
الف	+	-	-	-
ب	+	+	-	+
ج	+	+	-	-
د	+	+	+	+
هـ	+	+	-	+

+ : نتیجه مطلوب - : نتیجه نامطلوب

اگر شما جای استاد این دانشجویان بودید چه می‌کردید؟ به نظر می‌رسد نتایج قسمت اول رضایت بخش است. در مورد قسمت دوم به دانشجوی الف توضیح دهید که چرا کار او خوب نبوده است و چگونه می‌تواند آن را بهتر کند و در صورت امکان بعداً دوباره وی را در مورد قسمت دو ارزیابی کنید. نتایج قسمت سه نشان می‌دهد که فقط یک دانشجو به درک لازم رسیده است و شاید نیاز باشد این قسمت دوباره تدریس شود. در اینجا شما در مورد توانایی کاری خود به نوعی بازخورد دست یافته‌اید و نحوه تدریس خود را به شیوه‌ای مطلوب تغییر خواهید داد. در مورد قسمت چهار احتمالاً تکرار تدریس برای همه کلاس اتلاف وقت خواهد بود؛ زیرا تعداد دانشجویان ضعیف چندان زیاد نیست. ارزیابی به این شیوه بسیار وقت‌گیر است؛ اما یکی از با ارزش‌ترین کارهایی که به عنوان یک استاد می‌توانید انجام دهید راهنمایی تک تک دانشجویان است.

تعیین روش ارزیابی

برای دانستن اینکه چه شیوه‌ای را در چه زمانی باید به کاربرد، ابتدا باید دانست که هر روش چه مزایا و معایبی دارد و هر شیوه کدامیک از مقاصد اجرای آزمون را برآورده می‌کند. در حقیقت، تعیین‌کننده نوع آزمون، اهدافی است که در اولویت قرار می‌گیرد. پس از آن باید دید کدام آزمون هدف موردنظر را بیشتر تأمین می‌کند.

مثال:

الف: در انتهای دوره کارورزی جراحی، یک امتحان کتبی برگزار شد. در این امتحان دانشجویان هفت سؤال تشریحی را ظرف سه ساعت پاسخ می‌دادند. سپس از هر دانشجو ظرف پانزده دقیقه امتحان شفاهی گرفته شد.

ب: در خلال یک دوره آموزشی فیزیوپاتولوژی، دو هفته یکبار دانشجویان بیست سؤال چند گزینه‌ای را در مورد عناوینی همچون: نشانه‌های بیماری، شیوه‌های درمان و پیشگیری از بیماری، پاسخ داده با توجه به پاسخهای استاد، خودشان اوراقشان را تصحیح کردند.

ج: کارورزان در حال کارآموزی بخش گوش، حلق و بینی به مدت یک ماه فعالیت خود را تحت نظارت یک استاد ماهر انجام دادند و ناظر، گزارشی از کار آنها تهیه کرد.

حال به بررسی این سه امتحان توجه کنید:

کمک در یادگیری	پایایی	روایی	به صرفه بودن از نظر زمانی	روش
ضعیف	ضعیف	ضعیف	ضعیف	«الف»
خوب	بسیار خوب	بسیاری از مهارت‌های مهم* را نادیده می‌گیرد	خوب	«ب»
خوب	متوسط	بسیار خوب	ضعیف	«ج»

* مهارت‌های مهم: مهارت‌هایی است که دانشجو در شغل آینده خود به آنها نیاز دارد.

الف: روش «الف» تقریباً در تمام موارد ضعیف است. برای برگزاری آزمون و تصحیح اوراق امتحان تشریحی، زمان طولانی صرف شده، ضمناً مهارت‌هایی که در این امتحانها سنجیده می‌شود چندان مهم نیست. همچنین نمره دادن در امتحان تشریحی و شفاهی پایایی کمی دارد.

ب: تهیه سؤالات و ترتیب دادن امتحانات چند گزینه‌ای خیلی طول می‌کشد اما سؤالات را می‌توان سال به سال (با تغییرات اندکی) به کار برد و سرعت نمره‌گذاری کرد. ارزیاب می‌تواند تواناییهای مهم را اندازه بگیرد به شرطی که سؤالات، دقیق و براساس موضوعاتی طرح گردد که به دانشجویان تعلیم داده شده است. پایایی این امتحانات بسیار خوب است بنابراین در نمره‌دهی اشتباه کمتری پیش می‌آید. دانشجویان با دادن نمره به کارهای خود و شناخت دقیق اشتباهاتی که مرتکب شده‌اند آنها را برطرف می‌کنند.

ج: این شیوه زمان زیادی نیاز دارد؛ زیرا ناظر باید گزارش شخصی بنویسد ولی مهارت‌های مهم را می‌آزماید. پایایی نیز ممکن است کم باشد چون ناظران مختلف، معیارهای

گونگون را در نظر می‌گیرند. اما این ارزیابی به طرز بسیار مؤثری به یادگیری کمک می‌کند.

با توضیحات فوق اگر به دنبال برگزاری آزمونی باثبات هستیم بهترین روش، شیوه دوم است و اگر به دنبال آزمونی هستیم که باید مهارت‌های مهم را به خوبی اندازه‌گیری کند یا روا^۱ باشد باید از شیوه سوم استفاده کنیم؛ یعنی هدف از برگزاری آزمون و اولویت‌بندیهای ما نوع آزمونی را که باید به کار رود مشخص می‌کند.

باید توجه داشت که ارزیابی کردن آنچه شخص می‌تواند انجام دهد آسان، اما ارزیابی آنچه قصد دارد در آینده انجام دهد مشکل است. بخصوص در حرفه پزشکی، سنجش معلومات کنونی که دانشجو برای کسب نمره بروز می‌دهد سهل است. اما آیا بدین وسیله تعهد دانشجو در برابر جان بیماران و وجدان اخلاقی او نیز سنجش پذیر است؟

اندازه‌گیری^۲

پس از ارزیابی دانشجویان باید نتایج ارزیابی استنتاج گردد. یعنی باید به آنچه طی فرآیند ارزیابی بدست آمده، اعدادی واقعی نسبت داد تا افراد مورد ارزیابی، قابل مقایسه شوند و در نهایت براساس ارزیابی به ارزشیابی و قضاوت پرداخت. در واقع در سیر از ارزیابی به سمت ارزشیابی، فرآیند میانه‌ای به نام اندازه‌گیری وجود دارد که طی آن به نتایج ارزیابی، ارزش عددی می‌دهد.

اندازه‌گیری تعاریف مختلفی دارد که دو تعریف جامع‌تر آن عبارتند از:

۱- اندازه‌گیری یک فرآیند منظم برای اختصاص دادن اعداد به اشیا و رویدادها بر اساس قوانین تعیین شده و به عبارت دیگر کمی کردن یک متغیر پیوسته مثل هوش، استعداد یا طول است. در این تعریف نکات زیر وجود دارد:

الف) فرآیند اندازه‌گیری یک «فرآیند منظم» است یعنی روشهایی که استفاده می‌شود باید کاملاً روشن و استاندارد باشد تا هر شخص باتجربه‌ای بتواند آن را به کار گرفته، به همین نتایج دست یابد. این روشها باید قابل آموزش به دیگران نیز باشد.

ب) «کمی کردن»؛ یعنی مثلاً: اگر $A > B$ و $B > C$ در نتیجه $A > C$ خواهد بود در این صورت بین A و B و C ارتباط کمی معنی‌داری وجود دارد. نتایج اندازه‌گیری نیز باید دارای تفاوت‌های کمی معنی‌دار باشد. یعنی از لحاظ کمی و مقداری باید با هم قابل مقایسه

باشد.

ج) «متغیر»، خصوصیتی است که ارزشهای عددی مختلف دارد، مثل قد و وزن.
 د) «متغیر پیوسته» متغیری است که با اعداد حقیقی دارای تناظر یک به یک است و کلیه اعداد حقیقی (اعداد منهای بی‌نهایت تا مثبت بی‌نهایت) را می‌توان به آن نسبت داد. در مقابل متغیر پیوسته، متغیر گسسته قرار دارد که از راه شمارش بدست می‌آید و قابل انتساب به اعداد طبیعی (اعداد از ۱ تا مثبت بی‌نهایت) می‌باشد. مثل تعداد دانشجو که سه یا چهار نفر دانشجو صحیح ولی ۳/۵ دانشجو صحیح نیست.

۲- اندازه‌گیری، مجموعه‌ای از حقایق یا اطلاعات است که به وسیله تجربه، آزمایش و مشاهده استنباط شده، بیشتر در مورد اشیای مادی به کار می‌رود. اما امروزه فرض بر این است که هر چیز انتزاعی (هوش، استعداد، پیشرفت تحصیلی) را نیز می‌توان اندازه گرفت.

باید توجه داشت اندازه‌گیری و ارزیابی به منظور بیان کمیت و کیفیت خصیصه‌ها به کار می‌روند، اما بیان دقیق ویژگیهای رفتاری آدمی (مانند هوش، یادگیری، نگرش، دانش) بر خلاف متغیرهایی چون قد و وزن، امکان‌پذیر نیست و صرفاً می‌توان تخمینی از آنها به دست آورد، بدین دلیل است که معمولاً از واژه «ارزیابی» در این موارد استفاده می‌گردد.

انواع اندازه‌گیری

۱- اندازه‌گیری ملاک مرجع^۱

این اندازه‌گیری براساس ملاک مطلق، طراحی و انجام می‌شود؛ یعنی همه بر اساس یک حد نهایی سنجیده می‌شود نه نسبت به یکدیگر. مثلاً همه نسبت به نمره بیست سنجیده می‌شود. این نوع اندازه‌گیری تواناییهای خود دانشجو را بدون مقایسه با دیگران می‌سنجد. مثلاً وقتی فردی ۸۰ درصد سؤالات آناتومی را درست جواب داده به ۸۰ درصد از تواناییهای مورد نظر رسیده است. البته در اینجا مهم نیست که چه تعداد افراد، تواناییهای بیشتر یا کمتر از او دارند.

۲- اندازه‌گیری هنجار مرجع^۲

این اندازه‌گیری بر اساس ملاک نسبی طراحی و انجام شده، هدف آن تعیین جایگاه و رتبه دانشجو در مقایسه با دیگران است. در این نوع اندازه‌گیری، ارزش عددی و اندازه‌گیری

صرف مطرح نیست، بلکه میزان بزرگی و کوچکی آن نسبت به سایرین مطرح است. مثلاً: اگر گفته شود فرد «الف» از درس بیوشیمی نمرهٔ هیجده گرفته مفهومی ندارد بلکه باید تعیین شود که این فرد در چه رتبه‌ای از کلاس قرار گرفته است و چند درصد از دانشجویان بالاتر و چند درصد، پایین‌تر از او هستند.

مقیاس ۱

مقیاس، واحد اندازه‌گیری است و از طریق مقایسه با مقیاس، موضوع اندازه‌گیری شده ارزش عددی پیدا کرده، قابل بیان و فهم می‌شود مثلاً کیلوگرم یا میلی‌متر جیوه مقیاس انواع مختلف دارد که عبارتند از:

۱- **مقیاس اسمی**^۱. پایین‌ترین یا کم ارزش‌ترین سطح مقیاس، مقیاس اسمی است و آن طبقه‌بندی اشیا و رویدادها از طریق اختصاص شماره یا نام به آنها می‌باشد. مثلاً گروه‌های خونی یا شماره‌های روی پیراهن بازیکنان فوتبال.

۲- **مقیاس رتبه‌ای**^۲. این مقیاس، تعیین‌کنندهٔ موقعیت فرد بر حسب یک صفت در میان سایرین است. مقیاس رتبه‌ای چندان دقیق نیست، چون مقیاسی که نفر اول و دوم را جدا می‌کند، با مقیاسی که نفر سوم و چهارم را جدا می‌کند برابر نیست. مثلاً: مثلاً چهار نفر با نمرات ۱۰۰، ۹۷، ۹۰، ۸۰ به ترتیب رتبه‌های اول تا چهارم را کسب می‌نمایند و با آنکه فاصله آنها از نظر عددی برابر نیست، ولی از لحاظ رتبه‌ای برابر است.

۳- **مقیاس فاصله‌ای**^۳. در این مقیاس واحدها برابر و نقطهٔ صفر با موقعیت قراردادی (نه مطلق) وجود دارد. مثلاً دماسنج که صفر آن نشانهٔ فقدان حرارت نیست. نمرهٔ امتحانی دانشجویان نیز از این نوع است؛ یا جواب ندادن به سؤالات امتحان فیزیولوژی به این معنی نیست که فرد هیچ اطلاعی در این زمینه ندارد.

۴- **مقیاس نسبی**^۴. که پیشرفته‌ترین نوع اندازه‌گیری است دارای واحدهای برابر و صفر مطلق بوده، قابل تعمیم می‌باشد، مثلاً وزن و حجم. در این مقیاس مثلاً اگر فردی از پنجاه سؤال به چهل سؤال جواب صحیح بدهد می‌توان برآورد کرد که از ۲۵ سؤال به ۳۶ سؤال پاسخ درست می‌دهد.

1- Scale

2- Nominal Scale

3- Ordinal Scale

4- Interval Scale

5- Ratio Scale

ارزشیابی

ارزشیابی معمولاً هدف عالی و نهایی فرآیند ارزیابی و اندازه‌گیری است، یعنی آنچه به سنجش معنی می‌دهد ارزش دادن به آن است. ارزشیابی در واقع تجزیه و تحلیل نتایج آزمون یا هر اندازه‌گیری دیگری است که بر پایه آن بتوان به برنامه‌ریزی‌های صحیح و بازخوردهای مناسب و عملی دست یافت. به بیان دیگر، آزمون در صورتی روال منطقی خود را دارد که از راه ارزشیابی بتوان از نتایج آن به برخی راهنماییها، راهکارها و اصلاحات دست یافت. مهمترین تعاریفی که می‌توان برای ارزشیابی ارائه کرد عبارتند از:

الف) براساس نظر بیبای^۱ ارزشیابی، فرآیند جمع‌آوری و تفسیر نظامدار شواهدی است که در نهایت به قضاوت ارزشی با چشمداشت به اقدامی معین می‌انجامد.

در این تعریف چند نکته با اهمیت وجود دارد:

۱- «نظامدار»: یعنی کاملاً روشن و استاندارد به طوری که هر شخص باتجربه‌ای بتواند از آن استفاده کند.

۲- «تفسیر شواهد»: یعنی نشان دهیم آیا آنچه ارزشیابی شده در نهایت توانسته است به کیفیت لازم برسد یا خیر.

۳- «قضاوت ارزشی»: یعنی بررسی هدفها و مقیاسهای ارزشیابی.

۴- «با چشمداشت به اقدامی معین»: یعنی استفاده از نتایج ارزشیابی برای تصمیم‌گیریهای آینده.

در ارزشیابی می‌توان قضاوت و نظر شخصی خود را دخالت داد، چون ارزشیابی بیشتر جنبه توصیفی و کیفی دارد تا جنبه عددی و کمی. همانطور که گفته شد ارزیابی و اندازه‌گیری، مقدمه ارزشیابی است. اگر اطلاعات بدست آمده، تجزیه و تحلیل گردد عمل ارزشیابی صورت گرفته است. مثلاً: در ارزیابی، نمره آناتومی فردی، بیست برآورد شده است، حال تصمیم استاد مبنی بر استفاده از این دانشجو برای تدریس درس عملی آناتومی نوعی ارزشیابی می‌باشد.

ب) براساس نظر گیلبرت^۲ ارزشیابی یک فرآیند مستمر بر مبنای ضوابط بدست آمده از کار گروهی است که هدف آن سنجش فعالیت یادگیرنده، کارآیی استادان و کیفیت برنامه می‌باشد.

در این تعریف چند نکته با اهمیت وجود دارد:

- ۱- «فرآیند مستمر»: یعنی: انجام ارزشیابی همگام با آموزش و تدریس.
- ۲- «ضوابط بدست آمده از کارگروهی» یعنی: استفاده از تجربیات قبلی به منظور تجزیه و تحلیل نتایج فعلی. اگر آزمون برای اولین بار اجرا می‌شود ارزشیابی منطقی آن چندان عملی نخواهد بود.

ارزشیابی با ارزیابی و اندازه‌گیری متفاوت است. ارزشیابی فرآیندی کلی‌تر محسوب می‌گردد. علاوه بر آن هم توصیف کمی و هم کیفی از رفتار بدست می‌دهد؛ از دیگر سو، اندازه‌گیری ارزیابی پایه ارزشیابی هستند. مثلاً: وقتی فشار خون شخصی 160 mmHg است این رقم حاصل اندازه‌گیری است. حال اگر با توجه به این رقم بگوییم که او احتمالاً در آینده به بیماری قلبی مبتلا می‌شود ارزشیابی انجام داده‌ایم. تفسیر نمرات دانشجویان و یا رتبه‌بندی آنها در رده‌های عالی خوب، متوسط، ضعیف ارزشیابی محسوب می‌گردد ولی نمره دادن نوعی اندازه‌گیری است.

در ارزشیابی توجه به نکات زیر ضروری است:

- ۱- **تعیین هدف ارزشیابی:** هدف باید کاملاً واضح و منطبق بر اهداف آموزشی از پیش تعیین شده باشد.
- ۲- **انتخاب وسیله مناسب:** برای فراهم آوردن اطلاعات باید وسیله مناسب انتخاب کرد تا سنجش با دقت انجام پذیرد. مثلاً برای سنجش مهارت دانشجو در اندازه‌گیری فشار خون، آزمون کتبی یا شفاهی مناسب نیست، بلکه آزمون عملی کارآمد است.
- ۳- **تحلیل و تفسیر اطلاعات کسب شده:** جهت بهبود کیفیت آموزش، اطلاعات بدست آمده باید تجزیه، تحلیل و تفسیر گردد. در این امر توجه به انتخاب ملاکهای مناسب برای قضاوت از اهمیت برخوردار است.

انواع ارزشیابی^۱

ارزشیابی مشتمل بر چند نوع است که از این میان می‌توان موارد زیر را نام برد:

- ۱- ارزشیابی مجموعی^۲؛
- ۲- ارزشیابی تکوینی^۳؛

۱- یا توجه به آن که مبنای ارزشیابی، اندازه‌گیری و ارزیابی است، این دو مفهوم نیز به انواع ذکر شده در این بخش قابل تقسیم هستند.

۳- ارزشیابی تشخیصی^۱

۱- ارزشیابی مجموعی

این ارزشیابی در پایان هر نیمسال اجرا شده، هدف آن تعیین میزان یادگیری دانشجویان در طول یک دوره آموزشی است، که به منظور نمره دادن به دانشجو، قضاوت در مورد کار استاد و اندازه‌گیری سرعت و دقت دانشجو صورت می‌گیرد. در ارزشیابی مجموعی فقط نمونه‌ای از محتویات مندرج در جدول مشخصات^۲ آزمون طرح می‌شود.

۲- ارزشیابی تکوینی

این ارزشیابی در طول زمان تدریس، به منظور رفع مشکلات یادگیری دانشجویان و برطرف کردن نواقص آموزش صورت پذیرفته، جزئی از فرآیند یادگیری است که نباید آن را در نمره دادن و طبقه‌بندی دانشجویان دخالت داد. اهداف نهایی این نوع ارزشیابی عبارتند از:

الف) تعیین میزان رسیدن دانشجو به مهارت لازم؛

ب) تعیین اینکه فرآیند تدریس تا چه حد نیازمند تغییر است که جزء مهمترین اهداف است؛

ج) پیش‌بینی نتایج ارزشیابی مجموعی.

روش تهیه آزمونهای ارزشیابی تکوینی:

جدول مشخصات را تهیه کنید، بعد از تجزیه و تحلیل محتوای درسی که در طی مدت معینی باید آموخته شود، تعیین کنید کدامیک از عناصر درسی مهم بوده، کدامیک اهمیت چندانی ندارند. آنهایی را که چندان مهم نیستند باید در جدول بنویسید، اما برای طرح سؤال می‌توانید از آنها صرف‌نظر کنید. پس از برگزاری امتحان و تصحیح سؤالات، در صورتی که پاسخنامه‌ای ۸۰ درصد یا ۸۵ درصد پاسخ صحیح داشته باشد دانشجو به مهارت لازم رسیده است. به علاوه با نشان دادن اشتباهات به دانشجو، بازخوردی برای او حاصل می‌شود و وی می‌تواند حیطه‌هایی را که در آنها ضعف داشته، شناسایی کند و نسبت به

1- Diagnostic Evaluation

۲- جدول طرح آزمون (مشخصات) شامل دو ستون است: ستون عمودی موضوعات مورد امتحان و ستون افقی سطح مورد انتظار که دانشجو باید به آن برسد (دانش، درک، کاربرد،...). در مقابل هر یک از این ستونها تعداد سؤال لازم را می‌نویسند و از روی آن سؤال طرح می‌کنند. برای توضیحات بیشتر به پایان این فصل مراجعه شود.

تقویت آموخته‌های خود در آن زمینه‌ها تلاش بیشتری نماید.

۳- ارزشیابی تشخیصی

پیش از شروع درس، برای تعیین اطلاعات قبلی دانشجویان و تشخیص اشکالات و نارساییهای آنها، ارزشیابی تشخیصی صورت می‌گیرد. ارزشیابی تشخیصی می‌تواند بر نتایج ارزشیابیهای مجموعی قبلی متکی باشد و براساس آن دانشجویان تقسیم‌بندی شوند. می‌توان ارزشیابی تشخیصی را از طریق پیش‌آزمون انجام داد. پیش‌آزمون وسیله‌ای است که میزان اطلاعات قبلی دانشجو را در زمینه آنچه باید بیاموزد، اندازه می‌گیرد و با توجه به آن، نقطه شروع آموزش تعیین می‌شود؛ اگر چه نمی‌توان از معدل سال قبل برای تشخیص نقطه شروع تدریس استفاده کرد، می‌توان مثلاً از نمره فیزیولوژی (۱) برای شروع تدریس فیزیولوژی (۲) استفاده نمود؛ زیرا معدل نتیجه مجموعه امتحانات است. در امتحانات ارزشیابی تشخیصی بهتر است دانشجو ناشناس بماند بنابراین، نمره با شماره رمز اعلام می‌گردد. به هر حال نتایج ارزشیابی تشخیصی، به دانشجو در رفع نقاط ضعف خود کمک می‌کند.

آزمون

با توجه به مطالب ذکر شده می‌توان ارتباط بین آزمون، اندازه‌گیری و ارزشیابی را به صورت زیر ذکر کرد:

آزمون یکی از ابزارهای اندازه‌گیری و اندازه‌گیری زیر بنای ارزشیابی است. گرچه این سه، هر یک شرط لازم برای دیگری است، شرط کافی نیست. مثلاً در ارزشیابی لازم است آزمون و اندازه‌گیری قبلاً صورت گرفته باشد اما فرآیندهای دیگری نیز احتیاج دارد.

اگر آزمون را ساخت و اجرا، اندازه‌گیری را نمره‌دهی و ارزشیابی را قضاوت برپایه نمرات در نظر بگیریم متوجه خواهیم شد که اینها زیر مجموعه یکدیگرند و تا یکی انجام نشود صعود به دیگری مقدور نیست. این ارتباط را می‌توان در قالب شکل ۱ نمایش داد:

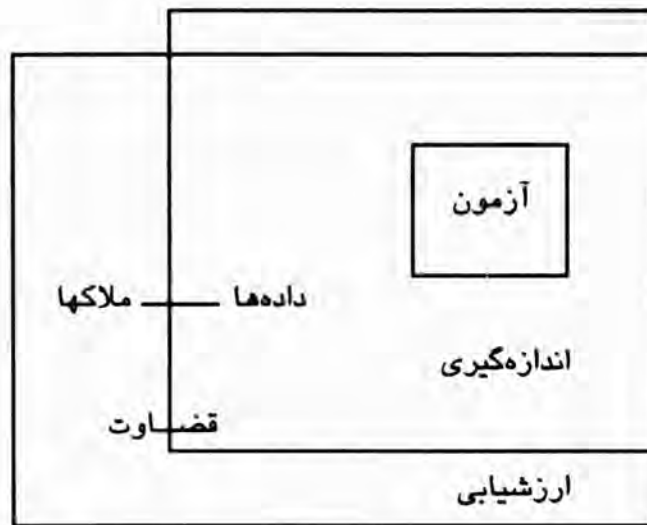
از سوی دیگر طراحی آموزشی دارای چهار مرحله است:

۱- تعیین اهداف درس؛

۲- چگونگی رسیدن به این اهداف؛

۳- گزینش محتوا، شیوه‌ها و رسانه‌های مناسب؛

۴- تعیین نظام ارزشیابی یا سنجش میزان پیشرفت تحصیلی.



شکل ۱

بنابراین هر جا سخن از آموزش است، آزمون و ارزشیابی امری حتمی و اجتناب‌ناپذیر می‌باشد. هر فرد آزمون‌گیرنده قبل از آنکه یک آزمون‌گیرنده خوب باشد باید بداند آزمون چیست، با چه اهدافی برگزار می‌شود و ویژگیهای یک آزمون خوب کدام است که در این قسمت این مطالب توضیح داده می‌شود.

تعریف آزمون:

در مورد آزمون تعاریفی گفته شده است که از آن جمله است:

۱- آزمون وسیله‌ای عینی^۱ و استاندارد شده است که برای اندازه‌گیری نمونه‌ای از رفتارها یا ویژگیها استفاده می‌شود. منظور از «عینی بودن» این است که نتیجه اندازه‌گیری آزمون یا به عبارتی اجرا، نمره دادن، تعبیر و تفسیر نتایج آزمون به دور از قضاوت و دیدگاههای شخصی آزمون‌گیرنده و بر اساس قواعدی معین و مشخص باشد. منظور از «استاندارد شده» این است که آزمون قبلاً در مورد اشخاص دیگر اجرا شده، نتایج آن مورد تجزیه و تحلیل قرار گرفته، روایی و پایایی آزمون تعیین شده باشد، مانند آزمونهای استاندارد شده هوش و یا پیشرفت تحصیلی.

۲- آزمون وسیله مخصوص اندازه‌گیری در آموزش و مجموعه‌ای از سؤالات است که برای عرضه به یک یا چند دانشجو در وضعیت خاص طرح‌ریزی شده است.

1- Objective

باید به این نکته توجه داشت که هدف از امتحان به هیچوجه غافلگیر کردن دانشجو نیست بلکه آزمون مجموعه سؤالاتی است که توسط استاد تهیه شده و برای ارزشیابی آموخته‌های دانشجو به کار می‌رود.

در آزمون، میزان تغییرات حاصل در جنبه‌های خاصی از رفتار که موردنظر اهداف آموزشی هستند در فواصل زمانی معین تعیین می‌شود و دانشجویان به صورت یکسان به تعدادی سؤال پاسخ می‌دهند و یا وادار به انجام یک فعالیت عملی می‌شوند. از آنجا که هدف، مقایسه کمی توانایی افراد است، ملاک داوری در امتحان، میزان نیل به اهداف درسی است. نتایج امتحان نه تنها استاد را قادر می‌سازد مقایسه‌ای بین دانشجویان انجام دهد بلکه در مورد بررسی کارآیی برنامه و وسایل آموزشی و هدفها نیز به او کمک می‌کند.

اهداف آزمون

استاد در درجه اول باید به صورتی تدریس کند که دانشجو به هدفهای از پیش تعیین شده برسد؛ در درجه بعدی باید پیشرفت تحصیلی را اندازه بگیرد یعنی میزان موفقیت در رسیدن به اهداف تعیین شده را نشان دهد و برای تعیین میزان موفقیت باید آزمونهایی ترتیب دهد. کاربرد کلیه ابزارهای اندازه‌گیری در اجرای آزمون و تفسیر نتایج بدست آمده، جزء مسؤولیت‌های هر استاد است.

نحوه برگزاری آزمون تحت تأثیر هدفهایی است که با انجام آن باید به آنها دست یافت. برای دستیابی به این اهداف صرفاً انجام آزمون کافی نیست بلکه بررسی تفاوت میان پیش آزمون^۱ و پس آزمون^۲ و تجزیه و تحلیل نیز برای رسیدن به این اهداف ضروری است. هدف از آزمون صرفاً مقایسه میان دانشجویان و تعیین تعداد مردودی و قبولی آنها نیست؛ اجرای آزمون و بررسی نتایج آن، بازخوردهایی برای استاد و دانشجویان دارد که جزء اهداف ارزیابی است و در مورد این اهداف، در ابتدای فصل توضیح داده شد. البته وقتی این نتایج از آزمون بدست خواهد آمد که دانشجو نسبت به امتحان نگرش مطلوبی داشته باشد. استفاده از آزمون به عنوان وسیله تهدید، باعث اضطراب در دانشجو شده، او را از یادگیری باز می‌دارد.

موارد استفاده از آزمونها

به طور کلی آزمونها را در این موارد به کار می‌بریم:

۱- تعیین نقطه آغاز تعلیم در دوره بعدی (پیش آزمون)؛

۲- ایجاد انگیزه در دانشجو (آزمون برای انگیزش)^۱؛

۳- تمرین و تکرار مطالب مهم درسی؛

۴- تعیین میزان پیشرفت تحصیلی^۲؛

۵- تعیین میزان موفقیت در تدریس.

۱- تعیین نقطه آغاز تعلیم در دوره بعدی

پیش آزمون باید وضعیت دانشجویان را در ابتدای دوره نشان داده، مشخص کند معلم باید از کجا تدریس را شروع کند. در حقیقت به کمک آزمون سابقه ذهنی و اطلاعات قبلی دانشجو مشخص می‌شود. در این امتحانات معمولاً سؤالات باید کوتاه طرح شود تا نحوه پاسخگویی بتواند سطح وسیع اطلاعات دانشجو را نشان دهد. سؤالات صحیح - غلط^۳ نیز مناسب بوده، برای ساخت، اجرا و تصحیح، زمان کمتری لازم دارد. در این آزمونها نباید سؤالات امتحان راجع به کلمات تخصصی طرح شود، بلکه باید دانسته‌های عمومی فرد را در آن زمینه خاص بسنجد. مثلاً ممکن است دانشجو مطالبی را بداند ولی به دلیل وجود لغات فنی در سؤال، نتواند به آن جواب دهد. زمان این امتحان نیز نباید از پانزده تا بیست دقیقه بیشتر شود.

۲- آزمون برای ایجاد انگیزه

در این آزمونها سؤالاتی مطرح می‌شود که بحث برانگیز بوده، باعث ایجاد مسائل جالبی می‌شود که دانشجو را به بحث در آن مورد یا جستجوی جواب برمی‌انگیزد. مثلاً: مراحل زندگی در نوزاد حیوان با ... شروع می‌شود.

جواب بیشتر دانشجویان تولد^۴ می‌باشد ولی پس از جواب به سؤال، این مسائل مطرح می‌شود: آیا جنین حرکت می‌کند؟ از چه زمانی؟ آیا نفس می‌کشد؟ چگونه؟ آیا ضربان قلب دارد؟ از چه زمانی؟

1- Test for Motivation
3- True - False

2- Educational Progress
4- Birth

پیش آزمونها نیز می‌تواند حاوی چنین سؤالاتی باشد که احتمالاً جواب آنها در طی دوره تدریس مشخص می‌شود.

مثال دیگر: صداهایی که با استتوسکوپ بر روی قفسه سینه شنیده می‌شود، در اثر ... ایجاد می‌شود.

جواب اکثر دانشجویان «بسته شدن دریچه‌های قلب» است. سؤالاتی که به دنبال آن در ذهن دانشجو مطرح می‌شود، عبارتند از: در چه زمانی از دوره قلبی این صداها شنیده می‌شود؟ در چه محلی از قفسه سینه این صداها واضح‌تر شنیده می‌شود؟ چند صدا حین یک دوره قلبی شنیده می‌شود؟ ویژگیهای هر صدا و روش تشخیص آنها از یکدیگر چگونه است؟

۳- آزمون برای تمرین و تکرار

این آزمونها حاوی سؤالاتی است که مطالب مهم درسی را مطرح می‌کند. این آزمون می‌تواند شفاهی یا کتبی باشد. سؤالات کوتاه پاسخ و به دور از بحث در صورتی که شفاهی باشد بهتر است، ولی به دلیل وقت‌گیر بودن مشکل ایجاد می‌کند؛ در ضمن باید سؤالات را از قبل آماده نمود که از بحثهای بی‌نتیجه جلوگیری شود. در صورتی که آزمون کتبی باشد تصحیح اوراق لزومی ندارد. اما دانشجویان اغلب علاقه دارند از نتیجه امتحان آگاهی حاصل کنند تا بتوانند اشتباهات و سوء تفسیرهای خود را شناخته، به اصلاح آن بپردازند.

۴- آزمون تعیین میزان پیشرفت تحصیلی

تعیین پیشرفت تحصیلی اغلب به وسیله نمرات دانشجویان انجام می‌شود. معمولاً چنین استنباط می‌شود که صاحبان نمرات بالاتر پیشرفت بیشتری دارند، در صورتی که باید به عوامل دیگری نیز توجه داشت؛ از جمله: نقاط ضعفی که در امتحان وجود داشته و اینکه برآورد امتحان تا چه حد همه سطوح مهم یادگیری و همه مطالب درسی را شامل شده است؟ آیا روا و پایا بوده است یا خیر؟ به طور مثال: ممکن است دانشجویی بتواند فشار خون بیمار را با فشارسنج اندازه بگیرد ولی نتواند تفاوت فشار سیستول و دیاستول را از نظر تئوری بیان نماید؛ یا دانشجوی دیگری بتواند تفاوت فشار سیستول و دیاستول را توضیح دهد ولی از نحوه تعیین فشار خون بی‌اطلاع باشد. در امتحان کتبی با وجود مهارت بیشتر، دانشجوی اول مردود می‌شود. پس با توجه به همه این عوامل و نه فقط نمره آزمون، باید میزان پیشرفت را اندازه گرفت.

۵- آزمون برای ارزشیابی میزان موفقیت در تدریس

اگر نمرات خوب کلاس زیاد باشد، نشان دهنده موفقیت تدریس است و برعکس. ولی نحوه تدریس تنها عامل مؤثر در نمرات آزمون نیست و عوامل دیگری هم مؤثرند. مثلاً درصد موفقیت زیاد می‌تواند نشان دهنده تدریس خوب و یا تجارب قبلی دانشجو در این مبحث باشد و موفقیت کم می‌تواند نشان دهنده تدریس ضعیف یا اشکال در نحوه طراحی آزمون باشد. پس برای تعیین موفقیت تدریس نباید فقط به نمرات آزمون تکیه کرد. ممکن است دانشجویی از دانش بیشتری برخوردار باشد ولی نتواند از عهده امتحان شفاهی برآید و برعکس.

قواعد کلی طراحی آزمون

برای طراحی آزمون ابتدا باید نوع سؤالات را مشخص نموده، سپس به وسیله جدول طرح آزمون، آن را طرح‌ریزی کرد. البته در پایان، ویرایش و بازبینی نیز لازم است.

انتخاب نوع پرسش مناسب

در انتخاب نوع پرسش مناسب برای استفاده در آزمون‌ها این نکات باید مورد توجه قرار گیرند:

الف) نوع پرسش انتخاب شده. سؤال باید سطوح یادگیری مورد نظر را تا آنجا که ممکن است به طور مستقیم اندازه‌گیری کند.

ب) کیفیت پرسش. هر نوع پرسش بخشی از آموخته‌های دانشجو را در یک زمینه خاص اندازه می‌گیرد و با این حال پرسشهای چندگزینه‌ای و تشریحی تا حدی می‌تواند انواع مختلف سطوح ساده و پیچیده را اندازه بگیرد.

براساس فرآیندهای ذهنی مورد اندازه‌گیری، سؤالات را به انواع زیر طبقه‌بندی می‌کنیم:

- ۱- **بازشناسی**^۱. دانشجو از میان چند مطلب ارائه شده در پرسش، براساس آنچه قبلاً آموخته یا تجربه کرده است به تشخیص و یادآوری می‌رسد. مثلاً: عصب عضله گلو تنوس ماگزیموس کدامیک از اعصاب گلو تنال فوقانی، گلو تنال تحتانی و ایتوراتور می‌باشد؟
- ۲- **فراخوانی**^۲. نوعی یادآوری است که دانشجو بدون دیدن پاسخ صحیح، براساس آموخته‌هایش پاسخ را می‌یابد. مثلاً: اعصاب عضله گلو تنوس ماگزیموس را نام ببرید.

۳- بازسازی^۱. دانشجو علاوه بر به یاد آوردن، نوعی سازماندهی نیز به صورت کتبی یا شفاهی انجام می‌دهد. مثلاً: درگیری عصب گلوتهال تحتانی از نظر کلینیکی چه مشکلی را ایجاد می‌کند؟

۴- کارنمایی. دانشجو نوعی مهارت را که بر اثر تکرار و تمرین به صورت عادت درآمده به اجرا می‌گذارد. مثلاً: در بیمار تخت شش کدامیک از سوفلهای قلبی شنیده می‌شود؟ (که سنجش دانشجو در شنیدن و تشخیص صداهای قلبی است.)

در سطح بازشناسی بهتر است پرسشهای چند گزینه‌ای یا صحیح - غلط یا جورکردنی^۲ را انتخاب کرد و در سطح فراخوانی از سوالات کوتاه پاسخ و سوالهای کامل‌کردنی بهره برد و در سطح بازسازی از پرسشهای تشریحی و در سطح کارنمایی از امتحانات عملی استفاده نمود.

ذکر این نکته ضروری است که هیچ نوع سؤالی را نمی‌توان برای همه موقعیتهای توصیه کرد. برخی سوالات برای اندازه‌گیری گروهی از اهداف ارجح هستند. مثلاً: برای سنجش توانایی در معاینه فیزیکی^۳ نمی‌توان از سوالات چند گزینه‌ای استفاده کرد. همچنین استفاده از چهار یا پنج نوع سؤال، وقت‌گیر خواهد بود. اگر به طور دقیق بیاندیشید و برای نوشتن سوالات وقت کافی به کار برید، دلیلی برای استفاده انواع سوالات در یک آزمون وجود نخواهد داشت. اگر مجبور به انجام این امر شدید، در مورد نحوه پاسخ دادن به سوالات و شیوه ارزش‌گذاری برای دانشجویان توضیح داده، با یک یا دو مثال آنها را راهنمایی کنید.

ج) نحوه انتخاب سؤال براساس روایی، پایایی و عملی بودن. در آزمونهای داخل کلاسی در صورتی که منتهی به کسب مدرک نمی‌شود و صرفاً برای کشف نارسایی‌های یادگیری است، باید آزمون پایا به کار برید؛ زیرا آزمون پایا با تکرار مجدد، نتایج یکسانی بدست می‌دهد و می‌توان نتیجه گرفت که سؤالاتی که اشتباه پاسخ داده شده‌اند حتماً به دلیل وجود اطلاعات ناکافی در آن زمینه بوده و در صورت سؤال و نحوه طرح آن اشکالی نبوده است. در مورد آزمون نهایی یک دوره که منجر به کسب مدرک می‌شود، باید آزمونی به کار برید که هم روا و هم پایا باشد؛ زیرا برای اعطای مدرک باید بدانیم دانشجو مهارت شغلی آینده را به خوبی آموخته است یا خیر. آزمون روا همان هدف

مورد نظر را ارزیابی می‌کند. مثلاً: در پایان دوره تخصص اطفال، گرفتن آزمون در زمینه سکتۀ مغزی ارزش چندانی ندارد؛ چون در نوزادان، سکتۀ مغزی شیوع چندانی ندارد و این آزمون روا نخواهد بود.

د) نحوه انتخاب سؤال براساس میزان مشکل بودن. سؤالاتی که به کار می‌برید باید از نظر میزان مشکل بودن متفاوت باشد، بخصوص اگر هدف، تمیز قائل شدن بین دانشجوی قوی و ضعیف است.

مراحل تدوین سؤالات

- با توجه به مراحل چهارگانه ذکر شده، جهت تدوین سؤالات طرح زیر پیشنهاد می‌گردد:
- ۱- در طی دوره تدریس، هر هفته چند پرسش در زمینه مطالب تدریس شده تهیه کنید.
 - ۲- برای هر یک از هدفهای بدیهی^۱ چند پرسش طرح کنید. همچنین معلوماتی را که برای یادگیری درس جدید لازم است تعیین کرده، برای هر کدام سؤال طرح نمایید.
 - ۳- تعداد پرسشهای تعیین شده، بیش از تعدادی باشد که بعداً در جدول مشخصات گنجانده می‌شود. در این صورت می‌توانید از بین آنها بهترین را انتخاب کنید.
 - ۴- هر پرسش را روی برگۀ جداگانه با اندازه یکسان، به عنوان مثال ۵×۸ سانتی متر بنویسید تا استفاده، تجدیدنظر و بایگانی آنها راحت‌تر صورت پذیرد.
 - ۵- پس از تهیه پرسشها و پاسخها، از افراد مطلع دیگری بخواهید آنها را بررسی نموده، نقاط ضعف آن را تا حد امکان رفع کنید.
- ذکر این نکته لازم است که طرح سؤالات در امتحانات بالینی مشکل است. عواملی چون: کاهش طول مدت بستری، بی‌میلی بیماران برای مراجعه به درمانگاهها جهت درمان سرپایی و بی‌میلی آنها در اجرای مکرر معاینۀ فیزیکی، فراهم آوردن مخلوطی از شرح موردهای^۲ کوتاه و بلند را برای طراحان امتحان مشکل کرده است.

جدول مشخصات آزمون

جدول مشخصات آزمون یا جدول طرح آزمون جدولی است که سطوح یادگیری را به محتوای درسی ربط داده، باعث می‌شود آزمون، سطوح و محتوا را به خوبی در برگیرد.

۱- هدفهای بدیهی اهدافی است که فرض می‌شود دانشجو از پیش آنها را آموخته و به آنها دست یافته است و جزء برنامه آموزشی ما برای تدریس نیست ولی دانستن آنها لازم است.

در این جدول، تعداد سؤالات مربوط به هر یک از سطوح یادگیری، همچنین تعداد سؤال تعلق گرفته به هر مطلب درسی، در هر سطح مشخص می‌شود. به طور کلی جدول مشخصات اهمیت نسبی هر هدف و تعداد سؤالات مربوط به آن را بر اساس میزان اهمیت فوق تعیین می‌کند. حال این سؤال پیش می‌آید که سطوح یادگیری چیست و چگونه تعیین می‌شود؟

سطوح یادگیری

هدفهای یادگیری در حوزه شناختی براساس طبقه‌بندی بلوم^۱ شامل شش سطح است که از ساده‌ترین به پیچیده‌ترین عبارتند از:

۱- بازشناسی و یادآوری^۲؛ ۲- فهمیدن^۳؛ ۳- به کار بستن^۴؛ ۴- تجزیه و تحلیل^۵؛ ۵- ترکیب^۶؛ ۶- ارزشیابی و قضاوت^۷.

برای رسیدن به هر سطحی از رشد فکری سپری کردن مراحل پیشین ضروری است که به صورت مختصر به شرح آن می‌پردازیم.

۱- **بازشناسی و یادآوری.** پایین‌ترین سطح یادگیری دانش است. در این سطح دانشجو باید بتواند از طریق بازشناسی و یادآوری شواهدی ارائه دهد حاکی از اینکه او پدیده‌ای را که در جریان آموزش تجربه کرده، به خاطر سپرده است. یادگیری در این سطح صرفاً جنبهٔ حفظی دارد. البته به خاطر سپردن مفاهیم و اصول را نمی‌توان به کلی بی‌اهمیت شمرد؛ زیرا پایه‌ای برای یادگیری در سطوح بالاتر است.

مثال: دانشجو بتواند بخشهای مختلف سیستم تحریکی هدایتی قلب را نام برده، نحوهٔ فعالیت الکتریکی قلب را توضیح دهد.

۲- **فهمیدن.** یادگیری در این سطح عبارت است از توانایی پی‌بردن به مفهوم یک مطلب و تبیین آن با جملاتی که خود شخص ساخته است، بدون آنکه میان آن مطلب با مطالب دیگر ارتباطی برقرار کند.

مثال: دانشجو بتواند قسمت‌های مختلف فعالیت الکتریکی قلب را با منحنی الکتروکاردیوگرام تطبیق دهد.

۳- **به کار بستن.** یادگیری در این سطح عبارت است از توانایی کاربرد اصول علمی، فرضیه‌ها،

1- Bloom

3- Comprehension

5- Analysis

7- Evaluation and Judgement

2- Recognition and Recall

4- Application

6- Synthesis

قضایا و دیگر مفاهیم انتزاعی، در موقعیت مناسب، بدون اینکه هیچ راه حلی ارائه شود.
مثال: دانشجو بتواند مشخص کند در چه بیمارانی استفاده از الکتروکاردیوگرام به تشخیص کمک می‌کند.

۴- تجزیه و تحلیل. در این سطح بر تجزیه مطالب به اجزای تشکیل دهنده آن و یافتن روابط بین اجزا و نحوه سازمان یافتن آنها تأکید می‌شود.

مثال: دانشجو بتواند تغییرات پاتولوژیک الکتروکاردیوگرام را با اختلالات بالینی بیمار ارتباط دهد.

۵- ترکیب. در این سطح یادگیرنده قسمت‌هایی از تجارب گذشته را با مطالب جدید در هم آمیخته، آن را به صورت یک کل تازه بازسازی می‌کند.

مثال: دانشجو بتواند با تفسیر الکتروکاردیوگرام و با کمک علائم بالینی بیمار تشخیص افتراقی داده، نوع بیماری فرد را مشخص کند.

۶- ارزشیابی و قضاوت. آخرین و عالیترین مرحله شناخت، ارزشیابی است. در این مرحله فراگیرنده، ناچار به اخذ تصمیم است.

مثال: دانشجو بتواند از بین شیوه‌های مختلف درمانی بهترین را انتخاب کند.

برای تعیین سطوح یادگیری این نکات را رعایت کنید:

الف) سطوح را به صورت رفتارهای قابل مشاهده بیان کنید. (توضیح دهید، بیان کنید، تعریف کنید، به کار برید و غیره) در این صورت می‌توان بین کسانی که می‌دانند و نمی‌دانند فرق گذاشت.

ب) سطوح و اهداف را به صورت کلی بیان کنید. مثلاً: «دانشجو بتواند عبارتهای اپونروز و تاندون را به طور صحیح تعریف کند» غلط است، زیرا کلی نیست و باید به جای کاربرد کلمات «آپونروز» و «تاندون» به صورت تک تک عبارت کلی «اصطلاحات» را به کار برد. یعنی جمله هدف به این صورت در می‌آید: «دانشجو باید بتواند اصطلاحات را به طور صحیح تعریف کند».

ج) سطوح و اهداف تعیین شده را به صورت مبهم بیان نکنید.

به منظور طرح‌نهایی سؤالات به این نکات توجه کنید:

۱- سطوح یادگیری و زمینه‌های محتوا را براساس اهمیت مطلب و مقدار زمانی که صرف آموزش آن شده است، تعیین کنید.

۲- به هر یک از موارد بالا (سطح و محتوا) ارزشی داده، ضریب دهید. این ضرایب در درجه

۳۰ آزمونهای پیشرفت تحصیلی در علوم پزشکی

اول با قضاوت استاد براساس اهمیت موضوع و در درجه دوم با توجه به تعداد صفحات و ساعات صرف شده و تأکید حین تدریس تعیین می‌شود.

۳- با توجه به ضرایب داده شده، تعداد پرسشهای آزمون را بین مباحث و سطوح تقسیم کنید. این جدول در حکم پیش‌آزمون است. پس از تهیه آن سعی کنید سؤالات را دقیقاً براساس محتوا و سطح مشخص شده در جدول تهیه کنید. توفیق بیشتر در این کار، سبب برگزاری آزمون مناسبتر خواهد بود.

۴- نکات دیگری از قبیل طول آزمون، کیفیت سؤالات و دستورالعملها، ترتیب پرسشها، چگونگی تصحیح و نمره‌گذاری، اصلاح نمرات، تفسیر نتایج، اعلام نتایج و تغییر در آزمون را (در صورت نیاز) مدنظر قرار دهید.

مثالی از جدول طرح آزمون

موضوعات مورد امتحان	سطوح دانش		درک		کاربرد		مرحله بالاتر		جمع
	عینی	ذهنی ^۱	عینی	ذهنی	عینی	ذهنی	عینی	ذهنی	
A-۱									X
B-۲									
C-۳									
D-۴									
جمع									Y

X. تعداد کل سؤالات مطرح شده در هر سطح راجع به یک موضوع (در این امتحان خاص راجع به موضوع A, X سؤال طرح شده است). Y. تعداد کل سؤالات موجود در یک سطح یادگیری (در این امتحان Y سؤال در حیطه دانشی به شکل عینی وجود دارد).

مزایای استفاده از جدول طرح آزمون

۱- در این روش آزمون روا بر اساس هدفهای پرورشی و محتوای درسی تهیه می‌شود؛ زیرا در طرح آزمون هیچ نکته مهمی نباید از قلم بیفتد. ضمن اینکه الزاماً نباید کلیه مطالب که در کلاس بحث شده در آزمون گنجانده شود. با استفاده از جدول مشخصات می‌توان به این هدف رسید.

۱- Subjective: در نمره‌گذاری سؤالات ذهنی (برخلاف سؤالات عینی) رأی و نظر تصحیح کننده مؤثر است. پرسشهای تشریحی از نوع ذهنی بوده، سایر انواع آزمون (چندگزینه‌ای، صحیح - غلط و ...) از نوع عینی است.

- ۲- با وجود جدول طرح آزمون اهمیت و ضریب یکسان به قسمت‌های مشابه داده می‌شود.
- ۳- بدون استفاده از جدول، سؤالات فقط از قسمت‌های آسان طرح می‌شود؛ زیرا راحت‌تر می‌توان از این قسمت‌ها سؤال استخراج کرد.
- گاهی برای طرح‌ریزی آزمون به جای جدول مشخصات، از چک‌لیست استفاده می‌شود. در این روش باید در ذهن خود، به تعدادی سؤال در مورد محدوده طراحی شده آزمون، وسعت محتوا، زمان و غیره پاسخ گوئیم. سپس بر اساس پاسخ این سؤالات و مشورت با افراد متخصص آزمون طراحی گردد. مثلاً: تأکید اصلی چیست؟ چند موضوع امتحان می‌شود؟ زمان هر موضوع چقدر است؟ وسعت هر موضوع چقدر است؟ ...

ویرایش آزمون

- پس از تهیه پرسش‌ها بررسی کنید که نکات زیر در آنها رعایت شده است یا خیر:
- ۱- آیا هر پرسش، سطح تعیین شده در جدول مشخصات را اندازه می‌گیرد؟ پاسخ این سؤال به کمک تجربه طراح سؤال بدست می‌آید و نظر شخصی در آن مهم است.
 - ۲- آیا نوع پرسش انتخاب شده برای اندازه‌گیری سطح موردنظر مناسب است؟ در این مورد باید معایب و مزایای هر یک از انواع پرسش‌ها را در نظر بگیرید.
 - ۳- آیا هر پرسش به صورت روشن و با زبان ساده بیان شده است؟ یعنی سؤال باید صریح و با الفاظ و کلمات ساده بوده، جملات آن تا حد امکان کوتاه باشد.
 - ۴- آیا پرسش فاقد قرینه لفظی است؟
- مثال: گره پیشاهنگ در قلب کدام است؟
- الف) دسته دهلیزی بطنی
 - ب) دسته هیس
 - ج) عضله دهلیزی
 - د) گره سینوسی دهلیزی
- وجود کلمه مشترک «گره» در صورت سؤال و گزینه (د) دانشجو را به سمت پاسخ صحیح هدایت می‌کند.
- ۵- آیا میزان دشواری پرسش‌ها برای دانشجو مناسب است؟ دشواری پرسش‌ها بستگی به سطح یادگیری موردنظر دارد. بهتر است میانگین دشواری پرسش‌ها^۱ در حد ۵۰ درصد باشد.

۱- دشواری پرسش: (ضریب دشواری) براساس نسبت دانشجویانی که به سؤال پاسخ صحیح داده‌اند به کل افراد

۶- آیا هر پرسش مستقل است و آیا به طور کلی در پرسشها مطالب مشترک وجود ندارد؟

۷- آیا پرسشهای آزمون همه بخشهای جدول مشخصات را در بر می گیرد؟

۸- آیا سؤال فقط دارای یک پاسخ درست و مشخص است؟

به منظور ویرایش سؤالات این اقدامات را انجام دهید:

۱- سؤالاتی را که قدرت تمیز^۱ ضعیف دارند حذف کنید. چون معمولاً سؤالات باید از لحاظ

آماری بتواند تفاوت موجود را میان گروههای قوی و ضعیف دانشجویان نشان دهد.

۲- گزینه‌های نادرست هر سؤال را - که با توجه به درجه دشواری سؤال دارای جذابیت

کمتر یا بیشتر از حد مطلوب است - جایگزین و اصلاح کنید.

۳- می‌توانید سؤالاتی را که بیش از حد آسان یا دشوار است یا با بقیه سؤالات هماهنگ

نیست، حذف کنید.

۴- در صورت امکان تمام پرسشهای متشابه را کنار هم بیاورید. این عمل چندین فایده دارد:

الف) برای هر گروه سؤالات یک دستورالعمل واحد بیان می‌گردد؛

ب) کارآیی دانشجو برای پاسخ‌گویی بیشتر می‌شود؛

ج) تصحیح و تجزیه و تحلیل نتایج آزمون آسانتر می‌شود.

۵- بهتر است سؤالات مربوط به یک سطح یادگیری (هدف آموزشی) را در کنار یکدیگر

بیاورید. این موضوع باعث می‌شود بتوانید سطحی را از یادگیری مشخص نمایید که برای

دانشجو مشکل ایجاد کرده، آموزشهای فردی یا گروهی را براساس آن طرح‌ریزی کنید.

در صورتی که در یک امتحان، ترتیب سؤالات بر اساس سطح یادگیری و ترتیب

براساس نوع سؤالات منطبق نباشد، استفاده از طبقه‌بندی براساس سطح یادگیری

مناسبت‌تر است؛ زیرا فواید آموزشی آن بیشتر می‌شود. ضمن اینکه در اغلب امتحانات به

علت محدود بودن نوع سؤالات این مشکل پیش نمی‌آید.

۶- سؤالات را از نظر سطح دشواری، از آسان به مشکل مرتب کنید. این روش اثر تسکینی یا

آرامش‌بخش مطلوبی بر دانشجو داشته، باعث می‌شود دانشجوی ضعیف در همان ابتدا

با سؤالات مشکل روبرو نشود. البته عده‌ای نیز معتقدند که امتحان باید به گونه‌ای

پراکنده و ترکیبی از سؤالات آسان و مشکل باشد. ولی به دلیل فوق ترتیب از آسان به

حساب می‌شود و این نسبت هر چه بالاتر باشد نشانه آسانتر بودن سؤال است که در فصل تجزیه و تحلیل توضیح داده خواهد شد.

۱- قدرت تمیز (ضریب تمیز) شاخصی است که تفاوت بین دانشجویان قوی و ضعیف را در پاسخ‌گویی به سؤالات آزمون نشان می‌دهد، که در فصل تجزیه و تحلیل به صورت مشروح توضیح داده خواهد شد.

مشکل بهتر است.

۷- تا حد امکان سؤالات تشریحی را همراه با سؤالات عینی با هم در یک جلسه آزمون مورد استفاده قرار ندهید.

۸- محل قرار گرفتن گزینه‌های صحیح را در سؤالات چند گزینه‌ای و صحیح - غلط به طور تصادفی انتخاب کنید.

۹- بهتر است برای آزمون‌ها پاسخنامه جداگانه‌ای تهیه نمایید.

۱۰- برای آزمون دستورالعمل خاصی را تعیین کنید، تا دانشجو با توجه به این دستورالعمل نسبت به پاسخگویی به سؤالات اقدام کند.

دستورالعملها باید ساده، مختصر و حاوی اطلاعات لازم درباره مطالب زیر باشد:

(الف) هدف آزمون؛

(ب) مدت آزمون؛

(ج) چگونگی ثبت پاسخها؛

(د) نحوه عمل در مورد حدس پاسخ و داشتن یا نداشتن نمره منفی.

وقتی در یک آزمون چندین نوع پرسش به کار می‌برید بهتر است یک دستورالعمل کلی در زمینه هدف آزمون، مدت آن و راهنمایی مربوط به وجود و یا فقدان نمره منفی بیان کرده، برای هر سنجش نیز یک دستورالعمل اختصاصی در زمینه چگونگی ثبت پاسخها بنویسید. در بعضی انواع پرسشها باید برای هر پرسش، دستورالعمل جداگانه‌ای به کار برید. همچنین در صورت آشنا نبودن قبلی دانشجویان با پاسخنامه باید نحوه استفاده آن را دقیقاً برای آنها مشخص کنید.

۱۱- اگر سؤالات آزمون را به چند دانشجو بدانید تا با صدای بلند پاسخ گویند، در این صورت مشخص می‌شود که از لحاظ لغات، اصطلاحات و نمودارها در کجا مشکل وجود داشته، درک منظور کدام سؤال مشکل‌تر است.

۱۲- سؤالات آزمون را به افراد متخصص در آن رشته نشان داده، نظر آنها را در مورد محدوده و سطح پرسشها اخذ کنید.

۱۳- آزمون را برای گروه نمونه اجرا کرده، میانه، میانگین، انحراف استاندارد، درجه دشواری و قدرت تمیز و ضریب پایایی آن را تعیین کنید.

بعد از طی این مراحل، می‌توانید آزمون را با اطمینان برای فراگیرندگان اجرا کرده، از نتایج آن برای مقاصد مختلف بهره بگیرید.

منابع:

- ۱- آذر، رسول. *روشهای ارزیابی آموزشی*. تهران: انتشارات نیما، ۱۳۶۹.
- ۲- بلوم؛ و همکاران. *راهنمای ارزشیابی تکوینی و مجموعی*. جلد اول. ترجمه ابراهیم کظیمی. تهران: انتشارات دانشگاه تربیت معلم، ۱۳۵۵.
- ۳- پاشا شریفی، حسن. *اصول روان‌سنجی و روان‌آزمایی*. تهران: انتشارات رشد، ۱۳۷۳.
- ۴- سیف، علی‌اکبر. *روانشناسی یادگیری و تدریس*. تهران: انتشارات دانشگاه سپاهیان انقلاب ایران، ۱۳۵۵.
- ۵- شهاب، گیتا. *معرفی ارزشیابی تشخیصی*. تهران: انتشارات دانشگاه سابق آزاد ایران، ۱۳۳۸.
- ۶- گرون لوند، نورمن. *آزمونهای پیشرفت تحصیلی*. تهران: انتشارات سازمان سنجش، ۱۳۵۶.
- ۷- گنجی، حمزه. *اندازه‌گیری و ارزشیابی در آموزش و پرورش*. تهران: انتشارات بعثت، ۱۳۷۳.
- ۸- لطف آبادی، حسین. *سنجش و اندازه‌گیری در علوم تربیتی و روانشناسی*. تهران: انتشارات فردوسی، ۱۳۷۳.
- ۹- مایزرما، ویلیام؛ و همکاران. *اندازه‌گیری و آزمون در تعلیم و تربیت*. ترجمه غلامرضا خوئی‌نژاد. مشهد: انتشارات مؤسسه چاپ و انتشارات آستان قدس رضوی، ۱۳۷۲.
- ۱۰- ولف، ریچارد. *ارزشیابی آموزشی (مبانی سنجش توانایی و بررسی برنامه)*. ترجمه علیرضا کیامنش. تهران: انتشارات دانشگاه تهران، ۱۳۷۱.
- ۱۱- هومن، حیدرعلی. *اندازه‌گیریهای روانی و تربیتی و فن تهیه تست*. تهران: انتشارات فردوسی، ۱۳۷۳.
- ۱۲- حوزه معاونت آموزشی مرکز توسعه آموزش پزشکی اصفهان. *گامهای مختلف ارزیابی اصفهان*: ۱۳۶۹.
- ۱۳- حوزه معاونت آموزشی مرکز توسعه آموزش پزشکی اصفهان. *رابطه ارزیابی با اهداف آموزشی اصفهان*: ۱۳۶۹.
- 14- Abbatt FR. *Teaching For better learning*. 2nd ver.ed. Geneva: World Health Organization, 1992.
- 15- Athurber W, Callette A. *Teaching Science*. 1959.
- 16- Beck DE. *Evaluating student performance in the experiential setting with confidence*. Am J Pharm Educ, 1995 fall; 59(2): 236-47.
- 17- Cox K. *The medical teacher, writing multiple choice question*. London: Churchill-Livingstone, 1982.
- 18- Ehadra D. *Weightage of formative examinations in certifying examination of pharmacology: An opininon poll and relative performance of learners*. Med Teach 1992; 14(2/3): 197-200.
- 19- Hill DA. *Formative assessment: A student spective*. Med Educ 1994; 28: 394-399.
- 20- James SC. *Designing tests for evaluating student achivement*. London: Longman, 1990.
- 21- Jenkins DM. *Dose incourse assessment and counselling improve exit performance?* Med Teach 1992; 14(2/3): 241-2.
- 22- Paol VK. *Assessment of clinical competence of undergraduate medical students*. Indian J pediatr 1994; Mar-Apr; 61(2): 145-51.
- 23- Vleuten V. *A written test as an alternative to performance testing*. Med Educ 1989; 23 (1): 97-107.

فصل دوم

طبقه‌بندی آزمونها

- طبقه‌بندی براساس نحوه ساخت آزمون
- طبقه‌بندی براساس نحوه تصحیح آزمون
- طبقه‌بندی براساس نحوه پاسخگویی
- طبقه‌بندی براساس نوع پرسش
- طبقه‌بندی براساس تفسیر نتایج
- طبقه‌بندی براساس موضوع
- طبقه‌بندی براساس هدف

فکر می‌کنید چند نوع آزمون وجود دارد؟ چند نوع آن را تجربه کرده‌اید؟ چه تفاوتها و شباهت‌هایی بین آنها می‌بینید؟ تا چه اندازه به خصوصیات آزمون‌ها آشنایی دارید؟

آزمون‌ها گوناگون و کاربرد آنها نیز متفاوت است به طوری که هر کدام در موقعیت خاصی بهترین نتایج را دارند؛ بنابراین، آزمون‌گیرنده باید بتواند با توجه به وضعیت، بهترین را انتخاب کند. این هدف وقتی میسر است که او با انواع آزمون‌ها، مشخصات هر یک، موارد استفاده آنها، مزایا، معایب و سایر ویژگیهای آنها آشنایی کامل داشته باشد؛ در غیر این صورت آزمون‌گیرنده و آزمون‌شونده، هیچکدام نتیجه مطلوبی بدست نمی‌آورند و نتایج این امتحان نیز مستند و دقیق نخواهد بود. بنابراین، برای اجرای یک آزمون دقیق و کارآمد، آگاهی از انواع آزمون‌ها هم برای آزمون‌شونده و هم برای آزمون‌گیرنده ضروری می‌باشد. در این فصل سعی خواهد شد انواع آزمون‌ها و مشخصات هر یک تا حد لزوم بیان گردد. امتحان جنبه‌های مختلفی دارد از جمله: طراحی، اجرا، انتخاب نوع پرسش، چگونگی پاسخ‌دهی، نحوه تفسیر نتایج و کاربرد که هر کدام از این ابعاد، اساس طبقه‌بندی آزمون قرار می‌گیرد. در ادامه مبحث طبقه‌بندی آزمون‌ها از جنبه‌های مختلف، بررسی می‌شود.

الف) طبقه‌بندی براساس نحوه ساخت آزمون

بر این اساس دو نوع امتحان شناخته شده است:

- ۱- آزمون معلم ساخته^۱؛
- ۲- آزمون استاندارد شده^۲.

۱- آزمون معلم ساخته

در این نوع آزمون، استاد با توجه به مطالب تدریس شده و اطلاعات لازم برای دانشجویان، آزمون را تهیه و اجرا کرده، میزان موفقیت خود و دانشجویان را می‌سنجد.

این امتحان از نظر سطح یادگیری موردنظر، محتوای درسی و چگونگی اجرا، بستگی کامل به دیدگاه، استاد دارد. بدین ترتیب، اگر موضوع درسی و هدف امتحان نیز یکسان باشد باز هم آزمونهای طرح شده به وسیله استادان مختلف، به خاطر دیدگاههای گوناگون، متفاوت خواهد بود.

آزمون معلم ساخته از نظر محتوا محدود بوده، صرفاً یک یا چند کلاس را در بر می‌گیرد؛ بنابراین نتایج آن هم در همان حد اعتبار و ارزش دارد و نمی‌توان برای ارزیابی وسیع دانشجویان از آن استفاده کرد. آزمون معلم ساخته را در موارد زیر می‌توانید بکار ببرید:

۱- برای اینکه اطمینان پیدا کنید به هدف مورد نظر خویش رسیده‌اید. (البته هدفهای محدود و در سطح یک یا چند کلاس).

۲- برای اینکه به دانشجو نمره بدهید. بدیهی است که در این آزمونها نمره خوب یا بد مفهوم ندارد، بلکه با توجه به ملاکهای موردنظر خود، برای تفسیر نتایج باید از عباراتی مانند «کاملاً یاد گرفته است»، «هنوز باید بیاموزد» و ... استفاده کنید.

۲- آزمون استاندارد شده

این آزمون توسط تعدادی از متخصصان آن موضوع و برای گروه مخصوصی طرح‌ریزی شده، هدف اصلی اجرای آن مقایسه افراد می‌باشد. هر آزمون استاندارد شده برای رسیدن به مرحله نهایی اجرایی و همچنین، قابل استناد شدن باید مراحل زیر را طی کند:

۱- ابتدا متخصصان فن، سؤالات موردنظر را با توجه به هدف آزمون و محتوای درسی انتخاب کنند.

۲- آزمون برای تعدادی از افراد اجرا شود. این افراد نمونه جامع از گروهی هستند که آزمون برای آنها طراحی شده است.

۳- آزمون از نظر سطح دشواری، قدرت تمیز و رتبه‌بندی دانشجویان بررسی شده، مورد تجزیه و تحلیل قرار گیرد و سؤالاتی که از نظر سختی و آسانی و قدرت تمیز، شرایط مطلوب را نداشته باشد، حذف یا اصلاح شود.

۴- افرادی که در آزمون شرکت کرده‌اند براساس معیارهای مختلف مثل سن، جنس و ... گروه‌بندی شده، با توجه به نتایج بدست آمده برای هر گروه امتیازهایی در نظر بگیرند که نشان‌دهنده میزان مهارت افراد در آن گروه باشد. سپس آن را به کل گروه موردنظر

تعمیم دهند. آنچه به این ترتیب بدست می‌آید آزمون معیار^۱ نامیده می‌شود.

۵- پس از تهیه آزمون معیار، نمره هر فرد پس از شرکت در امتحان با توجه به گروهی که در آن قرار می‌گیرد، با نمرات مربوط مقایسه شده، وضعیت فرد در بین افراد همگروه خود مشخص می‌شود.

مقایسه دانشجویان در مناطق مختلف به نحو صحیح، مهم‌ترین کاربرد آزمون استاندارد شده است. همچنین این آزمونها در بسیاری از زمینه‌های علمی کاربردهایی دارد که از جمله می‌توان به موارد زیر اشاره نمود:

۱- تعیین سطح یادگیری دانشجو: زمانی که می‌خواهید دانشجویانتان را رتبه‌بندی کنید و یا یقین حاصل کنید که آنها مقطع تحصیلی فعلی را به طور کامل و مطلوب به پایان رسانیده، شایستگی ورود به مقطع بالاتر را دارند، می‌توانید از این آزمونها استفاده کنید.

۲- تعیین نقاط ضعف و قوت دانشجو: با کمک این آزمون می‌توانید نقاط ضعف و قوت هریک از دانشجویان را در محدوده سن و مقطع تحصیلی وی مشخص کرده، با توجه به استعدادهايش او را به رشته تحصیلی خاصی راهنمایی کنید.

۳- تعیین میزان پیشرفت دانشجو: اگر دو امتحان یکسان با فاصله زمانی مشخص (حدود چند ماه) از دانشجو بگیرید با مقایسه نتایج آن می‌توانید به میزان پیشرفت وی در این مدت پی ببرید. در صورتی که پیشرفت، بسیار کند یا بسیار سریع باشد، توجه ویژه‌ای لازم است.

۴- تعیین موقعیت نسبی یک کلاس: با مقایسه نتایج هر یک از کلاسها می‌توانید آنها را رتبه‌بندی کرده، برنامه‌های آموزشی آینده را براساس این اطلاعات طراحی کنید. همچنین می‌توانید نقاط ضعف و قوت یک کلاس و میزان پیشرفت آن را به این ترتیب تعیین کرده، برای طراحی برنامه‌ها و تعیین میزان موفقیت تدریس در کلاس از آنها استفاده کنید.

البته آزمونهای استاندارد شده بی‌اشکال نیست و معایبی نیز دارد، که هنگام استفاده از آنها باید مدنظر قرار گیرد از جمله:

۱- نتایج این آزمونها به تنهایی برای رتبه‌بندی دانشجویان تعیین کننده نیست؛ زیرا ممکن است دانشجویی که شایستگی ندارد به مقاطع بالاتر برود و فقط به صورت اتفاقی و به علت آگاهی از روشهای آزمون بتواند نمره خوبی بدست آورد، یا برعکس.

۲- این امتحانات معمولاً نمی‌تواند همه سطوح یادگیری را بسنجد.

۳- از آنجایی که قبولی در این امتحانات اغلب کلید ورود به مقاطع تحصیلی بالاتر است، این

سوء تعبیر برای استاد و دانشجو پیش می‌آید که هر آنچه دانشجو برای ورود به مقطع تحصیلی بالاتر نیاز دارد در این امتحانات گنجانده شده است. بنابراین تمام تلاش استاد و دانشجو به موفقیت هر چه بیشتر در این امتحان معطوف می‌شود نه کسب مهارت‌های لازم برای فعالیت شغلی آینده؛ چون فقط محتوای موردنظر در این امتحانات برای آنها حائز اهمیت است. به همین دلیل، استادان برای بهبود تدریس یا تغییر در برنامه آموزشی خود تلاش نمی‌کنند، چون احتمال می‌دهند در نتیجه تغییر برنامه، دانشجویان در این امتحانات نمرات پایین کسب کنند. در ضمن دانشجویان نیز از تغییر برنامه استقبال نمی‌کنند و این مسائل باعث رخوت و رکود در آموزش می‌شود.

۴- آزمونهای استاندارد شده، برای مقاصد کلی طرح‌ریزی شده، جوابگوی طرحهای محلی نیست. این آزمونها بیشتر براساس وضعیت حاکم بر کشور یا منطقه سازنده آزمون، طرح‌ریزی شده، برای کشورهای دیگر چندان قابل استفاده نیست. حتی آزمونهایی هم، که به اصطلاح "بری از فرهنگ" خاصی می‌باشد در این زمینه بدون اشکال نیستند.

به خاطر تفاوت‌های ذکر شده، گاهی نتایج حاصل از دو امتحان، متناقض می‌شود به طوری که براساس آزمون استاندارد، فرد یا کلاسی برتر شناخته می‌شود، در حالی که ممکن است در آزمون معلم ساخته فرد یا کلاس دیگری برتر شود. در نتیجه نتایج این امتحانات باید بر اساس محتوای امتحان، هدف و ... تفسیر شود.

ذکر این نکته لازم است که آزمونهای معلم ساخته و استاندارد با هم تضاد نداشته، مکمل یکدیگر است، فقط تفاوت آنها این است که آزمون استاندارد بر اساس محتوای کلی و هدفهای یکسان بین نظامهای تحصیلی مختلف و با قصد مقایسه افراد و تبیین وضعیت یک فرد در میان همگروه‌های خود ساخته می‌شود، در حالی که آزمون معلم ساخته از نظر هدف و محتوا محدود به یک کلاس است.

ب) طبقه‌بندی براساس نحوه تصحیح آزمون

براین اساس، دو نوع آزمون قابل تشخیص است:

۱- آزمون عینی^۱؛

۲- آزمون ذهنی^۲؛

۱- آزمون عینی

آزمون عینی، آزمونی است که نتایج آن همیشه ثابت باشد و اگر افراد مختلف یا یک فرد در زمانهای مختلف آن را تصحیح کنند، در همه موارد نمره یکسان باشد؛ این آزمونها براساس کلید از پیش تعیین شده برای سوالات نمره‌دهی شده، قضاوت ذهنی یا وضعیت تصحیح‌کننده تأثیری بر آن ندارد. البته هر چه این شرایط بیشتر مدنظر قرار گیرد آزمون از عینیت بیشتری برخوردار خواهد بود.

آزمونهای عینی، انواع مختلف دارد از جمله: **آزمون با سوالات چند گزینه‌ای، صحیح - غلط، جور کردنی و کوتاه پاسخ^۱** که هر یک در فصول آینده به طور کامل بررسی می‌شود.

از مزایای مهم این امتحانات می‌توان موارد زیر را نام برد:

- ۱- می‌توانید در امتحان از تعداد زیادی پرسش استفاده کنید. بنابراین آزمون محدوده وسیعی را از مفاهیم درسی در برگرفته، روایی محتوا^۲ افزایش می‌یابد.
- ۲- نمره دادن به این سوالات، آسان، دقیق و عینی بوده، می‌توانید از ماشین برای تصحیح اوراق استفاده نمایید.

هنگام استفاده از انواع پرسشهای عینی باید به نکات زیر توجه داشته باشید:

- ۱- این نوع سوالات، معمولاً برای سنجش سطوح دانستن، درک، کاربرد، تجزیه و تحلیل به کار رفته، برای سطوح ترکیب و ارزشیابی کارآمد نیست، بخصوص اگر در طراحی سوالات دقت نشود آزمون تنها به ارزیابی اطلاعات سطحی ولی در حد وسیع پرداخته، اطلاعات عمقی و اساسی نادیده گرفته می‌شود.
- ۲- تهیه پرسش عینی خوب مشکل بوده، احتیاج به تخصص، دقت و مهارت کافی دارد. بنابراین باید مهارت کافی در این زمینه کسب کرده، زمان لازم را به آن اختصاص دهید.
- ۳- دقت کنید که نتایج این آزمونها صد درصد مبنا قرار نگیرد؛ زیرا در امتحانات عینی، حدسی جواب دادن دانشجو، سرعت عمل و مهارت وی در خواندن، همگی می‌تواند در نمره وی تأثیر داشته باشد. گاهی برای جلوگیری از حدسی جواب دادن و از بین بردن شانس دانشجو در این امتحانات گاهی از نمره منفی استفاده می‌شود؛ یعنی اگر دانشجو به سؤالی پاسخ نادرست بدهد، نمره مخصوص آن سؤال را از دست می‌دهد درصدی از نمره پاسخهای صحیح نیز از نمره وی کسر می‌شود.

1- Short Answer

۲- Content Validity. روایی محتوا، همبستگی بین توزیع سؤالا نسبت به قسمت‌های مختلف درس، مدت زمان تدریس و اهداف آموزشی می‌باشد.

۲- آزمون ذهنی

آزمون ذهنی، آزمونی است که در آن نمره‌دهی به صورت ذهنی است؛ یعنی اگر افراد مختلف در یک زمان و یا یک نفر در زمانهای مختلف یک برگه آزمون را تصحیح کرده، نمره دهند نمرات متفاوتی بدست آمده، نتایج آن ثابت نیست؛ زیرا قضاوت ذهنی و وضعیت روحی مصحح بر آن تأثیر می‌گذارد.

از انواع پرسشهای ذهنی می‌توان به سؤالات انشایی، تشریحی^۱ و شفاهی اشاره کرد.

این آزمونها محدودیتهای زیادی دارند که مهمترین آنها عبارتند از:

۱- این سؤالات از نظر نتایج آزمون ثبات لازم را ندارد؛

۲- این آزمون از نظر محتوا محدود بوده، به علت تعداد کم سؤالات، تنها بخش کوچکی را از محتوای تدریس را در بر می‌گیرد؛

۳- چون در این آزمون، خود دانشجو جواب را تهیه می‌کند توانایی وی در گفتار و نوشتار نیز ارزیابی می‌شود. پس اگر هدف اولیه آزمون، سنجش این مهارتها باشد، اطلاعات علمی فرد نباید مورد توجه قرار گیرد ولی اگر هدف، سنجش اطلاعات علمی فرد است این نحوه آزمون معمولاً بهترین نیست.

با همه محدودیتهای این نوع آزمونها به نحو گسترده‌ای در نظام آموزش عالی کاربرد دارد؛ بخصوص در مواردی که ارزیابی سطوح یادگیری پیچیده‌تر، مدنظر باشد.

ج) طبقه‌بندی براساس نحوه پاسخ‌گویی

براین اساس، سؤالات را می‌توان به دو دسته اصلی تقسیم کرد:

۱- گزیده پاسخ یا سؤالات انتخابی یا تستی^۲؛

۲- بازپاسخ یا سؤالات تأمینی یا کلی^۳.

۱- آزمون گزیده پاسخ

در این نوع سؤالات، دانشجو پاسخ صحیح را از بین تعدادی جواب پیشنهادی انتخاب می‌کند. این سؤالات بسیار شبیه سؤالات عینی است. در حقیقت آزمونهای عینی بیشتر گزیده پاسخ هستند.

1- Essay Question
3- Supply-type Item

2- Choice-type Item

مزایای سؤالات گزیده پاسخ به قرار زیر است:

۱- چون نمره‌گذاری آنها سریع، عینی و پایا می‌باشد روایی آنها نیز در بیشتر موارد بالا است.
۲- می‌توان تعداد زیادی سؤال را در یک امتحان قرار داد؛ بنابراین محدوده وسیعی از اطلاعات سنجیده می‌شود و روایی محتوا افزایش می‌یابد.

۳- در این سؤالات، زیبایی خط یا قدرت بیان دانشجو در نمره او تأثیر نمی‌گذارد.

۴- در پاسخ دادن، دانشجو نمی‌تواند حاشیه رفته، در نمره‌دهی مشکل ایجاد کند.

اغلب در این سؤالات، سطوح پایین یادگیری اندازه‌گیری می‌شود که این عیب بیشتر متوجه شیوه طرح سؤال است نه نوع سؤال؛ به طوری که در انواع خوب طراحی شده، بخصوص در مورد سؤالات چند گزینه‌ای می‌توان، سطوح یادگیری عالی را نیز اندازه گرفت.

سؤالات انتخابی چند نوع است:

۱- سؤالات چند گزینه‌ای؛

۲- سؤالات صحیح - غلط؛

۳- سؤالات جور کردنی؛

۴- سؤالات استدلالی؛

۵- سؤالات ترکیبی.

۲- آزمون باز پاسخ

در این نوع آزمون، دانشجو خود باید جواب صحیح را تهیه و بیان نماید. سؤالات باز پاسخ نیز شبیه سؤالات ذهنی است.

از ویژگیهای بارز این سؤالات می‌توان به موارد زیر اشاره کرد:

۱- این پرسشها اندازه‌گیری برخی سطوح یادگیری را به طور مستقیم امکان‌پذیر می‌کند،
هتل: سطوح ترکیب و ارزشیابی.

۲- چون جواب را دانشجو، خود تهیه می‌کند احتمال حدسی پاسخ دادن وی بسیار کاهش می‌یابد.

۳- طراحی این سؤالات تقریباً آسان است.

۴- از آنجاکه نمره‌گذاری این سؤالات به صورت ذهنی است، پایایی آن به میزان زیادی کمتر از نوع گزیده پاسخ بوده، اختلاف در مفاهیم کلمات و تعبیر و تفسیر آنها و حتی غلطهای املائی می‌تواند در نمره‌دهی تأثیر داشته باشد؛ بنابراین روایی نیز کاهش می‌یابد.

این سؤالات شامل انواع زیر می‌باشد:

- ۱- سؤالات کامل کردنی^۱؛
- ۲- سؤالات کوتاه پاسخ؛
- ۳- سؤالات تشریحی، شامل انواع محدود پاسخ و گسترده پاسخ.

د) طبقه‌بندی براساس نوع پرسش

با وجود تنوع سؤالات به کار رفته در آزمونهای مختلف می‌توان آنها را به سه دسته کلی تقسیم کرد:

- ۱- سؤالات گسترده یا بی ساختار؛
- ۲- سؤالات سازمان یافته؛
- ۳- سؤالات بسیار سازمان یافته.

۱- سؤالات گسترده یا بی ساختار

در این سؤالات دانشجو کمترین محدودیت را برای پاسخ گویی داشته، می‌تواند با حداکثر آزادی، افکار و احساسات خود را بیان کند. بنابراین پاسخها بسیار متنوع خواهد بود. این نوع پرسشها بیشتر برای کشف عقاید یا دیدگاههای خاص دانشجویان در زمینه‌های خاص کاربرد دارد. مثلاً: اگر بخواهیم احساس دانشجو را در زمینه یک قطعه شعر یا موسیقی یا نظری وی را در مورد چگونگی اداره بیمار^۲ یا رویکرد به بیماری جویا شویم، این سؤالات بهترین نوع به حساب می‌آید. البته به دلیل گسترده بودن پاسخ، پایایی لازم را نداشته، کاربرد آنها محدود است.

۲- سؤالات سازمان یافته

در این سؤالات یک بخش اصلی وجود دارد که هدف سؤال را بیان می‌کند و به دنبال آن سؤالاتی مطرح می‌شود که جوابهای کوتاه و مشخص دارد؛ بنابراین نتایج عینی‌تر است. این نوع سؤالات درک دانشجویان را از مفاهیم و اصول درسی می‌سنجد و نسبت به سؤالات گسترده پاسخ کاربرد وسیع‌تری دارد.

۳- سوالات بسیار سازمان یافته

در این پرسشها، دامنه پاسخهای صحیح یک سؤال، کاملاً محدود است و پاسخها از قبل مشخص و کدگذاری شده است و دانشجو باید جواب صحیح را انتخاب کند. به همین جهت این سوالات از پایایی بالایی برخوردار بوده، امروزه کاربرد وسیعی دارد.

۵) انواع آزمون از نظر تفسیر نتایج

نمرات آزمونها به دو صورت تفسیر می‌شود:

۱- آزمون ملاک نسبی یا هنجار مرجع^۱؛

۲- آزمون ملاک مطلق یا ملاک مرجع^۲.

۱- آزمون ملاک نسبی

آزمون ملاک نسبی یا هنجار مرجع برای تعیین تفاوت بین دانشجویان و رتبه‌بندی آنها به کار می‌رود. در این آزمون، عملکرد متوسط آزمون‌شوندگان ملاک است و به کافی یا ناکافی بودن مهارتها توجهی نمی‌شود. این آزمون نشان دهنده آن چیزی است که وجود دارد نه آنچه باید وجود داشته باشد. در آزمون ملاک نسبی، نمره هر فرد تنها در مقایسه با نمرات سایر افراد معنی دار بوده، به تنهایی نشان دهنده میزان مهارت فرد نیست. قبول و مردود شدن فرد در این امتحان بستگی به مهارت نسبی وی بین همکلاسه‌هایش دارد، مثلاً: ممکن است در یک امتحان نمره $\frac{25}{100}$ جزو نمرات عالی کلاس باشد در حالی که در امتحان دیگری نمره $\frac{60}{100}$ ضعیف باشد. این تفاوت بستگی به دشواری سوالات و سطح اطلاعات افراد دارد. در این آزمون باید پراکندگی وسیعی از نمرات وجود داشته باشد تا تمایز بین افراد امکان‌پذیر باشد. برای این منظور دشواری سوالات باید در حد متوسط پیش‌بینی شود. آزمونهای معلم ساخته، اغلب به شکل ملاک نسبی تفسیر می‌شود و هدف استاد رتبه‌بندی دانشجویان و تعیین میزان اطلاعات نسبی آنها می‌باشد.

۲- آزمون ملاک مطلق

آزمون ملاک مطلق مشخص می‌کند دانشجو در کدام زمینه یادگیری دارای تسلط شده است. از آنجا که آزمون بر سنجش یک مهارت خاص تمرکز دارد تمام سوالات آن باید

دقیقاً در همان زمینه و سطح یادگیری موردنظر، مرتبط باشد. در این آزمون سختی و آسانی سؤالات اهمیت ندارد چرا که تمام حیطه‌های یادگیری از مراحل ابتدایی ساده تا مراحل انتهایی دشوار باید مورد سؤال قرار گیرد و حذف سؤالات آسان موجب می‌شود اندازه‌گیری تسلط دانشجو به صورت ناقص انجام گیرد.

در این آزمون‌ها نمره هر فرد به طور مطلق و بدون توجه به نمرات دیگران تفسیر می‌گردد و قبولی و مردودی وی با توجه به نمره او در مقایسه با ملاکهای استاندارد مشخص می‌شود.

همانطور که ذکر شد هدف از اجرای آزمون ملاک مطلق مشخص کردن میزان موفقیت دانشجو در رسیدن به اهداف موردنظر است. پس باید قبل از هر کار باید اهداف مورد آزمون به صورت رفتارهای قابل مشاهده بیان شده، سؤالاتی که بر این رفتارها دلالت دارد در مرحله بعد در آزمون مطرح گردد در این صورت نتایج آزمون قابل قبول خواهد بود. در این آزمون نمره افراد با یک ملاک مطلق سنجیده می‌شود، بنابراین ممکن است همه افراد در آزمون قبول یا مردود شوند. از سوی دیگر عوامل مختلفی از جمله سختی و آسانی سؤالات و نحوه تصحیح آنها - در صورتی که سؤالات عینی نباشد - نیز می‌تواند نتایج آزمون را تحت تأثیر قرار دهد.

اما استانداردهای تسلط و ملاک مطلق که نتایج آزمون با آن مقایسه می‌شود چیست و چگونه تعیین می‌شود؟

حد تسلط بر حسب نوع امتحان می‌تواند سرعت یا دقت در کار یا درصد پاسخهای صحیح و یا مجموعه‌ای از آنها باشد. برای تعیین این استانداردها اگر خودتان در مشخص کردن آنها سعی و تلاش به خرج دهید، کاری مشکل و بی نتیجه را پیش گرفته‌اید؛ زیرا باید معیارهای مختلفی را مدنظر قرار دهید، در حالی که راهنمای روشن و مشخصی هم برای این کار در دست ندارید. البته در طرحهای کوچک و محلی، موضوع تفاوت می‌کند. مثلاً: در سطح یک کلاس می‌توان در مورد مباحث و سطوح یادگیری مختلف، درصد جوابهای گروهی دانشجویان را اندازه گرفته، در جدولی تنظیم کرد و با بررسی آنها تخمین زد که در این حالت، استاندارد باید در چه حدی قرار بگیرد تا حد تسلط دانشجو قابل قبول باشد. ولی به طور کلی بهتر است استانداردهای پذیرفته شده را به کار برد. برای تعیین استانداردها روشهای مختلفی وجود دارد که یکی از آنها روش ندلسکی^۱ است که مراحل آن در زیر ذکر

می‌گردد:

- ۱- گروهی از کارشناسان و صاحب‌نظران دارای صلاحیت در مبحث موردنظر، در آن همکاری می‌کنند.
 - ۲- هر یک از کارشناسان به طور جداگانه سؤالات آزمون را بررسی کرده، پاسخهایی را که احتمال دهند دانشجو می‌تواند با حداقل شایستگی آنها را به عنوان پاسخ نادرست حذف کند مشخص کرده، کنار می‌گذارند.
 - ۳- احتمال حدس زدن هر سؤال با توجه به گزینه‌های باقیمانده مشخص می‌شود؛ که این نسبت برابر است با $\frac{1}{\text{تعداد گزینه‌های باقیمانده}}$.
 - ۴- مجموع احتمالات بدست آمده در مورد تمام سؤالات آزمون، نمره‌ای است که از دانشجوی با حداقل شایستگی انتظار می‌رود.
 - ۵- در نهایت معدل نمرات موردانتظار کارشناسان را محاسبه کرده، عدد بدست آمده حداقل استاندارد قابل قبول می‌باشد.
- استانداردهایی که به این روش محاسبه می‌شود به وضعیت خاص خود مربوط است و هنگام کاربرد باید با توجه به موقعیت و تجارب مختلف، آنها را تغییر داد در این راستا توجه به نکات زیر راهگشا است:
- ۱- معمولاً بهتر است سؤالات چندگزینه‌ای به نحوی طراحی شوند که حد تسلط در آنها کسب حداقل ۸۵ درصد نمره کل باشد.
 - ۲- در موارد زیر بهتر است سطح را بالا ببرید:
 - الف- اگر این اهداف برای مراحل بعدی آموزشی، پایه و اساس قرار می‌گیرد؛
 - ب- اگر حیطه موردنظر مرتبط با سلامتی افراد است؛
 - ج- اگر تعداد سؤالات امتحان کم است.
 - ۳- در موارد زیر بهتر است سطح را پایین بیاورید:
 - الف- اگر اهداف آموزشی در دوره بعدی آموزش، دوباره تکرار می‌شود؛
 - ب- اگر اهداف آموزشی بسیار پیچیده و انجام آنها مشکل است.
- در ضمن، استاندارد تعیین شده باید با نوع سؤال انطباق داده شود. مثلاً: اگر در سؤالات کوتاه پاسخ، استاندارد حداقل ۸۰ درصد پاسخ صحیح باشد؛ احتمالاً در امتحان صحیح - غلط بهتر است آن را حدود ۹۵ درصد پاسخ صحیح قرار دهید. این تفاوت به دلیل اختلاف در میزان حدس زدن دانشجو است.

جدول ۱-۲. مقایسه میان خصوصیات آزمون ملاک نسبی و ملاک مطلق

آزمون ملاک مطلق	آزمون ملاک نسبی	
آزمون برای سنجش تسلط دانشجویان به کار می‌رود.	آزمون برای رتبه‌بندی دانشجویان به کار می‌رود.	۱- استفاده اصلی
مشخص کردن مهارتهایی که دانشجو می‌تواند انجام دهد.	سنجش تفاوت‌های فردی در پیشرفت تحصیلی.	۲- تأکید اصلی
عملکرد دانشجو با استانداردهای مشخص مقایسه می‌شود.	عملکرد دانشجو با دیگران مقایسه می‌شود.	۳- تفسیر نتایج
معمولاً بر وظایف و مهارتهای خاصی متمرکز است.	معمولاً سطح وسیعی را شامل می‌شود.	۴- محتوای دربرگیرنده
مشخصات جزئی برای سؤالات لازم است.	استفاده از جدول مشخصات لازم است.	۵- چگونگی طرح آزمون
پرسشهای انتخابی کاربرد کمتری دارد.	پرسشهای انتخابی مناسبتر است.	۶- نوع سؤال
سطح تسلط در مقایسه با استانداردهای قابل قبول مشخص می‌شود.	سطح تسلط با تعیین موقعیت فرد در گروه مشخص می‌شود.	۷- استانداردهای تسلط
همه پرسشهای لازم اعم از دشوار یا آسان برای مشخص کردن حد تسلط مطرح می‌شود، هیچ کوششی برای تغییر دشواری سؤالات انجام نمی‌شود و بسته به حیطة موردنظر ممکن است آسان یا دشوار باشد.	پرسشها طوری انتخاب می‌شود که حداکثر قدرت تمیز را ایجاد کند و برای ایجاد تفاوت‌های بیشتر بین نمرات آزمون، سؤالات آسان حذف شده، دشواری سؤالات در حد متوسط است.	۸- مراحل انتخاب پرسش
روشهای معمول تعیین پایایی به علت پراکندگی کم نمرات مناسب نیست.	روشهای معمول تعیین پایایی به علت پراکندگی وسیع نمرات کارآمد است.	۹- پایایی آزمون

پس از تعیین استانداردها در هر آزمون، نمره هر فرد با آن استانداردها سنجیده می‌شود. اگر دانشجو در برخی موارد تسلط کافی کسب نکرده باشد می‌توان مجدداً برنامه‌ریزی درسی کرده، نقاط ضعف او را برطرف ساخت.

در جدول ۱-۲، مقایسه میان خصوصیات آزمون ملاک نسبی و ملاک مطلق نشان داده شده است.

با توجه به این توضیحات می‌توان گفت: آزمون کنکور از الگوی امتحان ملاک نسبی پیروی می‌کند در حالی که آزمونهای پایان هر نیمسال و آزمون علوم پایه نوعی آزمون ملاک مطلق است.

(و) تقسیم براساس کاربرد

براساس کاربرد، آزمون‌ها به چند دسته تقسیم می‌شوند که عبارتند از:

- ۱- آزمونهای پیشرفت تحصیلی یا بررسی^۱؛
- ۲- آزمونهای هوش و استعداد^۲؛
- ۳- آزمونهای رغبت^۳؛
- ۴- آزمونهای شخصیت^۴؛
- ۵- پیش‌آزمون^۵؛
- ۶- آزمون مهارت^۶؛
- ۷- آزمون تشخیصی^۷.

۱- آزمونهای پیشرفت تحصیلی

این آزمون‌ها برای ارزیابی پیشرفت دانشجویان و کارآیی تدریس و آموزش و همچنین رتبه‌بندی دانشجویان به کار می‌رود. در هر نظام آموزشی ابتدا باید اهداف آموزشی را به صورت رفتارهای قابل مشاهده و قابل ارزشیابی بیان کرد. این اهداف به صورت جزئی و کلی براساس اولویتهای آموزشی دسته‌بندی می‌شود. نظام آموزشی مطلوب نظامی است که به هدفهای اصلی و اساسی خود دست پیدا کند وگرنه باید در آن تجدیدنظر شود. برای اینکه بتوان میزان موفقیت را در دستیابی به اهداف موردنظر ارزیابی کرد، احتیاج به آزمونی است که براساس هدفهای اولیه طراحی شده باشد و به آن آزمون پیشرفت تحصیلی می‌گویند. این آزمون خصوصیتی دارد که در طرح آن باید مدنظر قرار گیرد از جمله:

- ۱- سطوح یادگیری را مطابق هدفهای آموزشی اندازه بگیرد.
- ۲- تمام مطالب تدریس شده و اهداف را در برگیرد. که جهت حصول اطمینان از این موضوع، برای طراحی آزمون، باید از جدول مشخصات استفاده کرد.
- ۳- سؤالات با اهداف موردنظر متناسب باشد. با توجه به هدف و ویژگیهای تدریس و آزمون، باید به دنبال بهترین نوع سؤال بود.

1- Achievement Test

2- Aptitude Test

3- Tendency Test

4- Personality Test

5- Pre - test

6- Proficiency Test

7- Diagnostic or Discriminatory Test

۵۰ آزمونهای پیشرفت تحصیلی در علوم پزشکی

۴- طراحی آزمون به نحوی باشد که بتوان از نتایج آن براساس هدفهای پیش‌بینی شده استفاده کرد.

۵- پایایی داشته باشد؛ برای افزایش پایایی می‌توان تعداد سؤالات آزمون را زیاد کرد و کیفیت سؤالات را نیز بهبود بخشید. البته باید در تفسیر نتایج محتاط بود و توجه داشت که در هر امتحان، به‌رحال اشکالاتی وجود دارد. بنابراین باید برای هر دانشجو محدوده نمره را تعیین کرد مثلاً: فردی که در امتحان نمره شانزده گرفته است با توجه به وضعیت حاکم بر امتحان احتمالاً محدوده نمره وی بین پانزده تا هفده خواهد بود. به این ترتیب تفاوت‌های جزئی بی‌اهمیت می‌گردد و تفسیر نتایج به واقعیت نزدیکتر می‌شود.

۶- نتایج آزمون باید بررسی و تجزیه و تحلیل شده، میزان دقت و کارایی پرسشها با توجه به پاسخهای دانشجویان تعیین شود. به این ترتیب میزان صحت نتایج آزمون مشخص می‌گردد. پس از بررسی نتایج آزمون و مقایسه آنها با اهداف اولیه، باید تغییرات لازم در نحوه تدریس و نظام آموزشی صورت گیرد تا یادگیری دانشجویان اصلاح شده، ارتقا یابد. با توجه به توضیحات فوق مشخص می‌شود که هدف از این امتحانات بهبود یادگیری دانشجو، تهیه بازخورد آموزشی و ارزیابی سطح پیشرفت فرد در مقایسه با برخی متوسطهای جامعه می‌باشد. البته این موضوع باید قبل از امتحان برای دانشجو مشخص شده باشد؛ زیرا دیدگاه وی در زمینه هدف امتحان می‌تواند در نتایج مؤثر بوده، تفسیر نتایج آن را دچار اشکال و اشتباه کند.

۲- آزمونهای رغبت

این آزمونها برای تعیین میزان علاقه افراد نسبت به مشاغل، رشته تحصیلی، سایر فعالیت‌های زندگی و به منظور راهنمایی افراد به سوی شغل و رشته تحصیلی مناسب طراحی می‌شود. به این منظور در مورد فعالیت‌های مشاغل و رشته‌های تحصیلی مختلف سؤالاتی مطرح و از فرد خواسته می‌شود براساس علائق خود به آنها پاسخ داده، اظهارنظر کند. این پاسخها با ملاکهای معین سنجیده و ارزیابی می‌گردد و براساس نتایج آن تصمیم‌گیری می‌شود.

۳- آزمونهای شخصیت

این آزمونها برای سنجش و اندازه‌گیری میزان سازگاری عاطفی، روابط اجتماعی و

انگیزه‌های رفتار فرد به کار می‌رود و می‌توان بعضی از ویژگیهای شخصیت را مثل درون‌گرایی، برون‌گرایی، افسردگی و ... با کمک آن اندازه‌گیری کرد. این آزمون‌ها بر پایه خصوصیات رفتاری و روانی انسان است بنابراین نمی‌تواند کاملاً دقیق و پایا باشد. شناخت این آزمون‌ها نیاز به توضیح بیشتر دارد که از حیطه این بحث خارج است.

۴- آزمونهای هوش و استعداد

در این آزمون‌ها براساس قدرت حافظه، استدلال، گفتار، نوشتار و سرعت سؤالاتی مطرح می‌شود که آزمون‌شونده به آنها پاسخ می‌گوید. نتایج حاصل با ملاکهای جامعه سنجیده شده، میزان استعداد فرد به طور نسبی مشخص می‌گردد. این آزمون‌ها بسیار متنوع است. برای اطلاعات بیشتر به کتابهایی که در این زمینه نوشته شده، مراجعه کنید.

۵- پیش‌آزمون

این آزمون برای تعیین میزان آگاهیهای دانشجو قبل از شروع هر دوره طراحی می‌گردد. بنابراین اطلاعات پایه و ابتدایی که برای ادامه مطالعه در برخی زمینه‌های خاص موردنیاز است در این آزمون سنجیده شده، میزان آمادگی دانشجو برای ورود به دوره بعد مشخص می‌گردد. این آزمون‌ها اغلب از نظر دشواری در سطح پایینی بوده، تفسیر نتایج آن نیز به صورت ملاک مطلق می‌باشد.

۶- آزمون مهارت

این آزمون برای سنجش معلوماتی است که در دوره همه افراد باید نسبت به آن آگاهی کامل داشته باشند. مثل: دانستن فیزیولوژی و آناتومی اعضای بدن انسان برای دانشجوی بالینی پزشکی. آزمون مهارت به قسمت معینی از درس محدود شده، معمولاً دشوار نیست و تفسیر نتایج آن نیز به شکل ملاک نسبی می‌باشد.

۷- آزمون تشخیصی

این آزمون برای تعیین مشکلات یادگیری به کار می‌رود. در آزمون تشخیصی نمره کل اهمیت چندانی نداشته، پاسخ به سؤالات خاصی مورد نظر است که نشان‌دهنده اشتباهات رایج دانشجویان می‌باشد. مثل: به کاربردن کلمات با تلفظ مشابه و

معانی مختلف. در حقیقت، در این آزمونها سعی بر این است که قسمتی از مطالب درسی که برای اکثر دانشجویان دشوار بوده، شناخته شود تا بتوان بر روی آن قسمت، بیشتر تأکید کرد.

ز) انواع آزمون با توجه به مدت زمان

آزمونها بر اساس محدودیت زمانی در نظر گرفته شده، به دو دسته تقسیم می‌شوند:

۱- آزمون سرعت؛

۲- آزمون پیشرفت.

۱- آزمون سرعت^۱

در این آزمون بر سرعت دانشجو در پاسخ‌دهی به سؤالات تأکید شده، نمره‌دهی بر همین اساس صورت می‌گیرد. این آزمون مشتمل بر تعدادی سؤال با دشواری متوسط است که دانشجو باید در زمان مشخص و محدود به آنها پاسخ گوید. فقط افرادی در این آزمون موفق می‌شوند که دانش و سرعت لازم را به صورت توأم با یکدیگر در اختیار داشته باشند. اغلب، آزمونهای معمولی استاندارد شده و آزمونهای معلم ساخته و گاهی سؤالات کوتاه پاسخ آزمونهای سرعت قرار می‌گیرد. تذکر این نکته لازم است که هدف آزمونهای تحصیلی و کلاسی، تعیین میزان یادگیری و پیشرفت تحصیلی دانشجو می‌باشد نه سرعت آن، بنابراین بهتر است از آزمون سرعت استفاده نشود. البته شکی نیست که در بعضی مهارتها سرعت از اهمیت بسزایی برخوردار است. مثلاً: در برخورد با بیمار مبتلا به آپاندیسیت تشخیص و درمان باید با سرعت انجام شود در غیر این صورت جان بیمار به مخاطره می‌افتد. برای سنجش میزان مهارت فرد در چنین حالاتی آزمونهای سرعت بسیار کارآمد است.

۲- آزمون قدرت^۲

در آزمون قدرت بر میزان یادگیری و پیشرفت تحصیلی دانشجو تأکید می‌شود و نتایج آن نشان‌دهنده میزان قدرت و توانمندی دانشجو است. این آزمون دارای تعدادی سؤال است، که به ترتیب بر سختی آنها افزوده می‌شود؛ به نحوی که هر سؤال نسبت به سؤال

قبل از اطلاعات بیشتری را برای پاسخگویی می‌طلبید. آزمون‌شونده سؤالات را به ترتیب پاسخ گفته، به سؤالاتی می‌رسد که پاسخگویی به آنها نیاز به دانشی بالاتر از حد معلومات وی دارد. بنابراین از آن مرحله به بعد نمی‌تواند سؤالات را پاسخ گوید. آزمون‌گیرنده با بررسی تعداد سؤالاتی که دانشجو به آنها پاسخ صحیح داده، می‌تواند به سطح یادگیری و معلومات او پی‌ببرد. این آزمون از نظر زمانی نباید محدودیت داشته باشد، به طوری که اکثر دانشجویان بتوانند به همه سؤالات پاسخ گویند. اگر در امتحانی بیش از ۱۰ درصد دانشجویان نتوانند در پایان وقت به همه سؤالات برسند نشان‌دهنده این است که سرعت نقش مهمتری را در نمره‌دهی ایفا می‌کند و همانطور که قبلاً ذکر گردید، اگر در آزمونهای تحصیلی هدف، سنجش میزان دانش دانشجو باشد، امتحان سرعت کارآمد نخواهد بود.

منابع:

- ۱- شهاب، گیتا. معرفی ارزشیابی تشخیصی. تهران: انتشارات دانشگاه سابق آزاد ایران، ۱۳۴۸.
- ۲- کوهن، لوئیز؛ مانیون، لارسن. راهنمای عملی تدریس. ترجمه فاطمه شاکری. تهران: انتشارات جهاد دانشگاهی، ۱۳۷۲.
- ۳- گرون لوند، نورمن. آزمونهای پیشرفت تحصیلی. تهران: انتشارات سازمان سنجش، ۱۳۵۶.
- ۴- گنجی، حمزه. اندازه‌گیری و ارزشیابی در آموزش و پرورش. تهران: انتشارات بعثت، ۱۳۷۳.
- ۵- لطف آبادی، حسین. سنجش و اندازه‌گیری در علوم تربیتی و روانشناسی. تهران: انتشارات فردوسی، ۱۳۷۳.
- ۶- مایزما، ویلیام؛ و همکاران. اندازه‌گیری و آزمون در تعلیم و تربیت. ترجمه غلامرضا خوئی نژاد. مشهد: انتشارات مؤسسه چاپ و انتشارات آستان قدس رضوی، ۱۳۷۲.
- 7- Al - Saeid K . *Evaluation of undergraduate students in paediatrics at Kuwait University.* Med Teach 1995; 17(3): 289-297.
- 8- Bowd SB. *Teaching in the health-related professions.* Durham: Carolina Academic Press, 1997.
- 9- Gronlund NE. *How to construct achievement tests.* 4th rev. ed. New jersey: Perntice - Hall, 1988.
- 10- Mehrans WA, Lehman IJ. *Measurement and evaluation in education and psychology.* 3rd ver.ed. New York: Holt-Saunders: 1984.
- 11- Vleuten V. *A written test as an alternative to performance testing.* Med Educ 1989; 23(1): 97-107.

بخش دوم

انواع آزمون

فصل سوم

آزمونهای کتبی مرسوم

- سوالات چندگزینه‌ای
- سوالات صحیح - غلط
- سوالات صحیح - غلط چندگزینه‌ای
- سوالات جورکردنی
- سوالات کوتاه پاسخ
- سوالات تشریحی

سوالات چندگزینه‌ای^۱

هر کدام از ما حداقل یکبار در آزمون‌های با سوالات چندگزینه‌ای شرکت کرده‌ایم. شرایط حاکم بر آن آگاهی داریم. آزمونهای چندگزینه‌ای ظرف بیست سال گذشته جای خود را در میان انواع آزمونها باز کرده، به طور چشمگیری همگانی شده است و البته علت کثرت استفاده از سوالات چندگزینه‌ای فقط سابقه تاریخی آنهاست و بتدریج در آزمونهای دانشگاهی، سوالات صحیح - غلط چندگزینه‌ای^۲ جایگزین می‌شود. با آن که بسیاری از دانشکده‌ها ضرورت وجود چنین آزمون‌هایی را زیر سؤال می‌برند؛ هیچکس انکار نمی‌کند که این شیوه بهترین روش برای ارزیابی اطلاعات واقعی^۳ است.

طراحی یک آزمون استاندارد چندگزینه‌ای احتیاج به تبحر و صرف زمان طولانی دارد. اما برای یک آزمون گیرنده ماهر، استفاده از آن مزایای فراوانی دارد (هم از نظر عینی بودن آزمون و هم از نظر سهولت نمره‌دهی و تفسیر نتایج).

اغلب به دلیل عدم مهارت طراحان سؤال، این پیش‌داوری برای دانشجویان وجود دارد که حتماً در هر امتحان چندگزینه‌ای سوالاتی مبهم و دو پهلو وجود خواهد داشت؛ البته این نگرانی زیاد هم بی‌مورد نیست. این مسأله مربوط به طراحی سؤال است نه خود سؤال. حال به عنوان یک آزمون گیرنده خوب چه تدبیری بیاندیشیم تا سوالات چندگزینه‌ای ما استاندارد باشد؟

تعریف سؤال چندگزینه‌ای

سؤال چندگزینه‌ای شامل دو بخش است:

- ۱- ساقه یا ریشه^۴؛ که موضوع مورد پرسش را مطرح می‌کند؛
- ۲- چند گزینه پیشنهادی که یکی از آنها درست یا بهترین پاسخ است و گزینه‌های غلط توجه دانشجویانی را که به پاسخهای خود اطمینان کافی ندارند، به خود جلب می‌کند.

1- Multiple Choice Question (MCQ)

2- Multiple True/False

3- Factual

4- Stem

انواع سؤالات چندگزینه‌ای

طراحی سؤالات چندگزینه‌ای به شیوه‌های مختلف صورت می‌پذیرد که بدانها اشاره می‌گردد:

۱- ساده‌ترین نوع آن بلی - خیر^۱ است که البته این آزمون یک آزمون چندگزینه‌ای کامل نیست. به این نوع، سؤال صحیح - غلط^۲ اطلاق می‌شود که در جای خود مورد بحث قرار خواهد گرفت.

مثال: منشاء پتانسیل استراحت غشا، یونهای K و Na هستند. بلی □ خیر □

۲- سؤال با چندگزینه که تعداد گزینه‌ها بین چهار تا شش است و معمول آن پنج گزینه می‌باشد. در میان این گزینه‌ها یکی صحیح یا بهترین پاسخ است.^۳

مثال: دیابت از لحاظ ژنتیکی جزء کدام دسته از بیماریهای زیر است؟ (بهترین گزینه را انتخاب کنید).

الف) تراتورژنیک

ب) پلی ژنیک

ج) کروموزومی

د) تک ژنی

در این سؤال گرچه گزینه‌های «الف» و «ب» هم صحیح هستند ولی گزینه «ج» بهترین پاسخ است.

مثال: کدام یک از هورمونهای زیر به طور مستقیم روی هسته سلول هدف اثر می‌کند؟

الف) تیروکسین

ب) کورتیزول

ج) هورمون رشد

د) سوماتوماموتروپین

ه) آدرنالین

در این سؤال فقط گزینه «الف» صحیح و بقیه گزینه‌ها کاملاً غلط است.

۳- مرتب کردن چند انتخاب به بهترین روش و ترتیب ممکن.

مثال: برای معاینه کبد بهترین ترتیب عمل چیست؟

مشاهده^۱ - لمس^۲ - دق^۳ - سمع^۴

در این سؤال دانشجو باید برحسب اولویت، عبارتهای بالا را مرتب کند.

۴- تشخیص نوعی رابطه بین علت و معلول.

مثال: «مزوتلیوما یک تومور شاخص در کارگران معادن آزبست است چون آزبست،

سلولهای مزوتلیال را سوراخ می‌کند» در مورد این دو جمله چه می‌توان گفت؟

الف) هر دو صحیح بوده، رابطه منطقی نیز با هم دارند.

ب) هر دو صحیح هستند اما ارتباطی با هم ندارند.

ج) اولی درست و دومی غلط است.

د) هر دو غلط است.

اغلب صاحب نظران، این شیوه طرح سؤال را بخصوص تحت شرایط فشار امتحان،

خیلی گنج کننده می‌دانند.

۵- برای کاهش میزان شانس در آزمون چندگزینه‌ای نوع دیگری از سؤال نیز رایج است که

نمونه‌ای از آن در زیر می‌آید:

مثال: شکستگی^۱ فوقانی تنه استخوان فمور همراه است با:

۱- کوتاه شدن اندام

۲- چرخش داخلی بخش Proximal

۳- Abduction قطعه Proximal

۴- Proximal extention قطعه

اگر ۱ و ۲ و ۳ درست است A را انتخاب کنید.

اگر ۱ و ۳ درست است B را انتخاب کنید.

اگر ۲ و ۴ درست است C را انتخاب کنید.

اگر ۴ درست است D را انتخاب کنید.

اگر همه درست است E را انتخاب کنید.

البته بسیاری از کارشناسان این نوع سؤالات را منطقی نمی‌دانند به علت اینکه مثلاً

اگر دانشجویی بداند که گزینه ۱ غلط و گزینه ۲ صحیح است بیدرنگ و بدون نگاه کردن

به سایر گزینه‌ها، گزینه C را انتخاب می‌کند. عکس این حالت هم پیش می‌آید. یعنی دانشجویی که می‌داند گزینه‌های ۱ و ۳ صحیح است، ولی در درست بودن گزینه ۲ شک دارد نمی‌تواند پاسخ صحیح را انتخاب کند. ایراد دیگر آن، وارد شدن اهداف اضافی در این نوع سؤالات است، بدین معنی که دانشجو در تمام طول خواندن سؤالات باید به یاد داشته باشد که کدام گزینه‌ها غلط و کدام درست است تا از جمع‌بندی آنها پاسخ صحیح را انتخاب کند.

۶- سؤالات صحیح - غلط چندگزینه‌ای که بعداً در مورد آن توضیح داده خواهد شد.

قواعد طراحی سؤالات چندگزینه‌ای

توانایی در نوشتن سؤالات چندگزینه‌ای خوب، نوعی هنر است. گفتن اینکه «چگونه ننویسید» یا «چه چیزهایی در سؤال شما نباشد» بسیار آسانتر از آن است که بگوییم «چه نکاتی را در نوشتن سؤال رعایت کنید تا یک سؤال خوب طرح شود» یا اینکه نمونه یک سؤال خوب را به شما ارائه دهیم!

هر سؤال چندگزینه‌ای باید چهار جنبه داشته باشد:

۱- مربوط بوده، با موضوع تناسب داشته باشد؛^۱

۲- کوتاه باشد؛^۲

۳- قابل فهم باشد؛^۳

۴- تمیز دهنده بین خوب و بد باشد.^۴

برای طراحی این آزمون می‌توانید از سه منبع استفاده کنید: کتب مرجع، مقالات مروری و گزارش موردی.^۵ از تحقیقات اخیر پزشکی اصلاً استفاده نکنید، مگر اینکه نتایج آن به طور وسیع منتشر شده باشد.

کیفیت یک سؤال چندگزینه‌ای به شیوایی جوابهای انحرافی آن بستگی دارد. طرح جوابهای انحرافی بسیار مشکل‌تر از جوابهای صحیح است. بسیار سخت است که مطمئن شوید جوابهای انحرافی واقعاً و صد درصد غلط است.

سؤالات را باید به تدریج و سرفرصت بعد از هر جلسه تدریس، طرح کرده، در فیشهای ۸×۱۲ سانتیمتر نگهداری کنید. بعداً در موقع لزوم از بین این سؤالات، پرسشهای

1- Relevant

2- Short

3- Understandable

4- Discriminative

5- Case Report

امتحانی را انتخاب نمایید.

در طرح سؤال، ابتدا باید ریشه سؤال، بعد پاسخ صحیح و در نهایت پاسخهای انحرافی نوشته شود.

در این کتاب قواعد نوشتن سؤالات به سه بخش تقسیم شده، سپس هر بخش به تفصیل توضیح داده خواهد شد.

الف) قواعد مربوط به نوشتن ریشه سؤال

ب) قواعد مربوط به نوشتن گزینه‌ها

ج) نکات دیگر

الف) قواعد مربوط به نوشتن ریشه سؤال

۱- هر پرسش باید یکی از اهداف مهم درس را بسنجد.

مثال: سؤال زیر در آزمون واحد تغذیه داروسازی طرح شده است:

نام دیگر ویتامین ب ۱ چیست؟

الف) آنتی نوریتیک

ب) بیوتین

ج) نیاسین

د) پانتوتنیک اسید

اگر دانستن نام فرعی ویتامین ب ۱ جزء اهداف موردنظر در گذراندن درس تغذیه نباشد، سؤال مزبور نامناسب است.

۲- ریشه سؤال به دو نوع پرسشی و ناقص نوشته می‌شود که نوع اول بهتر است زیرا در نوع ناقص دانشجو مجبور است هر گزینه را به طور جداگانه در ریشه گذاشته آن را کامل کند. در این نوع بدون آنکه سطح بالاتری از یادگیری مورد ارزیابی قرار گیرد، سؤال پیچیده‌تر شده، نیاز به صرف وقت بیشتری دارد.

مثال: در ریه در اثر تحریک اعصاب پاراسمپاتیک بوجود می‌آید.

الف) تنگی قطر برونشها

ب) کاهش ترشح موکوس

ج) افزایش تعداد حرکات تنفسی

د) تنگی قطر رگهای خونی

مدت زمان خواندن این سؤال خیلی طولانی‌تر از وقتی است که سؤال به این شکل نوشته شود: «تحریک پاراسمپاتیک در ریه باعث کدامیک از حالات زیر می‌شود؟». اگر آوردن نقطه‌چین را ضروری می‌بینید، آن را در آخر عبارت بیاورید در این صورت باید ریشه سؤال به اضافه هر پاسخ، یک جمله کامل از لحاظ دستوری باشد.

مثال: تحریک شبکه عصبی آنتریک موجب ...

الف) تشدید حرکات مخلوط کننده روده باریک می‌شود.

ب) قطع حرکات مخلوط کننده روده باریک می‌شود.

ج) تغییر در حرکات مخلوط کننده روده باریک نمی‌شود.

د) تخفیف حرکات مخلوط کننده روده باریک می‌شود.

۳- ریشه سؤال باید کوتاه بوده، موضوع را مستقیماً مطرح کند. همچنین، پرهیز از به کار بردن مطالب غیر ضروری، زمان خواندن سؤال را به حداقل می‌رساند.

مثال: به سؤال زیر توجه کنید.

در مورد با آزمون توبرکولین (تست مانتو) که برای ارزیابی ایمنی نسبت به میکروب سل به کار می‌رود کدام گزینه صحیح است؟

وظیفه دانشجو است که مورد استفاده تست توبرکولین را بداند و توضیح راجع به آزمون توبرکولین در این نمونه سؤال ضرورتی ندارد.

۴- سؤال به زبان ساده بیان شود، طوری که به صورتهای مختلف، بخصوص به وسیله دانشجویان باهوش تفسیر نشود.

مثال: در مورد عضلات کدام عبارت صحیح است؟

الف) در انقباض عضله صاف، کالמודولین نقش تروپونین C را ایفا می‌کند.

ب) خروج کلسیم از شبکه سارکوپلاسمی، به طریق انتقال فعال و بازگشت آن به شبکه آندوپلاسمی به صورت انتشار است.

ج) محلهای فعال در عضله مخطط در روی رشته‌های میوزین قرار دارند.

د) فیبرهای عضلانی سریع در عضله مخطط ضخیم‌ترند و میتوکندری کمتری دارند.

در این سؤال که یکی از پرسشهای فیزیولوژی امتحان جامع علوم پایه می‌باشد گزینه‌های «الف» و «د» می‌تواند توسط دانشجویان خوب، گزینه‌های انتخابی باشد. در حالی که گزینه «د» به عنوان گزینه صحیح اعلام شده است. طراح سؤال به دلیل متفاوت بودن مکانیسم عمل کالמודولین و تروپونین C حتی با وجود یکسان بودن نقش

این دو گزینه «الف» را نادرست در نظر گرفته است.

۵- عوامل مشترک و کلمات یکسان گزینه‌ها را در ریشه سؤال بیاورید.

مثال: سومین قسمت دثودنوم:

الف) در مقابل کلیه راست است.

ب) در مقابل پانکراس است.

ج) در مقابل کولون عرضی است.

د) در مقابل معده است.

کلمه «در مقابل» بین تمام گزینه‌ها مشترک است و باید در صورت ساقه سؤال منظور شود.

۶- نکات دستوری را در ریشه سؤال و کلیه گزینه‌ها بخصوص گزینه‌های غلط رعایت کنید

وگرنه همین مطالب دانشجو را به سوی پاسخ صحیح راهنمایی می‌کند.

مثال:

In portal hypertension we may have an:

- a) esophageal varicosa
- b) pulmonary embolism
- c) coronary thrombosis
- d) sepsis

در این سؤال «an» دانشجو را فقط به سمت گزینه «a» هدایت می‌کند، بدون اینکه از درست یا غلط بودن بقیه گزینه‌ها اطلاع داشته باشد.

۷- ریشه؛ فقط یک مطلب را به طور روشن مورد سؤال قرار دهد. ریشه‌هایی که دو عبارت یا بیشتر را در بر می‌گیرد همراه‌کننده و مبهم است.

مثال: علل هیپوناترمی و هیپرکالمی کدام است؟

در این سؤال مشخص نیست که علل هیپوناترمی و هیپرکالمی را باید همزمان مورد بحث قرار داد یا به طور جداگانه.

۸- ریشه سؤال باید مثبت باشد. در صورتی که به کار بردن نوع منفی را ضروری می‌بینید

کلمه منفی را مشخص کنید مثلاً آن را بزرگتر یا پررنگتر بنویسید یا زیر آن خط بکشید.

از به کار بردن افعال منفی بخصوص منفی مضاعف خودداری کنید.

مثال: درباره منیسکهای زانو (غضروفهای مفصل زانو) کدام گزینه نادرست است؟

(الف) منیسکهای مدیال تقریباً به شکل نیم‌دایره است.

(ب) منیسک لترال تقریباً دایره‌ای شکل است.

(ج) منیسک مدیال در قدام پهن‌تر از خلف است.

(د) منیسک لترال متحرک‌تر از منیسک مدیال است.

در این سؤال حتماً باید زیر کلمه «نادرست» خط بکشید تا دانشجوی بااطلاع به خاطر بی‌توجهی به آن دچار اشکال نشود.

۹- هر پرسش، مستقل از سؤال دیگر بوده، پاسخ دادن به هیچ پرسشی مستلزم پاسخ به پرسش دیگر نباشد.

مثال:

۱- فشار سیستولی قلبی 120 mmHg و فشار دیاستولی آن 70 mmHg می‌باشد، فشار متوسط شریانی چقدر است؟

الف $\frac{250}{3}\text{ mmHg}$

ب $\frac{260}{3}\text{ mmHg}$

ج $\frac{240}{3}\text{ mmHg}$

د $\frac{250}{2}\text{ mmHg}$

۲- با توجه به سؤال بالا و در نظر گرفتن جریان خون به میزان $5/6$ لیتر در دقیقه، مقاومت رگهای محیطی را حساب کنید.

در این مثال پاسخ دادن به سؤال دوم مستلزم پاسخگویی به سؤال اول است و دانشجویی که قادر به یافتن پاسخ سؤال اول نبوده، به ناچار به سؤال دوم هم نمی‌تواند جواب دهد.

۱۰- پاسخ صحیح یک سؤال از روی سؤال دیگر مشخص نشود.

مثال: در آزمونی دو سؤال، به این صورت وجود دارد:

۱- کدام عبارت در مورد انتقال اکسیژن درست است؟

(الف) حمل اکسیژن در خون بیشتر توسط پروتئین‌های پلاسما انجام می‌شود.

(ب) علت اصلی خفگی به وسیله گاز مونواکسیدکربن اثرمهارى این گاز بر مرکز تنفسی است.

(ج) ضویب انتشار اکسیژن بیست برابر ضریب انتشار گاز دی‌اکسیدکربن است.

د) افزایش دما و افزایش فشار دی‌اکسیدکربن هر دو موجب انحراف منحنی تفکیک اکسی‌هموگلوبین به راست می‌گردد.

۲- کدامیک از عوامل زیر باعث انحراف منحنی تجزیه اکسی‌هموگلوبین به چپ می‌گردد؟
الف) افزایش دما

ب) افزایش دی‌اکسیدکربن

ج) افزایش هموگلوبین جنینی

د) افزایش غلظت آنزیم DPG - 2, 3 (۲ و ۳ دی فسفوگلیسرات)

در سؤال اول، گزینه «د» گزینه صحیح است و به راحتی با توجه به غلط بودن گزینه‌های دیگر انتخاب می‌شود. دانشجویی که توانسته صحیح بودن گزینه «د» را در سؤال اول تشخیص دهد، بی‌شک در سؤال بعدی، گزینه‌های «الف» و «ب» را انتخاب نمی‌کند و بین گزینه‌های «ج» و «د» به انتخاب خواهد پرداخت.

۱۱- وقتی در مورد یک سؤال عقاید مختلفی وجود دارد، باید مشخص کنید کدام نظریه مورد سؤال است.

مثال: در مورد رفتار، کدام گزینه صحیح است؟

الف) رفتار، ناشی از زمینه ناخودآگاه انسان است.

ب) رفتار تحت تأثیر محیط است.

ج) رفتار تحت تأثیر ناخودآگاه و محیط است.

د) رفتار ناشی از فطرت انسان است.

در این مثال هر گزینه‌ای می‌تواند به نظر یکی از روانشناسان صحیح باشد. بنابراین حتماً باید ذکر کنید که نظریه کدام روانشناس مورد سؤال است.

اگر درصد دقیق وقوع، مقدار و مانند اینها در هر موضوع در منابع مختلف فرق می‌کند، در سؤال خود از آنها استفاده نکنید.

۱۲- از بکار بردن علائم اختصاری چه در ریشه و چه در گزینه‌ها خودداری کنید.

مثال: در CMCC اختلال در کدامیک از سلولهای زیر وجود دارد؟

الف) B Cell

ب) T Cell

ج) NK Cell

د) فاگوسیت‌ها

در صورت این سؤال باید ذکر شود که منظور از CMCC، chronic muco cutaneous candidiasis (کاندیدیاژ جلدی - مخاطی مزمن) است. در هنگام امتحان، منحرف کردن ذهن دانشجو در زمینه مفاهیم اصطلاحات مختصر صحیح نیست.

۱۳- وقتی ریشه سؤال را به صورت یک مورد بیمار^۱ مطرح کرده، نوع بیماری را مورد سؤال قرار داده‌اید، حتماً مقادیر طبیعی را در ریشه ذکر کنید.

۱۴- سعی کنید کلماتی که در ریشه به کار می‌برید برای همه گزینه‌ها یک نقش دستوری واحد داشته باشد.

مثال: در این پرسش بیلی‌روبین برای دو گزینه نقشهای متفاوت دارد.

Bilirubin

a) Is a breakdown product of myoglobin.

b) concentration in the serum increases in cirrhosis.

۱۵- زمان افعال گزینه‌ها باید یکسان بوده، بهتر است در گزینه‌ها زمان حال را به کار برید.

۱۶- پاسخ به سؤال «کدامیک از موارد زیر» به جای آنکه انتخاب جواب صحیح یا غلط در بین تعدادی از گزینه‌ها باشد به طور منطقی انتخابی بین دو جواب است، از نظر دستوری معمولاً فعل این جمله (می‌باشد یا می‌باشند) است که دانشجو را راهنمایی می‌کند تا گزینه با فعل جمع یا مفرد را انتخاب کند. بنابراین در چنین مواردی از به کار بردن این اصطلاح در ریشه سؤال خودداری کنید.

ب) قواعد مربوطه به نوشتن گزینه‌ها

۱- هر یک از پاسخهای پیشنهادی باید برای دانشجویی که معلومات کافی در زمینه آزمون ندارد قابل قبول باشد. گزینه‌های انحرافی را برای دانشجویان غیر مطلع به صورت دقیق بنویسید و مطمئن شوید که حتماً یکی از گزینه‌ها کاملاً صحیح یا بهترین است.

مثال: عصب رادیال به کدامیک از عضلات زیر عصب می‌دهد؟

الف) گاستروکنمیوس

ب) سولئوس

ج) براکیورادیالیس

د) تیمیالیس قدامی

گزینه‌های «الف» و «ب» و «د» هر سه، عضلات پا بوده، در حالی که عصب رادیال در دست است به علاوه در عضله براکیورادیالیس تشابه لفظی با عصب رادیال وجود داشته دانشجوی غیر مطلع را به سمت جواب صحیح هدایت می‌کند.

گزینه‌های انحرافی را می‌توانید به روشهای زیر تهیه کنید:

الف) اشتباهات رایج دانشجویان را در گزینه قرار دهید. مثال این تصور غلط که: «تراتوم تستیکولار به طور معمول به غده‌های لنفاوی اینگوینال مناساز می‌دهد».

ب) بیانی شبیه بیان دانشجو در این گزینه‌ها بکار برید.

ج) از لغات مثبت خوشایند استفاده کنید. مانند: گاهی اوقات، ممکن است، در بعضی موارد.

د) از جمله‌های قالبی و پاسخهای به ظاهر علمی البته نه به صورت گمراه‌کننده استفاده کنید.

۲- گزینه صحیح و گزینه‌های غلط را تا حد امکان با طول یکسان بنویسید.

مثال: مرکز خشم در سیستم لیمبیک کدام است؟

الف) هسته Preoptic

ب) کیاسمای Optic (بینایی)

ج) هسته امیگدال - هیپوکامپ و هیپوتالاموس

د) تالاموس

در این سؤال گزینه «ج» که پاسخ صحیح است از تمام گزینه‌ها طولانی‌تر بوده، طبیعتاً توجه دانشجویان را جلب خواهد کرد.

۳- برای غلط جلوه دادن گزینه‌ها توضیحات اضافی ندهید. جمله در اصل باید صحیح یا غلط باشد.

مثال: همه گزینه‌های زیر در مورد گلوکز صحیح است بجز:

الف) یک قند شش کربنه است.

ب) قند مصرفی مغز است.

ج) نام دیگر آن آرابینوز است و یک مولکول آن سی مولکول ATP می‌دهد.

د) یک آلدوهگزوز است.

گزینه «ج» گزینه غلط بوده، با توجه به توضیح اضافی غلط، به راحتی به عنوان گزینه مورد نظر انتخاب می‌شود.

۴- تا حد امکان از بکار بردن «همه موارد فوق» اجتناب کرده، در مورد به کار بردن «هیچیک از موارد فوق» با احتیاط عمل کنید.

هشال: IgG (ایمونوگلوبولین G):

(الف) از جفت عبور می‌کند.

(ب) دو مولکول آن یک مولکول کمپلمان را فعال می‌کند.

(ج) یک Ig ناقص است.

(د) همه موارد فوق.

در این نمونه اگر دانشجویی بداند که دو مورد از گزینه‌ها صحیح است بی تردید گزینه «د» را انتخاب می‌کند. با این کار شما ضریب شانسی او را از ۲۵ درصد به ۵۰ درصد رسانیده‌اید.

۵- برای غلط جلوه دادن گزینه‌ها از قیود قطعی مانند «هرگز، همیشه و فقط» استفاده نکنید، چون به طور معمول جواب نادرست را نشان می‌دهد.

از جملات معروف Virgil Moon آن است که:

«همیشه به یاد داشته باشید که هرگز قید همیشه و هرگز را به کار نبرید».

هشال: در مورد پدیده هم انتقالی (Cotransport) کدام جمله صحیح است؟

(الف) سرعت انتقال مستقل از تعداد حامل است.

(ب) فقط دو یون را به طور همزمان انتقال می‌دهد.

(ج) در تمام انواع سلولهای بدن انجام می‌شود.

(د) از پدیده اشباع‌پذیری پیروی می‌کند.

در گزینه‌های «ب» و «ج» از کلمات قطعی استفاده شده که غلط بودن گزینه‌ها را نشان می‌دهد.

۶- از واژه‌های عام مثل «گاهی اوقات» و «ممکن است» استفاده نکنید چون به طور معمول جمله درست را نشان می‌دهد.

هشال: در مورد اندوکاردیت^۱ کدام گزینه صحیح است؟

(الف) انتروکوکسی از نوع آلفا همولیتیک، معمول‌ترین عامل باکتریایی آن است.

(ب) در موارد کشت مثبت اندوکاردیت به سه سری کشت خون برای جدا کردن پاتوژن احتیاج داریم.

(ج) اگر به وسیله استرپتوکوکسی ایجاد شود ممکن است بوسیله مجموعه پنی‌سیلین و

جنتامایسین درمان شود.

د) عوامل غیر باکتریایی اندوکاردیت بسیار نادر است.

- با توجه به قاطعیت در همه گزینه‌ها به جز گزینه «ج» که در آن عبارت «ممکن است» به کار رفته کاملاً مشخص است که گزینه «ج» پاسخ مورد نظر است.
- ۷- دو گزینه را که یکی خیلی جامع و دیگری خیلی جزئی است به طور متقابل به کار نبرید.

مثال: غلظت سرمی فروپروتئین آلفا در کدام مورد افزایش می‌یابد؟

۱- همه تومورهای بدخیم

۲- آدنوکارسینومای پانکراس

در این سؤال گزینه «۱» یک پاسخ کلی و گزینه «۲» در حقیقت جزئی از گزینه «۱» است، پس هر دو گزینه صحیح می‌باشد.

۸- از دو عبارت متضاد در گزینه‌ها استفاده نکنید چون سؤال را به شکل صحیح - غلط در می‌آورد.

مثال: در هنگام مصرف انسولین زیاد چه اتفاقی رخ می‌دهد؟

الف) قند خون کم می‌شود.

ب) قند خون زیاد می‌شود.

ج) کلسیم خون کم می‌شود.

د) پتاسیم خون کم می‌شود.

مشخص است که پاسخ صحیح، یکی از گزینه‌های «الف» و «ب» است، چون این دو گزینه متضاد یکدیگر هستند.

۹- از به کار بردن گزینه‌های مترادف و دربرگیرنده یکدیگر پرهیز کنید.

مثال: هورمون رشد از کجا ترشح می‌شود؟

الف) هیپوتالاموس

ب) هیپوفیز قدامی

ج) آدنو هیپوفیز

د) هیپوفیز

در این سؤال گزینه‌های «ب» و «ج» مترادف بوده، گزینه «د» نیز در برگیرنده گزینه‌های «ب» و «ج» می‌باشد بنابراین پاسخ صحیح سؤال مبهم است.

۱۰- از به کار بردن کلمات مشابه^۱ در سؤال که می‌تواند دانشجویان را به انتخاب پاسخ صحیح یا حذف گزینه‌های غلط راهنمایی کند، خودداری کنید.

مثال ۱: در ضربان قلب کدام گره پیشاهنگ است؟

الف) گره سینوسی دهلیزی

ب) دسته هیس

ج) فیبرهای پورکنژ

د) فیبرهای اتصالی

تشابه کلمه گره در ریشه و گزینه «الف»، صحیح بودن گزینه «الف» را حتمی می‌سازد.

مثال ۲: لوله قفسه سینه که برای بیمار هموتوراکس استفاده می‌شود برای خارج کردن

چیست؟

الف) عفونت

ب) هوا

ج) مایع

د) خون

در این سؤال، اصطلاح «همو» دانشجو را به راحتی به سمت گزینه «د» هدایت می‌کند.

۱۱- پاسخ صحیح را به صورت همان عبارتی که در کتاب آمده است به کار نبرده، از جملات کلیشه‌ای استفاده نکنید.

مثال: در مورد غده هیپوفیز کدام گزینه درست است؟

الف) عمل هیپوفیز قدامی مستقل از هیپوتالاموس است.

ب) مهمترین غده آندوکراین بدن است.

ج) در انسان از لوب میانی آن هورمون محرکه ملانوسیت‌ها ترشح می‌شود.

د) هورمون آنتی‌دیورتیک از هیپوفیز قدامی ترشح می‌شود.

گزینه «ب» یک عبارت استاندارد و قالبی است.

۱۲- گزینه‌هایی که در بردارنده دو عبارت باشد باعث ایجاد سردرگمی در دانشجویان می‌گردد. تا حد امکان از به کار بردن این نوع گزینه‌ها اجتناب کنید.

مثال: سرطان معده:

۱- به طور خاص در بیماران مبتلا به آنمی خطرناک^۱ اتفاق می افتد و از علایم آن، تأخیر تخلیه معده می باشد.
در این گزینه دانشجوی باید علاوه بر تشخیص صحت دو جمله رابطه آنها را نیز بررسی کند.

۱۳- برای طرح یک گزینه ساده، جمله ای را که به طور واضح صحیح است منفی نکنید.

مثال: آنورت بزرگترین شریان بدن نیست.

۱۴- گزینه ها را با حروف (الف و ب و ... یا a, b, ...) بخصوص زمانی که پاسخها عدد است مشخص کنید.

۱۵- اگر پاسخها عددی است ترتیب آنها را رعایت کنید؛ زیرا در غیر این صورت سبب گمراه شدن دانشجو می شود.

مثال: خط Subcostal (زیر دنده ای) در مقابل کدام مهره قرار دارد؟

الف) L1 ب) L2 ج) L3 د) L4

۱۶- نکات دستوری را در همه گزینه ها بخصوص گزینه های غلط رعایت کرده، آنها را از لحاظ دستوری، هماهنگ با ریشه بنویسید.

مثال: برنامه درمانی اسهال در مراقبتهای بهداشتی اولیه^۲ با کدام روش انجام می شود؟
الف) بیمار را در بخش مربوط بستری کنید و انجام اقدامات لازم.

ب) تجویز آنتی بیوتیک خوراکی یا تزریقی هر چه زودتر.

ج) اختلالات آب و الکترولیت به وسیله محلولهای خوراکی اصلاح می گردد.

د) قطع کلیه مواد خوراکی از جمله مایعات و جامدات تا بهبود اسهال به مدت چند روز.

گزینه «ج» که پاسخ صحیح است از لحاظ دستوری صحیح تر و خوشایندتر از بقیه گزینه ها بوده، راهنمای خوبی برای انتخاب محسوب می شود.

ج) نکات دیگر

۱- بهتر است تعداد گزینه ها بین چهار تا شش عدد باشد. البته هر چه تعداد گزینه ها بیشتر باشد، شانسی دانشجوی برای رسیدن به پاسخ صحیح از روی حدس کمتر می شود. اما پیدا کردن پنج گزینه قابل قبول برای یک پرسش غالباً دشوار بوده، اضافه کردن پاسخهایی

1- Pernicious Anemia

2- PHC: Primary Health Care

که آشکارا نادرست است نیز کیفیت پرسش را بهتر نکرده، باعث افت کیفی آن نیز می‌گردد.

۲- ثابت نگه‌داشتن تعداد گزینه‌ها در همه پرسشهای آزمون ضرورت دارد.

۳- محل جواب سؤالهای مختلف را به صورت تصادفی انتخاب کنید. برای اطمینان از این مطلب روی پنج فیش جداگانه حروف «الف» و «ب» و «ج» و «د» و «ه» را بنویسید و برای هر سؤال فیشها را مخلوط کرده یکی را انتخاب کنید تا محل گزینه صحیح مشخص گردد.

۴- ابتدا تعدادی سؤال کوتاه پاسخ^۱ به دانشجویان داده، از جوابهای غلط آنها برای طرح گزینه‌های انحرافی مناسب و جالب در آزمون چندگزینه‌ای استفاده کنید.

۵- گمان نکنید در آزمون چندگزینه‌ای فقط باید از مطالب جزئی سؤال طرح کنید، بلکه بهتر است پرسشها تا حد امکان کلی باشد.

۶- هر پرسش و هریک از گزینه‌ها را در یک خط افقی جداگانه بنویسید تا خواندن آن آسانتر شود.

۷- ریشه را به طور کامل بیان کنید، سپس گزینه‌ها را بنویسید. نوشتن سؤال به صورت زیر غلط است:

مثال: عضله:

الف) لاتیسیموس دورسی (latissimus dorsi)

ب) پکتورالیس ماژور

ج) دلتوئید

د) ساب‌کلاویوس

در ابدوکسیون مفصل شانه نقش دارد.

۸- سؤال نباید نظر شخصی دانشجو را بخواهد.

مثال: به نظر شما مناسب‌ترین اثر پمپ سدیم چیست؟

الف) ایجاد انقباض در عضله صاف

ب) جلوگیری از تورم سلول

ج) ایجاد پتانسیل غشا

د) انتقال ایمپالس عصبی

واضح است که دانشجویان نظریات مختلفی نسبت به مناسب‌ترین اثر پمپ سدیم دارند. سؤال شما باید قاطعیت داشته، گزینه موردنظر هم حتماً درست باشد.

۹- هدف از طرح پرسش باید سنجش دانشهای اصلی و کاربردی باشد. طرح یک سؤال مثل «معمول‌ترین گروه خون در بیماران مبتلا به سرطان معده چیست؟» آسان است ولی ارزش تشخیصی و درمانی ندارد. یا نیازی نیست که جراحان بدانند «کورتیکوتروپین یک تری پپتید است» در حالی که دانستن «منشا عصب جلدی خارجی ران» برای آنها بسیار حائز اهمیت است.

۱۰- سؤالات، همه باید در برگیرنده محتوای موردنظر بوده، بین سؤالات مربوط به بخشهای مختلف درس، تعادل برقرار باشد. وقتی آزمونی مانند امتحان جامع علوم پایه بخشهای مختلف دارد که هر قسمت طراح جداگانه می‌طلبد، احتمال تکراری شدن سؤالات زیاد است. مثلاً: در آسیب‌شناسی و میکروپشناسی امکان وجود سؤالات مشابه در زمینه سل و عفونت هلیکوباکتر وجود دارد. در این حالت باید کل سؤالات آزمون پس از طراحی توسط یک کارشناس یا گروه طراح سؤالات بررسی و در صورت لزوم بازنگری شود.

۱۱- اگر برای طراحی سؤال، از بانک سؤال استفاده می‌کنید، باید محدوده سرفصلهایی را که در برگیرنده سؤالات انتخابی است مورد توجه قرار دهید. به عنوان مثال: در درس آسیب‌شناسی اگر سؤالات را به طور تصادفی انتخاب کنید ممکن است به جای توزیع طبیعی، چهل و پنج سؤال در مورد نئوپلازی و فقط پنج سؤال در مورد حالات بیماری‌زا متابولیک و التهابی داشته باشید. اگر محدوده موضوع کنترل نشود، در تفسیر نتایج دچار مشکل می‌شوید.

۱۲- بهتر است موارد کاربردی در سؤال گنجانده شده، تا حد ممکن از پرسیدن مطالب حفظی خودداری شود. مثلاً به جای پرسش «عمل مهم املاح صفراوی چیست؟» بهتر است بپرسیم: «انسداد مجاری صفراوی باعث چه اختلالاتی می‌شود؟».

۱۳- طرح سؤال در زمینه‌های تغییرات روز علم پزشکی و کشفیات جدید و یا سؤالاتی که به اصطلاح تاریخ مصرف آنها گذشته است، موجب سردرگمی خواهد شد.

۱۴- برای ارتقای سؤال به سطوح بالاتر یادگیری و حیطة مهارت، بهتر است از تصاویر فتوگراف، رادیوگراف، شکل، نمودار، مطالب مقدماتی^۱، فیلم و صداهای قلبی و تنفسی

استفاده نمایید.

۱۵- از آنجا که یافتن منبع طرح سؤال مشکل است، می‌توانید از آزمونهای چندگزینه‌ای با سابقه و اجرا شده در یک دانشکده مهم و یا بانک سؤالات استفاده کنید. اما باید به عناوین دربرگیرنده، شیوه امتحانی آن و ضریب تمیز و ضریب دشواری توجه کرده، سؤالات را با توجه به هدفهای موردنظر خود انتخاب نمایید.

۱۶- معمولاً امتحاناتی مثل امتحانات جامع و تخصصی توسط گروهی از آزمون‌گیرندگان طرح می‌شود. بنابراین توافق برسر محتوا و کلمات صورت سؤال، خیلی مشکل و وقت‌گیر است. روش پیشنهادی آن است که آزمون‌گیرندگان، خود به سؤالات پیشنهادی پاسخ گویند تا مشخص شود آیا همگی برداشت یکسانی از صورت سؤال داشته، به جواب صحیح رسیده‌اند یا خیر. در پایان، سؤالات را بازبینی کرده، از لحاظ نکات دستوری، ابهام عبارات، علامتهای اختصاری، بیان مقادیر طبیعی اطلاعات آزمایشگاهی، جملات منفی، منفی در منفی و اشکالات املائی بررسی کنید.

مثال: اکنون به یک سؤال چندگزینه‌ای غیر قابل قبول توجه کنید:

۱- فیبروز کیستیک:

الف) هرگز اختلال در غدد اندوکرین ایجاد نمی‌کند (به کار بردن قید قطعی).

ب) به عنوان یک اختلال اتوزومی غالب در کروموزم پنج به ارث می‌رسد. (دو مطلب در آن سؤال شده است).

ج) بیماران، NaCl کمتری در عرق دارند. (وضعیت دستوری ریشه با این گزینه هماهنگی ندارد. باید به این صورت باشد: «بیماری است که در آن NaCl کمتری در عرق وجود دارد».)

د) در ۱۰ درصد موارد باعث سیروز کیسه صفرا می‌شود. (در گزینه‌ها نباید به یک سری اعدادی که فقط از تحقیقات به دست آمده و قطعی نیست استناد کنید.)

نکات اجرایی آزمون چندگزینه‌ای

قبل از اجرای هر امتحان چندگزینه‌ای توجه به نکات زیر ضروری است.

۱- استاندارد موردنظر امتحان: باید بدانید که قصد شما از اجرای آزمون چیست؟ بعد از اعلام نتایج می‌خواهید چه کنید؟ آیا هدف شما، جدا کردن دانشجویان توانا و ناتوان است؟ آیا می‌خواهید دانشجویان را مورد تشویق قرار دهید یا به آنها پیشنهاد استراحت بدهید؟ البته استاندارد آزمون نسبت به نوع آن متفاوت است. به عنوان مثال به طور

معمول در آزمونهای درون دوره^۱ هدف، اصلاح روند آموزش بوده، تعیین میزان پیشرفت دانشجو و هدایت دانشجو در مسیر جدید دارای اولویت است.

نمره قبولی برای آزمون چندگزینه‌ای باید بالا باشد، چون دانشهای پایه را اندازه می‌گیرد؛ یعنی آنچه را همه دانشجویان باید بدانند، اندازه می‌گیرد. پس بهتر است نمره قبولی حدود ۸۰ تا ۹۰ درصد نمره کلی باشد. بنابراین به جای آنکه امتحان مشکلی از دانشجویان بگیرد و نمره قبولی حدود ۵۰ تا ۶۰ درصد برای آن قائل شوید، امتحانی آسان گرفته، در عوض نمره قبولی را حدود ۸۰ تا ۹۰ درصد قرار دهید.

۲- تعیین تعداد سؤالات: تعداد قابل قبول سؤالات در یک ساعت امتحان بستگی به استاندارد امتحان و پیچیدگی آن دارد. آزمون چندگزینه‌ای حداقل ده سؤال دارد. چهل سؤال مناسب است اما شصت سؤال ارجحیت دارد. این آزمون با حدود پنجاه تا یکصد سؤال از پایایی بسیار بالایی برخوردار است.

۳- مدت زمان امتحان: برای امتحانات دانشگاهی یک سؤال در هر دقیقه مناسب است. به علاوه باید به ازای هر پنجاه سؤال، ده دقیقه دیگر به زمان امتحان اضافه کرد. (شصت دقیقه برای پنجاه سؤال) البته در برخی آزمونهای دانشگاهی به ازای هر سی سؤال، شصت دقیقه در نظر می‌گیرند.

۴- منبع سؤالات: اطلاع دانشجویان از منبع طرح سؤالات ضروری است. از سوی دیگر اگر از بانک سؤالی استفاده می‌کنید که بین ۶۰۰ تا ۸۰۰ سؤال دارد می‌توانید آن را چاپ کنید. طبیعتاً وقتی سؤالات را منتشر نمایید پاسخ‌گویی به آنها آسان‌تر خواهد شد. بنابراین در ارزیابی‌هایی که در آنها از سؤالاتی که قبلاً چاپ شده استفاده می‌کنید مورد فوق نکته توجه داشته باشید. البته فراموش نکنید که افراد با توانایی‌های پایین و اطلاعات اندک کمتر از افراد باهوش و زرنگ توانایی استفاده از چنین کتابهایی را دارند.

۵- کنترل کیفی امتحان: شیوه بررسی و برآورد کلی راجع به نحوه برگزاری آن را مدنظر قرار می‌دهد.

موارد استفاده از سؤالات چندگزینه‌ای

در موارد زیر سؤالات چندگزینه‌ای بیشتر مورد استفاده قرار می‌گیرند:

۱- وقتی تعداد دانشجویان زیاد باشد؛

- ۲- هدف آزمون اندازه‌گیری حجم وسیعی از اطلاعات باشد؛
- ۳- پایایی بالای نمرات مدنظر باشد، زیرا در صورت طراحی دقیق این آزمونها، پایایی قابل قبول است؛
- ۴- مهارت آزمون گیرنده در زمینه ساخت آزمون چند گزینه‌ای بیشتر از مهارت در تصحیح آزمونهای تشریحی باشد.

برای اندازه‌گیری کلیه سطوح یادگیری می‌توانید از آزمون چندگزینه‌ای استفاده کنید. البته سؤالات چند گزینه‌ای عادی برای سطوح پایین‌تر یادگیری به کار می‌رود و برای اندازه‌گیری سطوح بالاتر یادگیری به وسیله این سؤالات بهتر است از نمودار، عکس، نتایج آزمایشها، متن و غیره - که به مجموع آنها مطالب مقدماتی می‌گویند- استفاده کنید. این نوع آزمون به سنجش حیطه دانشی مطالب پرداخته، قادر به آزمودن جنبه‌های عملی آن نیست.

شاید گفته شود در این آزمون چون پاسخها در اختیار دانشجو است، جنبه سطحی معلومات سنجیده شده، به دانشجو امکان قضاوت شخصی داده نمی‌شود؛ در حالی که انتخاب پاسخ، عملی است که در زندگی، احتمال وقوع آن به اندازه احتمال خلق پاسخهاست. با آزمون چندگزینه‌ای نمی‌توانید مهارت در تصمیم‌گیری، توانایی برقراری ارتباط با دیگران و خلاقیت دانشجو را اندازه‌گیری کنید. مثلاً نمی‌توانید تشخیص دهید آیا دانشجو می‌تواند در یک موقعیت بالینی تصمیم قطعی گرفته، عمل کند و یا توانایی ارتباط صحیح با بیمار خود را دارد یا خیر. پس اگر چنین هدفی را دنبال می‌کنید این نوع سؤال تنها برای در برگرفتن قسمت کوچکی از دوره آموزشی دارای روایی بوده، برای سنجش سایر اهداف آموزشی لازم است از آزمونهایی اندازه‌گیری‌کننده مهارت استفاده نمود.

سؤالات صحیح - غلط^۱

به نوعی آزمون چند گزینه‌ای که دانشجو در آن باید از بین دو جواب بله و خیر یکی را انتخاب کند، آزمون با سؤالات صحیح - غلط می‌گویند.

مثال: شایعترین علت سندرم اجوف فوقانی^۱، برونکوژنیک کارسینوما است. ص □ غ □

یک جمله تنها در صورتی صحیح است که برای همه موارد درست باشد و حتی اگر در یک مورد هم نادرست باشد غلط محسوب می‌گردد.

مثال: شبکه مویرگی در بدن انسان بلافاصله به ورید ختم می‌شود.
پاسخ صحیح برای این سؤال «غ» می‌باشد. البته این موضوع در اکثر قسمت‌های بدن صدق می‌کند ولی استثناً هم دارد (گلو مریول کلیه).

انواع سؤالات ص - غ

۱- یک عبارت خبری که دانشجو باید درست یا نادرست بودن آن را تشخیص دهد.
مثال: در میوکاردیت نوع دیفتریک، توکسین باکتری باعث اختلال هدایت در سیستم هدایتی قلب می‌شود.

۲- دانشجو پس از تشخیص درستی و نادرستی جمله، صورت صحیح عبارت نادرست را بنویسد.

مثال: منیسک سمت مدیال زانو متحرک‌تر از منیسک لترال است. (غ)

جمله صحیح: منیسک سمت لترال متحرک‌تر از سمت مدیال است.

۳- نوع صحیح - غلط چند گزینه‌ای^۱ یا شبکه‌ای که شامل چندین سؤال ص - غ با ریشه یکسان است و در همین فصل به تفصیل مورد بحث قرار گیرد.

طراحی سؤالات ص - غ

در طرح سؤالات ص - غ باید نکاتی مدنظر قرار گیرند که از آن جمله است:

۱- در هر سؤال تنها یک موضوع واحد مرتبط با هدفهای اصلی را عنوان کنید:

مثال: چربیها در بدن و مواد غذایی بیشتر به صورت تری‌گلیسیرید هستند که از سه

مولکول گلیسرول و یک مولکول اسید چرب تشکیل شده است. (غ)

دانشجو در این سؤال برای یافتن پاسخ صحیح باید دو مطلب را بررسی کند. در حقیقت

این جمله دو سؤال بوده است که باید هر یک جداگانه مطرح گردد.

مثال: وزن مولکولی ایمونوگلوبولین $40 \cdot 10^4$ kd, G (کیلو دالتون) می‌باشد. (ص)

اگر یک پزشک در انجام وظایف آینده‌اش به دانستن وزن مولکولی ایمونوگلوبولین G

نیازی نداشته باشد طرح این پرسش برای دانشجوی پزشکی مناسب نخواهد بود.

۲- جمله موردنظر خود را طوری طرح کنید که کاملاً صحیح یا غلط بوده، ابهام نداشته،

دانشجوی مطلع را به اشتباه نیندازد.

مثال: سندرم کلیپل فایل^۱ با کم و زیاد شدن تعداد مهره‌ها ارتباط دارد.

این جمله به دو صورت تفسیر می‌شود: اگر منظور از «کم و زیاد شدن» تغییر تعداد مهره‌ها باشد عبارت مورد نظر صحیح و اگر مطلق «کم شدن و یا زیاد شدن» مدنظر باشد جمله غلط است؛ چون در این سندرم تعداد مهره‌ها کم می‌شود.

مثال: پریکاردیت فشارنده^۲ را می‌توانیم به صورت التهاب مزمن پرده پریکارد تعریف کنیم. پریکاردیت فشارنده یعنی التهاب مزمن پریکارد که با افزایش ضخامت و حتی کلسیفیکاسیون همراه است و باعث نارسایی قلب راست می‌شود. جمله مثال فقط پریکاردیت را تعریف می‌کند، پس کامل نبوده، غلط است ولی با توجه به کلمه «می‌توانیم» اگر دانشجو جواب «ص» را انتخاب کند، نمی‌توان به او ایراد گرفت.

۳- واژه‌های عام و کلی را به کار نبرید. واژه‌هایی مثل «معمولاً»، «اغلب»، «بعضی اوقات» و ... عمدتاً دلالت بر جمله صحیح دارد. بنابراین دانشجو بدون داشتن اطلاعات کافی پاسخ «ص» را انتخاب می‌کند.

مثال: سرطان ممکن است در سنین پیری ایجاد شود. (ص)

با به کار بردن «ممکن است» در این جمله، دانشجوی غیر مطلع هم به راحتی جواب «ص» را انتخاب می‌کند.

۴- واژه‌هایی مثل هرگز، همیشه، اصلاً، فقط، هیچکدام و ... که اغلب بر جمله نادرست دلالت دارند را به کار نبرید.

مثال: در تست تحمل گلوکز قند پایین‌تر از 140 mg/l همیشه دلیل بر عدم وجود دیابت است. (غ)
۵- جمله‌ای که انتخاب می‌کنید در اصل صحیح یا غلط بوده، با به کار بردن توضیحات اضافی یا قیود، آن را غلط نکنید.

مثال: اگر هدف شما این است که بیان قانون فرانک استارلینگ را مورد سؤال قرار دهید و سؤال زیر را طرح کنید:

«طبق قانون هنری، مقدار خون ورودی به قلب، میزان خون خروجی را تعیین می‌کند.» (غ)
دانشجو اگر کوچکترین اطلاعی راجع به قانون هنری داشته باشد بدون رسیدن به هدف مورد نظر شما، به جواب درست می‌رسد. یعنی دانشجو ممکن است قانون فرانک استارلینگ را نداند ولی بتواند به این سؤال پاسخ دهد و نمره بگیرد، در نتیجه شما به هدف خود نخواهید رسید.

۶- با به کار بردن کلمات منفی مانند «نباید و نه» جمله کاملاً صحیح را به جمله غلط تبدیل نکنید.

مثال: غده هیپوفیز بخشهای قدامی و خلفی ندارد. (غ)

در این سؤال می‌توان بخشهای غده هیپوفیز را اشتباه ذکر کرد تا جمله مفهوم غلط پیدا کند، اما منفی کردن فعل برای نادرست کردن یک جمله، صحیح نیست.

۷- از به کار بردن عبارتهای منفی و منفی در منفی تا حد ممکن خودداری کنید و در صورت استفاده از کلمه منفی، آن را مشخص نمایید. مثلاً زیر آن خط بکشید.

مثال: در حالت استراحت، دریچه‌های غیر فعال شدن سدیم باز نیست. (غ)

مثال بالا را به این صورت می‌توانید اصلاح کنید:

در حالت استراحت، دریچه‌های غیر فعال شدن سدیم بسته است. (غ)

۸- عین جمله کتاب را به عنوان سؤال صحیح - غلط به کار نبرید.

۹- تا آنجا که می‌توانید پرسشها را قابل فهم و کوتاه طرح کنید، زیرا در این صورت مطمئن می‌شوید که پاسخهای دانشجویان براساس دانش آنهاست نه توانایی در خواندن سؤالات.

مثال: اگر آنتی‌بیوتیکی بتواند عامل سیگمای مولکول «RNA پلی‌مراز» یک باکتری را غیر فعال کند، این آنتی‌بیوتیک با جلوگیری از شروع سنتز RNA باکتری را می‌کشد. (ص)

این سؤال بیش از اندازه طولانی شده است و اشکال دارد.

۱۰- در مورد مفاهیمی که هنوز بین صاحبان نظر اتفاق نظر وجود ندارد، حتماً ماخذ پاسخ دادن به سؤال را ذکر کنید.

۱۱- از مطالب مقدماتی شامل جدول، نقشه، شکل و ... برای اندازه‌گیری سطوح پیچیده‌تر یادگیری استفاده کنید.

۱۲- برای سنجش رابطه علت و معلول، اصل رابطه را صحیح و غلط کنید نه خود جملات را؛ به صورتی که دو جمله علت و معلول هر دو صحیح بوده، دانشجو در مورد صحیح یا غلط بودن ارتباط بین آن دو تصمیم بگیرد.

مثال: در جریان پتانسیل عمل، کانالهای سدیمی باز می‌شوند؛ چون سدیم در خارج سلول بیشتر از داخل است. (غ)

در این سؤال هر دو جمله صحیح بوده ولی بیشتر بودن سدیم خارج سلولی دلیل بر باز شدن کانال سدیم در جریان پتانسیل عمل نیست، پس این جمله غلط است.

مثال: سلولهای گره پیشاهنگ قلب، انقباض خود به خود دارند، زیرا دارای پتانسیل

استراحت است. (غ)

در این سؤال، هر دو جمله صحیح است، ولی پتانسیل استراحت سلولهای گره پیشاهنگ قلب دلیل انقباض خودبخود آنها نیست؛ پس این جمله غلط است.
۱۳- طول جملات درست را تا حد ممکن برابر با جملات نادرست طرح کنید؛ زیرا تمایل به طولانی نوشتن جملات غلط بیشتر است.

۱۴- پاسخهای ص - غ را به صورت تصادفی انتخاب کرده، سعی کنید جوابهای «غ» کمی بیش از «ص» باشد زیرا دانشجوی مردد بیشتر پاسخ «ص» را انتخاب می‌کند.

۱۵- برای اینکه سؤال روا باشد به جای بیان یک حقیقت^۱، از یک نظریه^۲ استفاده کنید و در صورت امکان سؤال را به صورتی طرح کنید که برای حل آن نیاز به فهم نظریه باشد؛ زیرا سؤال از حقیقت بیشتر جنبهٔ حفظی دارد، اما نظریه، سطوح بالاتری از معلومات دانشجوی را می‌سنجد و نیاز به تجزیه و تحلیل علمی دارد.

مثال: حقیقت: یون اصلی در ایجاد پتانسیل عمل، سدیم است.

نظریه: تغییرات پتانسیل عمل بیشتر به دریچه‌های ولتاژی سدیم مربوط است.

سؤال: برای متوقف کردن مسیر پیام عصبی باید دریچه‌های ولتاژی سدیم را بلوک کرد.

۱۶- بهتر است از دانشجویان بخواهید با حروف «ص» یا «غ» پاسخ سؤال را مشخص کنند، یا روی حروف «T» و «F» که خودتان نوشته‌اید علامت بزنند. نه اینکه آنها را بنویسند چون حروف «T» و «F» شبیه هم بوده، ممکن است در دست‌خطهای ناخوانا اشکال ایجاد کند.

۱۷- از دانشجو راجع به قضاوت شخصی او سؤال نکنید. اگر در صورت سؤال عبارت «به نظر شما» بیاید هر پاسخی که دانشجو بدهد درست خواهد بود، زیرا شما فقط نظر او را خواسته‌اید و کاری به درست بودن یا نبودن آن ندارید.

۱۸- کلمات زیر را در سؤالات ص - غ فقط در صورتی به کار برید که همان معنی مقابل آن را داشته باشد:

همه: همهٔ اعضای گروه، صد درصد، بدون استثنا

همیشه: در همه حال، در صد درصد مواقع رخ می‌دهد، امکان ندارد که رخ ندهد.

و: به علاوه، یک جملهٔ مرکب که بوسیله رابط «و» به هم متصل می‌شوند.

- فقط و فقط در صورتی صحیح است که همه جملات صحیح باشد.
- مثال: در مهندسی ژنتیک پلاسمیدها و فاژها برای حمل قطعات کوچک DNA به کار می‌روند.
- در اینجا جمله برای پلاسمید صحیح ولی برای فاژ غلط است، بنابراین کل جمله غلط می‌باشد.
- زیرا: پدیده‌ای که علت دیگری است و در نبود آن معلول ایجاد نمی‌شود.
- هر: هرچیز ممکن، بدون استثنا
- کم: بیش از یکی ولی نه خیلی (بین دو تا ده)
- غیرممکن: هرگز اتفاق نمی‌افتد، احتمال وقوع صفر است.
- قطعی: حتماً اتفاق می‌افتد، احتمال وقوع یک است.
- غیر وابسته: تحت تأثیر قرار نمی‌گیرد. در صورتی دو واقعه غیر وابسته هستند که وقوع یکی به دیگری ارتباط نداشته باشد.
- خیلی: تعداد زیاد، بسته به موقعیت.
- بیشتر: از نظر تعداد یا مقدار زیادتر باشد.
- بیشترین: از نظر تعداد یا مقدار در گروه، زیادترین است.
- هیچوقه: هرگز اتفاق نمی‌افتد.
- هیچکدام: صفر.
- فقط: بی‌همتا، واحد.
- فقط اگر: نمی‌تواند اتفاق بیفتد مگر اینکه.
- ممکن: می‌تواند اتفاق بیفتد، احتمال وقوع بیش از صفر است.
- برخی: حداقل یکی.
- یا: به عنوان جایگزین. یک جمله مرکب که بوسیله رابط «یا» به هم متصل می‌شود در صورتی صحیح است که یکی از جملات صحیح باشد و در صورتی که همه غلط باشد جمله نادرست است.
- مثال: در مهندسی ژنتیک پلاسمیدها یا فاژها برای حمل قطعات کوچک DNA به کار می‌روند. این مثال صحیح است؛ زیرا این جمله برای پلاسمیدها صحیح می‌باشد و چون رابط «یا» است کل جمله صحیح می‌باشد.

مزایا

پرسشهای ص - غ دارای محاسنی است که از آن جمله به موارد زیر می‌توان اشاره نمود:

- ۱- چون می‌توانید تعداد زیادی سؤال را در زمان کوتاه مورد آزمون قرار دهید، حجم زیادی از محتوای تدریس را در بر می‌گیرد.
- ۲- اگر بتوانید آزمون چندگزینه‌ای را به شکل ص - غ در آورید، چون هر گزینه به صورت یک سؤال جداگانه نمره دارد از نظر نمره‌گذاری دارای برتری است، ولی این کار احتمال حدس زدن پاسخ صحیح توسط دانشجو را دو چندان می‌کند. پس با توجه به ضرورت، نوع پرسش را انتخاب کنید.

معایب

- ۱- مهمترین عیب این نوع سؤال این است که احتمال حدس زدن پاسخ صحیح در آن بسیار بالاست (حدود ۵۰ درصد).
برای حل این مشکل می‌توانید از مقیاس زیر استفاده کنید:

صحيح	احتمالاً صحيح	نمی‌دانم	احتمالاً غلط	غلط	
+۲	+۱	۰	-۱	-۲	برای گزینه‌های صحيح:
-۲	-۱	۰	+۱	+۲	برای گزینه‌های غلط:

- ۲- تهیه آن مشکل است؛ چون جملات باید کاملاً درست یا نادرست باشد. اغلب با به کار بردن قیود و توضیحات اضافی جمله‌ها درست یا نادرست می‌شود و این قیود و توضیحات دانشجو را به همت جواب راهنمایی می‌کند.
- ۳- وقتی دانشجو جمله غلط را انتخاب می‌کند، نمی‌توانید استنباط کنید که درست آن را می‌داند؛ البته برای حل این مشکل می‌توانید از وی بخواهید صحیح جمله را در مقابلش بنویسد.
- ۴- وقتی دانشجو سؤال را اشتباه پاسخ می‌دهد، نمی‌توانید افکار اشتباه او و منشأ آن را مشخص کنید، ولی در آزمون چندگزینه‌ای از روی انتخاب گزینه‌های غلط به نوع اشتباه دانشجو پی می‌برید.

۵- معمولاً این نوع سؤالات را برای سطوح بالای یادگیری نمی‌توانید به کار برید؛ بخصوص وقتی سؤالات با دقت طرح نشود به سنجیدن مطالب حفظی و سطحی محدود شده، اندازه‌گیری یادگیری عمیق را مشکل می‌کند. طرح پرسش ص - غ به معنای انتخاب یک جمله از بین جملات کتاب درسی نیست زیرا در این صورت دانشجو فقط مطالب و جملات را حفظ کرده، به دنبال فهم آن نخواهد بود.

سؤالات صحیح - غلط چندگزینه‌ای

سؤالات صحیح - غلط چند گزینه‌ای (شبکه‌ای) نوعی آزمون چند گزینه‌ای است با یک ریشه و چند گزینه که دانشجو باید درستی یا نادرستی هر یک از گزینه‌ها را مشخص کند.

این سؤالات رایجترین نوع سؤالات چند گزینه‌ای است که به آنها سؤالات تشخیصی پاسخ نیز گفته می‌شود. این نوع سؤالات را زمانی به کار برید که یک سؤال چند گزینه‌ای با چند پاسخ صحیح داشته باشید یا طراحی جوابهای انحرافی مناسب مقدور نباشد.

در پزشکی معمولاً به دانشجو اجازه می‌دهند جلوی برخی از انتخابها جواب «نمی‌دانم» بنویسد. این حالت بیشتر در مواقعی موردنظر است که اطمینان داشتن پزشک و قاطعیت او در هنگام تصمیم‌گیری برای ما بسیار حائز اهمیت باشد. در این حالت «نمی‌دانم» نمره صفر داشته در حالی که پاسخ نادرست یک نمره منفی دارد.

مثال: کدام یک از علائم زیر مشخصه استئوپروز هستند؟ برای هر پاسخ «ص» یا «غ» را انتخاب کنید.

الف)	اجسام مهره‌ای کلاپس شده	ص	غ	نمی‌دانم
ب)	نواحی Looser برجسته	ص	غ	نمی‌دانم
ج)	فقدان تراکولاهای استخوان	ص	غ	نمی‌دانم
د)	کاهش غلظت Ca سرم	ص	غ	نمی‌دانم
ه)	ضعف ماهیچه‌ای Proximal	ص	غ	نمی‌دانم

برای طرح این نوع سؤال یک موضوع اصلی و براساس آن چند مفهوم^۱ و برای هر مفهوم چند زیر شاخه، به صورت زیر وجود دارد:

موضوع اصلی:		بیماریهای متابولیک استخوان		
بیماری پاژه	استئومالاسی	استئوپروز	مفهوم:	
علایم بالینی	علایم بالینی	علایم بالینی	زیر شاخه:	
علت	علت	علت		
تشخیص	تشخیص	تشخیص		
پیکیری	پیکیری	پیکیری		

اعضای NBME^۱ در سال ۱۹۹۱ تصور می‌کردند که این نوع سؤالات فقط در حد یادآوری است و اعلام کردند بهتر است به جای آن از شکل معمول چندگزینه‌ای استفاده کنند، در حالی که کاربرد اصلی این پرسشها در زمینه «مهارت در موضوع»^۲ است. برای اینکه دانشجویی در موضوعی مهارت داشته باشد اول آنکه باید خصوصیات و شاخصهای آن موضوع را بداند، دوم آنکه بین خصوصیات این موضوع با موضوعهای مشابه تمیز قابل شود. به عنوان نمونه در مثال قبل باید استئوپروز و استئومالاسی را تفکیک کند. در این نوع سؤالات، جملات درست بیان‌کننده شاخصهای موضوع موردنظر و جملات نادرست نشان‌دهنده خصوصیات موضوعات مشابه است.

برای طرح این پرسشها از نوع چندگزینه‌ای استفاده می‌شود، به این صورت که همان ریشه را نوشته، برای طرح پنج گزینه باید موقعیتهای مشابه را به دو موضوع مختلف تعمیم داد. به این طریق، سؤالات از حد یادآوری به سطح درک ارتقا می‌یابد. هتلاً در طرح سؤال قبل گزینه‌های نادرست، خصوصیات استئومالاسی است، زیرا می‌خواهیم بدانیم آیا دانشجوی توانایی تفکیک بین این دو موضوع مشابه را دارد یا خیر.

اگر این نوع سؤالات را جایگزین پرسشهای جورکردنی کنیم ارزش خود را از دست می‌دهد. یعنی اگر به جای طرح سؤال صحیح - غلط چند گزینه‌ای که در آن از استئومالاسی سؤال گردید و ویژگیهای استئوپروز هم در میان گزینه‌ها آورده شد از دانشجو خواسته شود خصوصیات استئومالاسی و استئوپروز را در یک ستون با ستون مقابل جور کند ممکن است فرد بدون هیچ اطلاعی از استئوپروز که منظور سؤال است و فقط با اطلاع از

۱- National Board of Medical Examinations: هیاتی است که امتحانات ورود به مرحله تخصصی را به صورت ملی (National) در سطح ایالات متحده برگزار می‌کند.

استئومالاسی به جای تمیز بین پاسخها، نوعی حذف انتخابی انجام دهد. از آنجا که در این نوع سؤالات، تسلط به موضوع اهمیت دارد، به طور معمول دانستن تمام جوانب موضوع معیار تلقی می‌گردد، بنابراین از نمره‌گذاری ملاک مرجع استفاده می‌گردید.

سؤالات جورکردنی ۱

سؤال جورکردنی نوع تغییر یافته‌ای از سؤالات چندگزینه‌ای است که در آن تعدادی ریشه و تعدادی پاسخ، هر یک در ستونی مشخص قرار گرفته، دانشجو باید برای هر ریشه، از ستون دوم یک پاسخ را انتخاب کند. هر پاسخ ممکن است یک یا چندین بار انتخاب شود یا اینکه اصلاً انتخاب نشود.

مثال: در ستون (الف) نام هورمون و در ستون (ب) مکانیسم عمل مشخص شده است. برای هر یک از بخشهای گروه (الف)، مورد مناسب را از گروه (ب) انتخاب کنید. حرف پاسخ انتخابی خود را در مقابل ریشه مورد نظر بنویسید. هر یک از پاسخها را می‌توانید برای یک بار یا چندین بار انتخاب کرده یا اصلاً انتخاب نکنید.

<u>الف</u>	<u>ب</u>
۱- هورمون آنتی دیورتیک (ADH)	a . cGMP
۲- هورمون رشد	b . پروتیین کیناز
۳- کورتیزول	c . Ca و کالمودولین
۴- انسولین	d . رسپتور هسته‌ای
۵- تیروکسین	e . رسپتور سیتوپلاسمی
	f . cAMP

طراحی سؤالات جورکردنی

در طرح این سؤالات به نکات زیر توجه کنید:

۱- تنها از مطالب متجانس و مربوط به هم استفاده نمایید به طوری که همه پاسخها برای همه پرسشها قابل قبول باشد.

۲- سؤالاها و بخصوص پاسخها را تا حد ممکن مختصر کنید تا زمان خواندن آن کوتاهتر شود؛ زیرا دانشجو هر سؤال را فقط یک بار می‌خواند ولی پاسخها چندین بار خوانده

می‌شود.

۳- تعداد پاسخها را بیشتر از تعداد پرسشها قرار دهید. در ضمن پاسخها را طوری انتخاب کنید که بعضی از آنها بیش از یک بار تکرار شود. با این عمل احتمال حدس زدن پاسخ صحیح را کاهش می‌دهید.

۴- در دستورالعمل، ارتباط بین دو ستون و نحوه پاسخ‌گویی دانشجو را بیان کرده، دقیقاً مشخص کنید که هر جواب ممکن است چندین مرتبه انتخاب شود.

۵- پاسخها را به صورت منظم و براساس حروف الفبا، ترتیب زمانی یا ... مرتب کنید.

۶- تعداد سؤالات بین پنج تا دوازده پرسش مناسب است. اگر تعداد پرسشها کمتر باشد احتمال شانس افزایش می‌یابد و اگر بیشتر باشد وقت‌گیر و خسته‌کننده خواهد بود.

۷- دقت کنید قرینه‌های لفظی و نکات ظریفی را که موجب آشکار شدن پاسخ می‌شود به کار نبرید. بررسی کنید که همه پاسخها برای هر پرسش قابل قبول باشد.

۸- همه پرسشها و پاسخها را در یک صفحه قرار دهید. این عمل موجب دقت عمل و سرعت خواندن دانشجو می‌شود.

مثال: نمونه‌ای از یک سؤال جور کردنی با اشکالات زیاد:

به پرسش زیر پاسخ دهید:

- ۱- دستگاهی که مسئول تنظیم وزن در طی هفته‌ها و ماهها و سالهاست.
- ۲- چربی ۲۰ درصد بالاتر از نرمال است. افزایش وزن a)
- ۳- کنترل درازمدت وزن. تری گلیسیریدها b)
- ۴- وزنی که بین نرمال و چاقی است. اشتها c)
- ۵- سیستمی که در ارتباط با گرسنگی و سیری است. ۲۵۰۰ کالری d)
- ۶- تعداد کالری در یک پوند. دستگاه متحرک e)
- ۷- سلولهای مخصوصی که چربی ذخیره می‌کنند. متابولیک f)
- ۸- تعداد کالری که یک شخص عادی در روز مصرف می‌کند. چاقی g)
- چین پوستی h)
- هیپوتالاموس i)
- بافت چربی j)
- ۲۲۰۰ کالری k)
- دستگاه وزنی l)

این پرسش اشکالات زیادی دارد از جمله:

- ۱- دستورالعمل پاسخ‌دهی بیان نشده است.
 - ۲- ترتیب منطقی در پاسخها وجود ندارد.
 - ۳- در بسیاری از سؤالات، اشاره مستقیم وجود دارد از جمله: در سؤالات ۱ و ۵ راجع به دستگاه^۱ سؤال شده که فقط گزینه‌های ۱ و e با آن مرتبط است. همچنین در سؤالات ۶ و ۸ در مورد میزان کالری سؤال شده است. بنابراین تنها جوابهای d و k می‌تواند انتخاب شود. شماره ۴ در مورد وزن سؤال کرده است. بنابراین تنها گزینه‌های g و a به آن مربوط است. در این موارد بهتر است از سؤالات ص - غ استفاده شود.
- مثال: برای هر یک از پرسشهای ستون «الف»، گزینه‌ای از ستون «ب» را انتخاب کنید.

<u>الف</u>	<u>ب</u>
۱- عضله تراپیوس	۱- C3-C4-C5
۲- استرنوکلیدوماستوئید	۲- شریان اپی‌گاستریک فوقانی و تحتانی
۳- رکتوس	۳- عصب اکسسوری
۴- عضله ساب‌کلاویوس	۴- شریان ساب‌کلاوین

بعضی از اشکالات متعدد مثال فوق عبارتند از:

- ۱- مطالب متجانس نیستند چون در همه پاسخها باید یا شریان یا عصب باشد.
- ۲- تعداد پرسشها و پاسخها مساوی است.
- ۳- از قرینه‌های لفظی استفاده شده است؛ مثلاً در مورد عضله ساب‌کلاویوس و شریان ساب‌کلاوین.
- ۴- تعداد سؤالات کم است؛ چون تعداد مطلوب آنها بین پنج تا دوازده عدد می‌باشد.
- ۵- دستورالعمل کامل نیست؛ زیرا نحوه پاسخ‌گویی را بیان نکرده، در ضمن مشخص نیست آیا دانشجو می‌تواند یک جواب را چندین بار انتخاب کند یا خیر.

مزایا

این سؤالات محاسن و معایبی دارد که در هنگام استفاده باید به آنها توجه نمود:

- ۱- آنها را می‌توانید در اندازه‌گیری درک تعاریف و ارتباط بین دو موضوع به کار برید.

۹۰ آزمونهای پیشرفت تحصیلی در علوم پزشکی

۲- با استفاده از پرسشهای جور کردنی مطالب زیادی را می‌توانید در یک سؤال واحد بگنجانید. بنابراین مختصر و با صرفه بوده، بخش وسیعی از مفاهیم درسی را در بر می‌گیرد.

۳- این سؤالات نیازی به یافتن گزینه‌های انحرافی ندارد.

۴- چون همهٔ پاسخها باید گزینه‌های قابل قبول برای هر یک از پرسشها باشد، یافتن و طرح چنین سؤالی معمولاً دشوار است.

۵- این سؤالات اغلب مطالب مهم را بیان نکرده، تعداد محدودی از هدفها را در بر می‌گیرد.

۶- اگر این سؤالات به نحو صحیح طرح نشود، پرسشها و پاسخها ناهمگون شده، در نتیجه برای هر پرسش تنها یک یا دو پاسخ قابل قبول وجود خواهد داشت که شانس دانشجو را برای حدس زدن پاسخ صحیح زیاد می‌کند.

سؤالات کوتاه پاسخ^۱

فرآیند تدریس در دهه‌های اخیر به طور کلی تغییر کرده، تحولی از تدریس صرف، استاد محور^۲ و یکطرفه به سمت تدریس با مشارکت دانشجو^۳ به وجود آمده است و مبنای تدریس، بر خودآموزی قرار گرفته است. دانشجویان تشویق شده‌اند که به شیوهٔ گروهی یا به تنهایی کار کرده، دیدگاههای خود را برای حل مسألهٔ طرح شده ارائه کنند.

آنچه در کسب دانش با ارزش است، تفسیر و کاربرد دانسته‌هاست. بنابراین، شیوه‌های آزمون باید چنین قابلیت و توانایی را اندازه بگیرد، نه اینکه فقط دانش فرد را بسنجد.

اگر چه پرسشهای چندگزینه‌ای را می‌توان برای سنجش قدرت حل مسأله به کار برد؛ سؤالات کوتاه پاسخ بهترین راه حل برای روبرو کردن دانشجو با موقعیت واقعی حل مسأله در زندگی است و اصلاً این سؤالات به همین منظور ابداع شده است.

پرسشهای کوتاه پاسخ تنها نوع سؤال باز پاسخ عینی بوده که دانشجو، خودش جواب را تهیه می‌کند. این پرسشها شبیه سؤالات چندگزینه‌ای است اما گزینه ندارد.

سؤالات کوتاه پاسخ را معمولاً به جای سؤالات تشریحی بکار می‌برند. به این ترتیب که یک سؤال تشریحی را با چند سؤال کوتاه پاسخ جایگزین می‌کنند تا میزان درک دانشجو را از جنبه‌های مختلف مشکل طرح شده به طور جداگانه، کشف کنند.

هر چه سؤالات کوتاهتر باشد، پایایی آزمون بیشتر و نمره‌دهی دقیقتر خواهد بود، چون پاسخها کوتاه است.

مثال:

۱- ناقل بیماری مالاریا چیست؟

۲- محل زندگی آن کجاست؟

۳- روش مبارزه با آن چگونه است؟

در سؤال اول ممکن است دانشجو نام حشره را اشتباه گفته باشد، اما کافی است اطلاعات کلی راجع به روشهای کنترل ناقل داشته باشد. تا بتواند بخشی از سؤال آخر را جواب دهد. در صورتی که اگر دانشجو دانسته‌های متعددی راجع به یک موضوع داشته، تنها قسمت کوچکی از اطلاعات مورد نیاز برای پاسخ به پرسش چند گزینه‌ای را نداند مثل این است که هیچ چیز نمی‌داند و نمی‌تواند گزینه صحیح را پیدا کند. اما با طرح سؤالات کوتاه پاسخ که دانشجو خود، جواب را تهیه می‌کند او حداقل می‌تواند اطلاعات خود را نشان داده، به جواب سؤال نزدیک شود.

معمولاً این نوع سؤال وقتی به کار می‌رود که:

۱- هدف، تهیه پاسخ توسط خود دانشجو باشد نه انتخاب پاسخ از میان تعدادی گزینه؛

۲- طراح سؤال در تهیه گزینه برای آزمون چندگزینه‌ای مشکل داشته باشد.

این پرسشها شکل و محتوای جواب را محدود کرده، به دانشجو توان درک منظور استاد را می‌دهد.

سؤالات کوتاه پاسخ را می‌توانید برای سنجش سطوح عمیق‌تر یادگیری نیز به کار برید.

سطح یادآوری: محل شروع^۱ و محل اتصال^۲ عضله‌ای خاص را در اندام بیان کنید؟

سطح درک: حین pushoff کدام عمل آناتومیک انجام می‌شود؟

سطح کاربرد: در فلج عصب کامون پرونیال مشترک^۳ کدام مرحله از راه رفتن مختل می‌شود؟

کاربرد سؤالات کوتاه پاسخ

از این سؤالات برای تفسیر تصاویر، حل مسأله و نیز خواندن نمودارها و عکسهای رادیوگرافی استفاده می‌شود. کاربرد اصلی این سؤالات وقتی است که لازم باشد فرد مورد

1- Insertion

2- Origin

3- Common Peroneal

آزمون با موارد عینی‌تر روبرو شود نه اینکه صرفاً برای گذراندن امتحانات مقداری اطلاعات را حفظ کند.

مثال: با توجه به ارتباط موارد نامبرده، شماره ۳ را نام ببرید.

انتروبیوس ورمیکولاریس^۱ - آسکاریس لومبریکوئیدس^۲ -

این دو کرم در گروه نماتودها قرار دارند و پاسخ دانشجو نیز باید از این دسته باشد.

نحوه ساخت

۱- نکات مهم درس را مورد سؤال قرار دهید (اطمینان یابید پاسخ سؤال شما یک هدف مهم است).

۲- سؤال را به صورتی طرح کنید که پاسخ به آن مستلزم کاربرد کمترین کلمات باشد (برای اطمینان از این موضوع خودتان یا فرد دیگری قبل از برگزاری آزمون به سؤال پاسخ دهید).

۳- صورت سؤال را کوتاه و رسا و با زبانی ساده طرح کنید.

۴- سعی کنید بیشتر از سؤالات پرسشی استفاده کنید تا کامل کردنی، زیرا این عمل باعث بیشتر آشکار شدن منظور شما می‌شود و احتمال یک پاسخ داشتن پرسش شما افزایش می‌یابد.

۵- جمله‌های انتخاب شده، مستقیماً از متن کتاب نباشد.

۶- مطمئن باشید سؤال فقط یک پاسخ داشته، جمله‌بندی سؤال پاسخ را القا نمی‌کند. برای اطمینان از این موضوع، پیش از برگزاری آزمون آن را به چند نفر دهید، تا مطمئن شوید همه یکسان پاسخ می‌دهند.

مثال: در هموتوراکس با قرار دادن لوله چه چیزی را از ریه خارج می‌کنند؟

با توجه به کلمه «هم» دانشجو مستقیم به جواب «خون» هدایت می‌شود.

مثال دیگر: در پاسخ به سؤال «توسط الکتروکاردیوگرام کدام ضایعه را می‌توان تشخیص داد؟» پاسخهای متفاوتی داده می‌شود. از جمله «هیپرتروفی قلب، آریتمی، جریان ضایعه، انفارکتوس میوکارد و ...».

۷- در پرسشهای کامل کردنی بیش از یک جای خالی آن هم در آخر جمله به کار نبرید؛ چون دانشجو مجبور به تجزیه و تحلیل بیشتر می‌شود، بدون آنکه سطح یادگیری مورد اندازه‌گیری را بالا ببرد.

۸- بهتر است همراه سؤالات کوتاه پاسخ، سؤالات چندگزینه‌ای را هم به کار ببرید.

- ۹- در مورد پاسخهای عددی، واحد مقیاس مورد استفاده و میزان دقتی را که باید در محاسبات رعایت شود، مشخص کنید. مثلاً در مورد یک تاریخ، دقت محاسبه (سال، روز، ماه) را مشخص کنید.
- ۱۰- تصحیح سؤال را به عهده افراد غیر مطلع نگذارید زیرا اگر دانشجو کلمه‌ای مترادف بکار برد حق او ضایع می‌شود.
- ۱۱- در پرسشهای کامل کردنی، کلمات مهم و با اهمیت را به عنوان جواب قرار دهید.
- ۱۲- طول جای خالی را در تمام سؤالات به یک اندازه انتخاب کنید نه براساس طول پاسخ.
- ۱۳- قرینه لفظی ندهید. مثال: استفاده از a یا an قبل از جای خالی یا کم یا زیاد بودن طول جای خالی متناسب با طول پاسخ.
- ۱۴- در سؤالات کامل کردنی، اطلاعات کافی را جهت پرکردن جای خالی در متن سؤال بگنجانید، به طوری که دانشجو را بدون فریب دادن به سمت جواب صحیح هدایت کند. مثال: داروی انتخابی درمان عفونت استرپتوکوکی ... است.
- ۱۵- اگر در این سؤالات از تصاویر استفاده کنید کار ارزشمندی برای بهبود فرایند ارزیابی انجام داده‌اید.

مزایا

- ۱- در سؤالات کوتاه پاسخ، احتمال حدس زدن وجود ندارد.
- ۲- هیچ‌گونه راهنمایی جهت هدایت به سمت پاسخ به دانشجو ارائه نمی‌شود.
- ۳- تهیه و طرح سؤالات نسبتاً آسان است.
- ۴- پاسخ به سؤالات نیاز به مهارتهای ارتباطی پیچیده ندارد.
- ۵- آزمون کامل کردنی برای ارزیابی یادآوری واقعیاتی مثل تاریخ، اسم داروها، دستور مصرف آنها و اطلاعات کلی دیگر مفید است.

معایب

- ۱- آزمون کوتاه پاسخ دانشجو را در تفسیر موضوع محدود می‌کند؛ به همین دلیل برای جلوگیری از سرگشتگی دانشجو باید در انتخاب کلمات دقت نمود.
- ۲- در این سؤالات، اولین پرسش، موضوع اصلی را مورد سؤال قرار می‌دهد بنابراین، سؤالات به صورت قدم به قدم طراحی می‌گردد، بدین معنی که سؤال اول تنها یک جواب

دارد. اما همین طور که به سؤالات بعد می‌رسید امکان اظهار نظر بیشتری وجود دارد و لازم است سؤالات پیچیده‌تر و جوابهای نسبتاً صحیحی ارائه کنید، بدون آنکه جواب سؤال اول را اشتباه بدهید.

۳- چون نمره‌دهی این سؤالات به راحتی از نوع عینی به نوع ذهنی در می‌آید، بهتر است سؤال تحلیلی بکار نبرده، از سؤال مستقیم^۱ استفاده کنید.

۴- املائی کلماتی که در پاسخ به کار می‌رود مشکل ایجاد می‌کند. مثلاً گاهی فردی مطلبی را می‌داند ولی بدلیل ضعف در نوشتن کلمات، املائی کلمات را نادرست می‌نویسد که اگر این جواب، درست تلقی شود در نتیجه آزمون مشکل ایجاد می‌کند و اگر املائی کلمه مدنظر قرار نگیرد، مشخص نمی‌شود که کلمه دارای املائی غلط تا چه حد می‌تواند پاسخ صحیح را برساند. در این موارد پیش‌بینی نحوه نمره‌دهی الزامی است.

مثال: نام دیگر گوانتیدین چیست؟

پاسخ صحیح Ismelin است که اگر دانشجو Ismelan بنویسد، معلوم نمی‌شود آیا جواب صحیح را نمی‌دانسته یا در املائی کلمه مشکل داشته است.

با توجه به معایب این سؤالات، استفاده از آنها را به موارد زیر محدود کنید:

۱- اگر احتمال می‌دهید در صورتی استفاده از پرسشهای چند گزینه‌ای استفاده کنید، دانشجو با دیدن گزینه‌ها فوراً جواب را حدس می‌زند.

۲- یکی از اهداف شما اندازه‌گیری توانایی یادآوری مطالب توسط دانشجو باشد.

سؤالات تشریحی^۲

سؤالات تشریحی از گروه آزمونهای ذهنی است. در آزمونهای دانشگاهی به ندرت از سؤالات تشریحی استفاده می‌شود؛ زیرا معمولاً کاربرد این نوع سؤال باعث نارضایتی شدید دانشجویان می‌گردد. اما آیا هیچوقت از خود پرسیده‌اید که چرا دانشجویان اغلب از پرسش تشریحی گریزان بوده، به جای پاسخ‌گویی به هفت تا هشت سؤال تشریحی و رهایی از امتحان ترجیح می‌دهند قریب به شصت تا هفتاد سؤال چندگزینه‌ای را جواب بدهند؟!

آزمون تشریحی نوعاً با ابهامات فراوانی روبروست. فرض کنید به عنوان یک طراح سؤال قصد دارید از درسی که یک نیمسال تدریس آن را بر عهده داشته‌اید با هدف به حداقل رساندن پاسخ‌گویی حدسی دانشجویان، آزمون تشریحی برگزار کنید.

اگر قرار باشد همه آنچه با قبول زحمت فراوان تدریس کرده‌اید به شکل تشریحی مورد سؤال قرار دهید با ضیق وقت مواجه می‌شوید و اگر بخواهید با در نظر گرفتن وقت محدود، تعداد سؤالهای خود را کم کنید مجبور می‌شوید از بسیاری مطالب مهم تدریس شده چشم‌پوشی نمایید که در این صورت روایی امتحان را به خطر انداخته‌اید.

پس چه باید کرد؟ در اینجا بهترین راه‌حل، حذف صورت مسأله است. یعنی بهتر است این قانون را بپذیریم که «فقط در صورت ناچارى از سؤالات ذهنی استفاده کنیم». به عنوان یک آزمون‌گیرنده چه وقت مجبوریم از سؤالات تشریحی استفاده کنیم؟ سؤالات عینی اکثر اوقات (حدود ۷۰ تا ۸۰ درصد موارد) راهگشا بوده، برای طرح آزمون می‌توان آنها را به کار برد. اما این پرسشها هرگز نمی‌توانند سطوح پیشرفته یادگیری را در سطح تجزیه و تحلیل یا ترکیب اندازه بگیرند. پس هر گاه هدف آن باشد که دانشجویان از مطالب تدریس شده استنتاج کرده، از نتایج آن برای حل مسائل پیچیده استفاده کنند ناچار باید به سؤالات تشریحی متوسل شویم.

پرسش تشریحی چیست؟

پرسش تشریحی سؤالی است که فرد مورد آزمون، خود باید پاسخ آن را تهیه کرده، سازمان‌بندی کند و با یک روال منطقی بنویسد. مهمترین ویژگی این نوع آزمون آزادی فرد مورد آزمایش در پاسخ‌گویی است. بنابراین، موارد کاربرد این نوع آزمون در امتحاناتی است که هدف از اجرای آنها بررسی قدرت درک مطلب، طرز تفکر، سبک نویسندگی، نشانویسی و قدرت نتیجه‌گیری از طیف وسیعی از معلومات باشد.

از آنجا که در این نوع سؤال نمی‌توان هیچ الگوی خاصی به عنوان جواب انتخاب کرد برای تصحیح پاسخها هم نیاز به فردی است که کاملاً راجع به موضوع امتحان مطلع باشد. (بر خلاف پرسشهای چندگزینه‌ای که در آن فرد مصحح نیاز به اطلاعات و دانش خاصی نداشته، تصحیح با استفاده از یک کلید از پیش تهیه شده صورت می‌گیرد.)

سؤالات تشریحی به دو دسته تقسیم می‌شوند:

۱- سؤالات تشریحی با جواب محدود؛

۲- سؤالات تشریحی گسترده پاسخ.

۱- سوالات تشریحی با جواب محدود

در این دسته از پرسشها، آزمون‌گیرنده با بکار بردن کلماتی خاص در صورت سؤال مثل «نام ببرید، تعریف کنید، دو یا چند علت ذکر کنید» و یا با دادن دستورالعملهای خاص در صورت سؤال، پاسخ را محدود می‌کند.

مثال:

۱- علل افزایش قندخون را متعاقب کاهش انسولین نام ببرید.

۲- اجزای مختلف سیستم تحریکی هدایتی را نام ببرید.

این نوع پرسش معمولاً توان ندارد مهارتها و استعدادها را به خوبی نوع گسترده پاسخ محک بزند. یعنی اندازه‌گیری قابلیت‌هایی مثل فهم و درک و کاربرد که بخش‌های سطحی‌تر یادگیری هستند با آن امکان‌پذیر است و بیشتر، استعداد فرد را در به خاطر آوردن داده‌ها و وقایع می‌سنجد.

البته می‌توان برای سنجش توان دانشجو در تفسیر و کاربرد اطلاعات با این نوع سؤال از پرسشهای علت و معلول و یا ذکر یک روش یا یک رابطه ریاضی برای پرسش با توجه به دانسته‌های قبلی استفاده کرد.

ممکن است بتوان با محدود کردن جای‌خالی برای پاسخ‌گویی، پاسخ دانشجو را محدود کرد اما اغلب تفاوت در دستخط افراد چنین تدبیری را با شکست مواجه می‌کند.

در پرسشهای تشریحی با جواب محدود، دامنه تغییر پاسخ دانشجو بسیار زیاد است و به سختی می‌توان تصمیم گرفت که کدام پاسخ قبول و کدام مردود است. به مثال نمونه زیر و پاسخهایی که دو دانشجو به آن داده‌اند توجه کنید:

• «وقتی انسولین افزایش می‌یابد قند خون چه تغییری می‌کند؟»

پاسخ دانشجوی الف: کاهش می‌یابد

قبول

پاسخ دانشجوی ب: تغییر کرده، نوسان می‌یابد.

مردود اما قابل قبول

پاسخ دوم پاسخ دقیقی به این پرسش نیست ولی هرگز نمی‌توان ادعا نمود که پاسخ غلطی است. معمولاً در مواجهه با چنین پاسخهای زیرکانه دانشجویان است که ضعف این نوع سؤال نمایان می‌شود.

۲- سوالات تشریحی گسترده پاسخ با جواب نامحدود

در نوع اخیر، دانشجو کاملاً می‌تواند کیفیت و روش پاسخ را به میل خود انتخاب کند.

شاید بتوانید دانشجو را از نظر صفحه و زمان محدود کنید، اما در مورد متن پاسخ هیچ محدودیتی وجود ندارد. این نوع سؤال امکان انتخاب، سازمان‌بندی، ادغام و ارزشیابی ایده‌ها را فراهم می‌کند.

برخلاف پرسشهای دسته اول که بیشتر برای ارزیابی سطوح یادگیری سطحی‌تر کاربرد دارد، نوع دوم (گسترده پاسخ) بیشتر در سطوح یادگیری عمقی‌تر مثل ترکیب و ارزشیابی کاربرد داشته، به دانشجو اجازه خلاقیت می‌دهد.

مربوط کردن پرسشهای محدود پاسخ به سطح یادگیری موردنظر آسانتر از گسترده پاسخ است. گرچه پرسشهای محدود پاسخ پایاتر از پرسشهای تشریحی و انشایی (گسترده پاسخ) است، این پرسشها نمی‌تواند جانشین آن شود (چون قادر به اندازه‌گیری توانایی‌های یادگیری در سطوح بالا نیست).

کاربرد آزمون تشریحی

آزمون تشریحی در موارد زیر به کار می‌رود:

- ۱- هرگاه خواهان دانستن دیدگاههای جدید دانشجویان در مورد یک موضوع هستید (این آزمون بیشتر جنبه نظرخواهی دارد).
 - ۲- در صورتی که می‌خواهید دانشجویان با الحاق مجموعه اطلاعات خود راه حل مسأله را کشف کنند، یا در پی آزمایش روشهای جدید حل مسأله هستید.
 - ۳- وقتی می‌خواهید قدرت نویسندگی، خلاقیت و ارزشیابی کلی فرد را از مطالب آموزشی بسنجید.
 - ۴- هرگاه یکی از اهداف تدریس، توانایی دانشجو در تهیه جواب توسط خود او بوده است (خودش متن جواب را بنویسد).
 - ۵- مواقعی که می‌خواهید دانشجویان را در مورد بالاترین سطوح یادگیری مورد ارزیابی قرار دهید.
- تذکره مؤکد: اگر هدف شما از برگزاری آزمون تشریحی هیچیک از مقوله‌های مذکور نیست از به کاربردن پرسش تشریحی اجتناب کنید چون کاستیهای آن بسی بیش از فایده آن است.
- نکته آخر: اگر از این نوع پرسش در یک آزمون پایان دوره استفاده می‌کنید، برای گرفتن نتیجه بهتر، آن را با سؤالات عینی همراه نمایید (مخلوطی از هر دو نوع سؤال را بکار ببرید).

طراحی پرسشهای تشریحی

حین طرح سؤال تشریحی باید نکات زیر را مورد توجه قرار دهید:

۱- قبل از طرح سؤاها اهداف کلی و جزئی تدریس را مدنظر قرار دهید. این نوع سؤال از لحاظ تعداد، بسیار محدودیت دارد پس لازم است ابتدا برای تک تک اهداف کلی، سؤال طرح کرده، اگر فرصتی باقی ماند به طرح سؤال برای اهداف جزئی تر بپردازید. به مثال زیر توجه کنید:

فرض کنید موضوع تدریس بخشی از مبحث «بیماریهای مری» است و شما مصمم هستید از متن زیر که قسمتی از تدریس بوده، سؤال طرح کنید:

«مری در انسان ۲۵ سانتیمتر طول داشته دارای سه تنگی است که مبنای تعیین فاصله هر کدام، مسافت آن تا دندان پیش است. مری بسیار مستعد کانسر می باشد که از سرطانهای شایع بوده، معمولاً خوردن غذاها و مایعات داغ از عوامل خطر آن به حساب می آید. بیشتر بدخیمی های مری در محل تنگی های آن اتفاق می افتد».

سؤال آزمون: مری چند سانتیمتر طول دارد؟ فاصله تنگیهای آن چند سانتیمتر بوده، چگونه تعیین می شود؟

آیا هدف از تدریس شما بیماریهای مری نبود؟ آیا با این پرسش توانستید چیزی از دانسته های فرد راجع به بیماریهای مری را تخمین بزنید؟ پس چرا با طرح چنین سؤالاتی فکر می کنیم معلومات دانشجو را بهتر سنجیده، تعدادی از سؤالات آزمون را بیهوده هدر می دهیم. اگر اهداف کلی و مهم تدریس را هنگام طرح سؤال فراموش کنیم فقط روایی آزمون را زیر سؤال برده ایم.

۲- به جای استفاده از عبارات مبهم مثل «به نظر شما، شما چطور فکر می کنید، هر چه می دانید بنویسید و ...»، از کلماتی استفاده کنید که به طور واضح تکلیف دانشجو را مشخص کند. مثلاً در سؤال از او بخواهید که دو مورد را با هم مقایسه کند و یا از عباراتی مثل: «توضیح دهید، دلایل آن را بنویسید، پیش بینی کنید، استفاده کنید، شرح دهید، ربط دهید، تفسیر کنید، فهرست کنید، مقایسه کنید، مختصر کنید؛ چرا و ...» استفاده کنید.

بد نیست چند مورد از سؤالات مبهم و غیر صریح را مرور کنیم:

مثال: «راجع به کار هورمون تیروئید چه می دانید؟»

بسیاری از استادان گمان می کنند با طرح چنین سؤالی می توان فهمید چه کسی واقعاً

درس را خوب یاد گرفته است؛ چون به گمان آنها هیچ نوع قرینه لفظی که هدایت‌کننده دانشجو به سوی جواب باشد در صورت سؤال وجود ندارد.

اما آیا هیچ فکر کرده‌اید که دانشجو در جواب به این سؤال در چه تنگنایی گرفتار می‌شود. او برای راضی کردن شما باید اثر هورمون تیروئید را بر چه چیز توضیح دهد. اثر آن را بر قلب بگوید یا بر گردش خون و یا متابولیسم؟!

ممکن است گفته شود که باید تمام اعمال آن را بنویسد، آیا بهتر نیست به جای درگیر کردن دانشجو در ده صفحه از کتاب فیزیولوژی گایتون، سؤالات تشریحی واضح و روشن راجع به هر عنوان به او بدهید.

مثال: اثر هورمون تیروئید را بر گردش خون توضیح دهید؟

در این سؤال سیر تفکر دانشجو مشخص شده، بنابراین، شکی نیست که بهتر از مثال قبل است.

مثال: به نظر شما چرا در ارتفاعات آب بدن کم می‌شود؟

این پرسش نیز نارساست، چون هر کس برای خود دیدگاههایی دارد. در صورت به کار بردن عبارت «به نظر شما» باید منتظر بیان نظر افراد چه درست و چه غلط باشید و نمی‌توانید به آنها ایراد بگیرید، چون صرفاً نظرشان را خواسته‌اید. پس با صراحت به دانشجو بگویید کدام قسمت از نظر او را جویا هستید. مثلاً برای او توضیح دهید که با توجه به عملکردهای تنفسی، چرا در ارتفاعات آب بدن کم می‌شود.

۳- قبل از طرح هر سؤال دقت کنید که آیا سؤال شما در حد تجزیه و تحلیل است یا در حد ترکیب. البته اطمینان از این امر مشکل است. برای نیل به این هدف می‌توانید خودتان یکبار پاسخ سؤال را به تفصیل نوشته، بعد ببینید واقعاً این پاسخ شما در حد تحلیل و ترکیب است یا خیر. البته این امر به میزان مهارت شما بستگی دارد.

۴- اگر در صورت سؤال مجبور به استفاده از کلماتی مثل: «که، چه، چه وقت، نام ببرید، کجا و ...» شدید، بدانید که سؤال در سطوح پایین یادگیری بوده، باید آن را با یک سؤال عینی عوض کنید. در طراحی سؤال بهتر است به توصیه‌های زیر توجه کنید:

۱- برای طرح سؤال در حد بازشناسی و یادآوری از سؤالات عینی استفاده نمایید؛

۲- در سطوح درک، کاربرد و تفسیر از سؤالات عینی یا تشریحی از نوع محدودپاسخ استفاده کنید؛

۳- در طراحی پرسش در سطوح تجزیه، ترکیب و ارزشیابی، پرسشهای تشریحی

گسترده پاسخ را بکار ببرید.

۵- اگر به جای طرح یک سؤال گسترده پاسخ از چند سؤال محدود پاسخ استفاده کنید پایایی را افزایش داده، به کیفیت آزمون خود کمک کرده‌اید.

مثال: عمل هورمون تیروئید را بر دستگاہ گردش خون توضیح دهید؟ (سؤال گسترده پاسخ)

این سؤال را می‌توانید با سؤالات زیر عوض نمایید:

۱- هورمون تیروئید چه اثری بر روی میزان جریان خون دارد؟

۲- هورمون تیروئید چه اثری بر روی قدرت انقباضی و ضربان قلب داشته، چگونه این اثر را اعمال می‌کند؟

۳- هورمون تیروئید چه اثری بر روی فشارخون و فشار نبض دارد؟

۴- در بیماران تیروئیدی در حالت هیپوتیروئیدی هریک از موارد فوق چه تغییری می‌کند؟

۶- قبل از طراحی سؤال اعمال ذهنی را که دانشجو برای یافتن جواب باید از آنها استفاده کرده، به سطح یادگیری مورد نظر برسد نوشته سپس براساس آن به طرح سؤال بپردازید. پیشنهاد می‌شود در این مرحله، پاسخ سؤال را خودتان یکبار نوشته، ببینید آیا از تک تک مراحل مورد نظر گذشته‌اید یا خیر.

۷- در برگه سؤالات، امتیاز (یا میزان نمره) هر سؤال را مشخص کنید تا ضمن آگاه ساختن دانشجو، برای تصحیح اوراق با مشکل کمتری مواجه شوید.

۸- پیش از برگزاری آزمون در مورد اختصاص نمره به عواملی مثل شیوه نگارش، غلطهای املائی، روانی لغات و ... تصمیم بگیرید. اگر قصد دارید به این موارد نمره‌ای اختصاص دهید حتماً در برگه آزمون این موارد را ذکر نمایید.

۹- برای نظارت دقیق و کامل بر برگزاری و تصحیح این نوع آزمون، سعی کنید وقتی تعداد دانشجویان اندک هستند از این شیوه استفاده نمایید.

۱۰- هنگام برگزاری آزمون تشریحی باید مطمئن باشید که این طریقه امتحان هیچ لطمه‌ای به طرح سؤال از کل محتوای درس نمی‌زند.

۱۱- برای جلوگیری از طولانی شدن بی‌مورد جواب دانشجو می‌توانید تعداد صفحه یا زمان آزمون را کاهش دهید، در صورت به کار بردن چنین تدبیری اطمینان حاصل نمایید که وقت لازم را برای تفکر و نگارش در اختیار دانشجو قرار داده‌اید.

۱۲- در سه حالت می‌توانید از سؤالات انتخابی استفاده کرده، امکان انتخاب به دانشجو بدهید.

اول: اندازه‌گیری مهارت نوشتن؛ یعنی وقتی هدف از اجرای آزمون، سنجش اطلاعات نظری نیست و فقط خواهان سنجش قدرت نوشتن دانشجویان هستید مثل امتحانات گزارش‌نویسی.

دوم: اندازه‌گیری خلاقیت دانشجو.

سوم: مورد سؤال قرار دادن نتایج پژوهش دانشجو وقتی که او علائق خود را با مطالعه مستقل دنبال کرده است. مثل زمانی که چند موضوع برای تحقیق ارائه کرده، در آزمون نهایی راجع به همه این موضوعات سؤال طرح می‌کنید. البته هر کس فقط سؤالهای مربوط به موضوع تحقیق انتخابی خود را پاسخ می‌دهد. اما به دلایل زیر تا حد امکان سؤال اختیاری به دانشجو ندهید.

الف: مقایسه بین دانشجویان غیرممکن می‌شود چون سؤالاتی که به آنها پاسخ گفته‌اند یکسان نیست.

ب: دانشجویان باهوش یا درس‌خوان معمولاً خود را با سؤالات مشکل‌درگیر می‌کنند.

ج: دانشجویان فقط تعداد خاصی از عناوین را مطالعه کرده، هنگام امتحان به سؤالات مربوط به آن عناوین پاسخ می‌گویند.

د: گاهی دانشجو واقعاً نمی‌داند کدام سؤال را به طور کامل می‌تواند پاسخ دهد بنابراین مقداری از پاسخ سؤال را نوشته سپس متوجه می‌شود اطلاعاتش برای آن کافی نبوده است که باعث هدر رفتن وقت او می‌شود.

ه: طرح سؤالاتی با دشواری یکسان بسیار مشکل خواهد بود. تنها در صورت طرح سؤالات با دشواری یکسان می‌توان اطمینان داشت هرکس، هر نوع انتخابی در امتحان انجام دهد باز هم دشواری امتحان برای همه یکسان بوده است.

بنابراین همه باید به سؤالات یکسانی پاسخ گویند؛ زیرا اجازه انتخاب سؤال، پایایی و روایی آزمون را به خطر می‌اندازد.

معایب آزمونهای تشریحی

آزمونهای تشریحی با محدودیتهایی روبرو بوده است و معایبی نیز دارند، از جمله:
۱- توان دانشجو در گزافه‌نویسی و غلو کردن در پاسخ و همچنین روانی قلم او بشدت در پاسخهای او اثر می‌گذارد.

مثال: استاد ادبیات انگلیسی دانشکده‌ای، دانشجویان را برای بازدید یک اثر تاریخی برد و از

آنها خواست به طور تشریحی این مکان را برای او توصیف کنند. یکی از دانشجویان که اصلاً در این بازدید شرکت نکرده بود به صورت زیر پاسخ داد و بهترین نمره را از استاد گرفت و استاد زیر برگه او چنین نوشت «این بهترین گزارشی بود که در تمام دوران تدریس از دانشجویانم دریافت کرده‌ام».

این دانشجوی زیرک فقط با استفاده از صحبت‌های دیگران و بهره‌گیری از کمی استعداد و تخیل، گزارش زیر را تهیه نمود:

«هوای دل‌انگیز پاییزی جانم را نوازش می‌کند و ترنم باران، باطنم را صفا می‌بخشد و چه جایی بهتر از اینجا برای تفکر به خود، که اینجا روزگاری خانه با شکوه آدمیانی بوده است چون من و تو، که با غرور و نخوت بر سنگفرش آن گام می‌نهادند و گمان می‌کردند که حتی اگر هفت طبقه افلاک بشکافت اریکه قدرت آنها را خلی نیست. از در بزرگ و باشکوه این مکان قدم به درون می‌گذارم و بر تقدیر شوم مردمانش تأسف می‌خورم. چه هوای دل‌انگیزی در راهروهای تاریک و تو در توی آن وجود دارد. خودم را در لباس بلند شوالیه‌ای احساس می‌کنم که صدای مستحکم قدم‌هایش زمین این کاخ عظیم را به لرزه می‌آورد و ...» معمولاً ما خیلی زود تحت تأثیر ترفندهای چنین دانشجویانی قرار می‌گیریم و چون از این حالت گریزی نیست بهتر است اصلاً خود را با چنین نوع سؤالاتی درگیر نکنیم.

۲- نمره‌دهی آن چندان پایا نیست.

۳- با توجه به محدودیت تعداد پرسشها، این آزمون چندان روا نیست، زیرا نمی‌تواند کل محتوای درس را در برگیرد.

۴- طرح سؤال مرتبط با سطح یادگیری موردنظر بسیار مشکل است. البته بیشتر وقتها در مورد طرح هر نوع سؤال مرتبط با سطوح عالی یادگیری دچار مشکل می‌شویم.

نمره‌گذاری آزمون تشریحی

به طور کلی دو روش برای نمره‌گذاری در آزمونهای تشریحی وجود دارد:

۱- روش نقطه‌ای^۱: این شیوه بیشتر در نوع سؤالات تشریحی محدود پاسخ به کار می‌رود. در این جا هر پاسخ را به طور مستقل با کلید مقایسه کرده، به نکات آن نمره می‌دهند.

۲- روش دسته‌بندی^۱: در این شیوه ابتدا در مورد هر سؤال برگه‌های کلیه دانشجویان را می‌خوانند. سپس آنها را در چهار دسته عالی، خوب، متوسط و ضعیف قرار داده، هر گروه را به طور جداگانه نمره می‌دهند. یعنی در اینجا علاوه بر مقایسه پاسخ دانشجویان با کلید تصحیح، با پاسخ سایر دانشجویان همگروه نیز مقایسه شده، ارزش نمره‌ای پاسخ او تعیین می‌شود.

بهرتر است در نمره‌دادن به آزمونهای تشریحی نکات زیر را در نظر بگیرید.

۱- باید کلید تصحیحی تهیه کرده، هر جزء اطلاعات کوچک و مهم را در آن نمره‌گذاری کنید.
۲- میزان نمره اختصاص یافته به سبک نگارش و رعایت اصول آن مثل نشانه‌گذاری را از پیش تعیین کنید.

۳- هنگام تصحیح، ابتدا یک سؤال را در همه برگه‌ها نمره‌گذاری کرده، به قضاوت یکسانی در مورد همه پرسشها برسید، سپس به سراغ سؤالات بعدی بروید.

۴- قبل از تصحیح دقیق سؤال، ابتدا می‌توانید نظری اجمالی به پاسخ دانشجو انداخته، یک برآورد کلی بدست آورید. این کار به نمره‌گذاری جزء جزء اطلاعات کمک می‌کند.

۵- هنگام تصحیح روی نام دانشجو را بپوشانید. چون زمینه فکری قبلی و تصور ذهنی شما از دانشجویان ضعیف و قوی نمره‌گذاری را متأثر می‌کند.

۶- برگه‌ها را در یک دوره زمانی، به طور مداوم تصحیح نکنید، زیرا دانشجویانی که در پایان این دوره برگه آنها را تصحیح می‌کنید به علت خستگی شما نمره کمتری می‌گیرند.

۷- در برگه دانشجو اشکالات او را به صورت ریز نوشته، اطلاعاتی در مورد پاسخ به او بدهید (با نوشتن تذکر و توضیح درون برگه امتحان).

۸- اگر در آزمونهای تشریحی، پاسخ هر سؤال را به اجزای کوچک تقسیم کرده، برای هر جزء نمره‌ای در نظر بگیرید نمره‌گذاری پایاتر است. ایراد این شیوه آن است که شاید دانشجو همه اجزای موردنظر شما را بدون برقراری هیچگونه ارتباط منطقی بین آنها در پاسخ خود به کار برده باشد یا به استنتاجهای لازم نرسیده باشد. مثلاً در یک امتحان توصیفی مثل انشا ممکن است فرد، تمام اصول نگارش را رعایت کرده، تک تک جملاتش با ارزش و معنی‌دار باشد در این صورت اگر شما به این قسمتها جداگانه نمره دهید او نمره خود را خواهد گرفت در صورتی که شاید جملات به کار برده شده اصلاً به هم مرتبط نباشد.

۹- پس از امتحان اگر کلید تصحیح را به دانشجو نشان دهید می‌توانید به او نوعی بازخورد داده، در یادگیری به وی کمک کنید.

طراحی آزمونهای تشریحی

- در خاتمه توجه شما را به نکات زیر در طراحی آزمون تشریحی جلب می‌نمایم.
- ۱- زمانی که در امتحان چندگزینه‌ای توسط دانشجو صرف می‌شود، زمان خواندن و فکر کردن است؛ اما زمان به کار رفته در آزمونهای تشریحی، زمان نوشتن و فکر کردن است، چون نوشتن بیشتر از خواندن طول می‌کشد طبیعتاً زمان هر سؤال در امتحان تشریحی به طور آشکار طولانی‌تر از آزمون چندگزینه‌ای است.
 - ۲- در آزمونهای تشریحی، پاسخهای مختصر و ناکامل به سؤال نیز تا حدی نمره می‌گیرند، یعنی اطلاعات اندک دانشجو هم راجع به موضوع ارزش دارد، اما در آزمونهای چندگزینه‌ای اطلاعات نسبی بی‌ارزشند^۱ و آنچه اهمیت دارد اطلاع مطلق از موضوع بوده، باعث هدایت فرد به گزینه صحیح می‌شود. پس شاید آزمون تشریحی از این لحاظ مزیتی بر آزمون چندگزینه‌ای داشته باشد.
 - ۳- نمره‌گذاری امتحان تشریحی مشکل‌تر از چندگزینه‌ای است؛ ولی در عوض، طراحی سؤالات چندگزینه‌ای دقت و زمان بیشتری لازم دارد. پس برای انتخاب بین این دو نوع امتحان بهتر است به تبحر خود توجه کنید و در نظر بگیرید آیا طرح سؤال برای شما آسانتر است یا نمره دادن دقیق.
 - ۴- امتحان تشریحی نقاط ضعف دانشجو را به خوبی آزمون چندگزینه‌ای مشخص نمی‌کند چون دانشجویان در این آزمونها با انتخاب جوابهای انحرافی خوب طراحی شده، می‌توانند منشا اشتباه خود را نشان دهند، اما در آزمون تشریحی چنین امکانی کمتر وجود دارد.
- نکته آخر: آنچه ما در سؤالات چندگزینه‌ای به دانشجویان القا می‌کنیم تشویق حدس زدن است، اما در سؤالات تشریحی امکان حدس نیست. در عوض ممکن است دانشجو تشویق به اضافه‌نویسی و گزافه‌نویسی می‌شود.

۱- البته در صورتی که در نمره‌دهی، از حد نصاب قبولی (Minimum Pass Level) استفاده شود تا حدی این اطلاعات نیز ارزشگذاری می‌گردند. شیوه محاسبه MPL در فصل تجزیه و تحلیل مورد اشاره قرار گرفته است.

سؤالات تشریحی تغییر یافته ۱

نوعی از سؤالات تشریحی که بیشتر برای سنجش قدرت حل مسائل پزشکی به کار می‌رود، سؤالات تشریحی تغییر یافته نام دارند. مسائل طرح شده در این سؤالات شبیه یک مورد بیمار حقیقی است. همچنین می‌توان در آن از منحنی‌ها (مثل الکتروکاردیوگرام، الکتروانسفالوگرام و ...) یا تصاویر رادیولوژی هم استفاده کرد. این سؤالات می‌تواند سطوح مختلف یادگیری را از فراخوانی گرفته تا سطوح پیچیده‌تر، ارزیابی نماید.

طرح سؤالات تشریحی تغییر یافته بسیار وقت‌گیر است و طراح آن نیاز به تخصص کافی دارد ولی با توجه به هزینه و زحمت انجام ارزیابی عملی این سؤالات زمینه انجام نوعی ارزیابی کتبی از قابلیت‌های عملی را فراهم می‌سازد.

این سؤالات یک سلسله مراتب تصمیم‌گیری بالینی را ارائه می‌دهند و از این نظر نوعی آزمون اداره بیمار^۲ محسوب می‌شوند. در این سؤالات دانشجو نمی‌تواند به صفحات قبلی برگردد و مطالب نوشته شده را تغییر دهد. دفترچه سؤالات اغلب شش تا ده صفحه دارد. در صفحه اول راهنمایی‌های لازم در مورد امتحان آمده است، مثلاً تعداد سؤال، زمان آزمون و تذکر در مورد اینکه دانشجو نباید به صفحات قبلی برگردد و پاسخها را تغییر دهد. در صفحه دوم یک مورد بالینی به صورت مختصر و کوتاه بیان شده است و در انتهای همین صفحه تعدادی سؤال با محل کافی برای نوشتن پاسخ طرح شده است. در صفحه بعدی اطلاعات بیشتری در مورد موضوع بیان گشته است که در واقع بخش دیگری از مشکلات بیمار هستند و سؤالات دیگری هم در همان زمینه مطرح گردیده است.

پس از طرح سؤال، آزمون‌گیرنده باید از میان پاسخهای مختلف ممکن در مورد هر سؤال، بهترین را تعیین و ارزش و نمره پاسخهای دیگر را هم در راهنمای تصحیح و نمره‌گذاری بیان کند. ممکن است به همه قسمت‌ها ارزش یکسانی داده شود و یا بعضی قسمت‌ها بیشتر یا کمتر از بقیه نمره بگیرند. در هر حال نمره هر قسمت باید از قبل مشخص شده باشد. مجموع نمرات هر بخش نمره حاصل شده از این آزمون خواهد بود. به مثالی از این نوع سؤالات توجه کنید.

راهنمایی روی جلد

۱- این آزمون از سه سؤال تشکیل شده است. شما پانزده دقیقه فرصت دارید تا به آنها

پاسخ بدهید.

۲- قبل از شروع امتحان صفحات را ورق نزنید و مطالب صفحات را نخوانید.

۳- به صفحات قبلی برنگردید و مطالب نوشته شده را تغییر ندهید.

صفحه اول:

زن ۲۵ ساله‌ای به علت تپش قلب و پرنوشتی به اورژانس مراجعه کرده است. سردرد صبحگاهی و پلی‌اورمی دارد. در سابقه خود دفع یک سنگ ادراری را ذکر می‌کند که در آنالیز، آن را از دسته اوراتی گزارش کرده‌اند. در معاینه عمومی هوشیار است و در معاینه ریه‌ها و شکم مشکلی ندارد. در بررسی آزمایشگاهی میزان قند خون ناشتا در حدود 262mg/dl می‌باشد.

سؤال ۱:

با توجه به موارد ذکر شده در بالا به نظر می‌آید بیمار مشکلات متعددی دارد. قطعی‌ترین تشخیصی فعلی خود را بنویسید؟

صفحه دوم:

در بررسی‌های بعدی میزان کراتینین سرم 0.9 و اسیداوریک سرم 7.5mg/dl گزارش شده است.

سؤال ۲:

الف) با توجه به اطلاعات قبلی، آیا نتایج آزمایشگاهی به دست آمده قابل قبول است؟
ب) آیا علت سردرد صبحگاهی و پلی‌اورمی بیمار را می‌توان به ناراحتی کلیوی وی نسبت داد؟ چرا؟

صفحه سوم:

در آزمایش ادرار، قند(+++) و استون منفی بود و آزمایش آنتی‌ژن HBS مثبت بود.

سؤال ۳:

الف) چگونه این نتایج را در این حالت تفسیر می‌کنید؟

ب) به این بیمار چه توصیه‌ای می‌کنید؟

منابع:

- ۱- آذر، رسول. روشهای ارزیابی آموزشی. تهران: انتشارات نیما، ۱۳۶۹.
- ۲- پاشا شریفی، حسن. اصول روان‌سنجی و روان‌آزمایی. تهران: انتشارات رشد، ۱۳۷۳.
- ۳- غفوری، علی. جزوه تست‌های چند جوابی (MCQ). تهران: حوزه معاونت آموزشی مرکز توسعه آموزش پزشکی.
- ۴- گرون لوند، نورمن. آزمونهای پیشرفت تحصیلی. تهران: انتشارات سازمان سنجش، ۱۳۵۶.
- ۵- گنجی، حمزه. اندازه‌گیری و ارزشیابی در آموزش و پرورش. تهران: انتشارات بعثت، ۱۳۷۳.
- ۶- مایزما، ویلیام؛ و همکاران. اندازه‌گیری و آزمون در تعلیم و تربیت. ترجمه غلامرضا خوئی‌نژاد. مشهد: انتشارات مؤسسه چاپ و انتشارات آستان قدس رضوی، ۱۳۷۲.
- ۷- هومن، حیدر علی. اندازه‌گیری‌های روانی و تربیتی و فن تهیه تست. تهران: انتشارات فردوسی، ۱۳۷۴.
- ۸- حوزه معاونت آموزشی مرکز توسعه آموزش پزشکی اصفهان. جزوات کارگاههای آموزشی در مورد روش ارزشیابی در مرکز توسعه آموزش پزشکی. اصفهان: ۱۳۶۹.
- 9- Abbatt FR. *Teaching For better learning*. 2nd ver.ed. Geneva: World Health Organization, 1992.
- 10- Clderhead J. *Exploring teacher's thinking*. Cassel Educ 1987; 104-125
- 11- Cox K, R Christion E. *The medical teacher, writing multiple choice question*. London: Churchill- Livingstone, 1982.
- 12- Dixon RA. *Evaluation and improving multiple choice paper and T/F questions in public health medicine*. Med Educ 1990; 28(5): 400-8.
- 13- Jenkins DM. *Does incourse assessment and counselling improve exit performance?* Med Teach 1992; 14(2/3): 241-2.
- 14- Katz FM, Fnow R. *Assessing health worker's performance, a manual for training and supervision*. World Health Organization Series, 1980; 72: 23-41.
- 15- Lead MC. *Immediate postexamination review of multiple choice question*. Teach & Learn in Med 1995; 6: 67-70.
- 16- Lowe D. *How to do it? set a multiple choice question examination*. BMJ 1991; 302-7.
- 17- Reece D. *How to do it*. London: BMJ publising group, 1995.
- 18- Richardson R. *The multiple choisce T/F question: what dose it measure and what could it measure?* Med Teach 1992; 14 (2/3): 201-5.
- 19- Robinson BV. *MCQ in pharmacologi*. London: Churchill Livingstone, 1989.
- 20- Shick J. *Tantalizing text book, part II: T/F, matching, completion, essay*. Health Educ 1980. APR/May; 4, 5: 18-22.
- 21- Tony; Buzen. *Use your head*. London: Brithish broad casting corporation, 1974.
- 22- Webber H. *Structural short answer questions: an alternative examination method*. Med Educ 1992; 26: 58-62.

فصل چهارم

آزمون اداره بیمار

آزمون اداره بیمار^۱

در طی دوران کارورزی، دانشجو باید بیماری را که تازه به بیمارستان آورده شده معاینه کند و از او اطلاعاتی کسب نموده، در مورد او تصمیماتی بگیرد و حتی در صورت نیاز دستور آزمایشاتی را بدهد که نتایج آن در تشخیص و درمان مؤثر باشد. گاهی مشکلاتی در حین تشخیص و درمان به وجود می‌آید که با توجه به آنها کارورز باید تصمیم جدیدی بگیرد تا نهایتاً درمان مطلوب حاصل شود. اما چگونه می‌توانیم توانایی دانشجو را در انجام این اعمال ارزیابی کنیم.

تا سال ۱۹۶۳ آزمون‌گیرندگان سعی می‌کردند دانشجو را با بیمار واقعی روبرو کرده، سؤالاتی را مطرح سازند. بعدها با افزایش تعداد دانشجویان، به جای بیمار از فیلمهای صامت و رنگی از بیماران استفاده کردند. آزمونهای شفاهی نیز توسط آزمون‌گیرندگان با سؤالات چندگزینه‌ای جایگزین شد.

امروزه به جای فیلمهای رنگی از شیوه^۲ «ارائه مورد^۱» استفاده می‌شود. در این شیوه ابتدا اطلاعاتی در مورد بیمار عرضه شده، سپس سؤالات موردنظر بر مبنای آن، طرح می‌گردد. به طور معمول با کاربرد این شیوه دو هدف دنبال می‌شود:

- ۱- آیا دانشجوی بالینی می‌تواند نشانه‌های بیماری را به طور مطلوب پیداکند.
- ۲- آیا دانشجوی بالینی می‌تواند با توجه به علائم، بیماری را تشخیص داده، درمان رضایت بخش را انتخاب و در مورد آن قضاوت نماید.

برای دستیابی به چنین اهدافی و ارزیابی کارایی بالینی پزشکان شیوه‌های شبیه‌سازی^۳ از بهترین روشهای موجود است و در آزمونهای اداره بیمار نیز بر همین اساس از ارائه شرح حال بیماران واقعی استفاده می‌گردد.

آزمونهای اداره بیمار در حقیقت نوع توسعه یافته‌ای از سؤالات کوتاه پاسخ می‌باشد که مانند سؤالات چندگزینه‌ای جنبه عینی دارد. این روش اولین بار در آمریکا به سال ۱۹۶۱

1- Patient Management Problem (PMP)

2- Case Presentation

3- Simulation

توسط NBME ابداع و اجرا شد. این آزمونها برای ارزیابی دانشجویانی که قرار است پرستاران و پزشکان آینده باشند بسیار مناسب است.

در این امتحان یک موقعیت بالینی همانند موقعیت واقعی رویارویی پزشک با بیمار شبیه سازی می شود. البته بیمار جای خود را به یک گروه اطلاعات نوشته شده داده است که دانشجو باید بر مبنای آنها به حل مسأله برسد. ممکن است دانشجویی پس از بررسی به مطالعات آزمایشگاهی و روشهای تشخیصی نیاز پیدا کند، که باید در اختیار او قرار گیرد. در دفترچه سؤال بلافاصله بعد از شرح حال بیمار، فهرستی از روشهای ممکن اداره بیمار قرار دارد که بعضی از این روشها صحیح و لازم بوده، بقیه نادرست است. دانشجو باید از بین این روشها، مورد درست را انتخاب کند؛ بعد از آنکه او یک مورد را انتخاب کرد، باید به دفترچه جواب مراجعه کند (جوابها گاهی جلوی سؤال نوشته می شود). در آنجا یک سری خانه های تیره می بیند که هر خانه با یکی از انتخابهای او هماهنگ است. البته روی جوابها جوهر کشیده شده، که با پاک کن مخصوص، پاک می شود. وی می تواند خانه مطابق با انتخابش را با یک پاک کن، پاک کند. در این حال نتایج تصمیم او هویدا می شود. اکنون اگر تصمیم اشتباهی گرفته باشد، اشتباه او آشکار می گردد. با عوض شدن وضعیت شکل تازه ظاهر می شود و دانشجو برای این مشکل و متناسب با این وضعیت جدید باید تصمیمات جدیدی بگیرد و به همین ترتیب امتحان ادامه می یابد. این کار بین چهار تا شش بار صورت می گیرد. در این فرآیند دانشجو هر قدم که پیش می رود اطلاعاتش افزایش یافته، قدم بعدی به قدم قبلی وابسته می شود. به دلیل استفاده از خانه های تیره، دانشجو نمی تواند اشتباه خود را جبران یا در امتحان تقلب کند. اما تقلب در امتحان وقتی که جواب سؤالات با یک ورقه پوشانده شده یا در پشت یکسری کارت نوشته شده باشد پیش می آید. البته اگر دانشجو جواب را اشتباه انتخاب کرد می تواند تصمیم خود را عوض کرده، بیمار را به صورت دیگری اداره کند. گرچه او اشتباه را جبران می کند اما نمره منفی از وی کسر خواهد شد؛ در حالی که چنین کاری در واقعیت امکان پذیر نیست، مثلاً اگر بیماری بمیرد واقعیت را نمی توان با تصمیم دیگری تغییر داد اما در آزمون اداره بیمار این کار ممکن است.

مثال: یک مرد ۴۵ ساله به خاطر درد در لگن سمت راست و استخوان هیپ هنگام راه رفتن، در بیمارستان پذیرش شده است. او در طی سال گذشته ۱۲ کیلوگرم کاهش وزن داشته است، از قدرت او برای کار کردن کاسته شده است. سه ماه قبل از پذیرش، دچار عفونت حاد دستگاه تنفسی فوقانی شده که آن را به صورت درد در استخوان و خشکی مفاصل ذکر کرده است. در آن زمان او سستی عمومی، خشکی و صدا کردن استخوانهای دست را نیز داشته و صحبت کردن هم به خاطر خشکی فک و لبها برای او مشکل بوده است.

این بیمار بیست سال قبل دچار علائمی مشابه شده بود که آن را به صورت درد در تمام بدن توصیف می‌کرد. در آن زمان بیماری وی استئوپروز تشخیص داده شده و در بخش روماتولوژی تحت درمان قرار گرفته بود. در طی این سالها او نسبتاً سالم بوده است.

معاینه فیزیکی

دمای بدن 37°C ، ضربان قلب ۸۰ در دقیقه و منظم، فشار خون $120/80\text{ mmHg}$ ، وضعیت تغذیه خوب است. ریه‌ها در دق و سمع سالم، اندازه و صداهای قلب طبیعی است و سوفل از آن به گوش نمی‌رسد. شکم برآمده بوده، هیچ توده یا ارگانی قابل لمس نیست. کشاله ران راست به هنگام لمس درد داشته ولی هیچ توده‌ای حس نمی‌شود. ادم ۲+ در پاها وجود دارد ولی در انتهای پاها، انگشتان طبیعی است. معاینه عصبی مورد غیر طبیعی نشان نداده، درد شدید در استخوان هیپ و ناحیه لگن و پاها در حین راه رفتن وجود دارد.

مطالعات آزمایشگاهی

خون: هموگلوبین $10.0\text{ g}/10.0\text{ mlit}$ ، هماتوکریت ۳۵٪، لکوسیتها $6800/\text{mm}^3$ ، نوتروفیل پلی‌مورفونوکلتر ۶۰٪، لنفوسیت ۳۴٪، منوسیت ۵٪، ائوزینوفیل ۱٪، اریتروسیت $4,000,000/\text{mm}^3$.

ادرار: وزن مخصوص ادرار 1.015 ، $\text{pH} = 5/5$ ، پروتئین (+۱)، قند منفی، گلبول سفید ۳-۴ عدد، گلبول قرمز ۱-۲ عدد (high-power field)، باکتری و کریستال دیده شد. در رادیوگرافی قفسه سینه، ریه‌ها کاملاً تمیز به نظر می‌رسید.

سؤال اول

با توجه به مهم بودن بودن اطلاعات بیمار برای رسیدگی به وضعیت او در مورد کدامیک از موارد زیر سؤال می‌نمایید؟

اطلاعات قابل دستیابی	اطلاعات و نتایج فرضی
۱- اشتها	اشتهای طبیعی
۲- مصرف مرکبات، آب میوه، یا غذاهای حاوی اسید اسکوربیک	مصرف روزانه آب پرتقال تازه
۳- تعداد، حجم، محتویات مدفوع	طبیعی
۴- انتقال خون، تزریقات	منفی
۵- تماس با مردم هنگام سرفه، تب یا عفونتهای شناخته شده	تماس نداشته
۶- ریکتز در خردسالی، مصرف ویتامین د، تماس با نور خورشید	ریکتز ندارد، مصرف طبیعی، تماس داشته
۷- جراحی گلو، عفونت استرپتوکوکی، اثبات یا تشخیص عفونت یا بیماری کلیوی در خردسالی	منفی
۸- تاریخچه تورم و قرمزی مفاصل	منفی
۹- سابقه فامیلی، بدقوارگی اسکلتی یا درد استخوان	یک خواهر مبتلا به آرتریت دارد

سوال سوم:

اکنون شما باید چه چیزهایی را اندازه گیری کنید؟

اطلاعات قابل دستیابی	اطلاعات و نتایج فرضی
۱۰- ترانس آمیناز سرم	۳۰ واحد = (AST) و ۴۰ واحد = (ALT)
۱۱- کلسیم و فسفر سرم	کلسیم ۴٫۶ mg/dl و فسفر ۱٫۸ mg/dl
۱۲- آلکالین فسفاتاز سرم	۵۸ U/L (نرمال ۱۳-۵)
۱۳- سدیم، پتاسیم، کلرور و بیکربنات سرم	و $HCO_3 = 27 \text{ Eq/lit}$, $Cl = 105 \text{ meq/lit}$ $K = 4 \text{ meq/l}$, $Na = 142 \text{ meq/l}$
۱۴- نیتروژن و اوره خون	۱۴ mg/l
۱۵- آهن سرم و TIBC	$Fe = 40 \mu\text{g/l}$, $TIBC = 280 \mu\text{g/l}$
۱۶- اسید فسفاتاز سرم	۱ واحد (k. A)
۱۷- کلیرانس فسفات	۲۳ ml/min، نرمال ۱۰-۵ ml/min
۱۸- تیر آتی استریتولیزین C	(۱۲۵ tod)
۱۹- دفع بروم سولفالتین	۵٪ در ۴۵ دقیقه
۲۰- اسیداوریک سرم	۶ mg/l cc

سؤال سوم:

با توجه به یافته‌های قبلی کدامیک از روشهای تشخیصی زیر را برای به دست آوردن اطلاعات مفید، مناسب می‌دانید؟

اطلاعات قابل دستیابی	اطلاعات و نتایج فرضی
۲۱- پیلوگرام داخل وریدی	طبیعی
۲۲- رادیوگرافی از اسکلت	دیمینرالیزه شدن عمومی و شکستگی کاذب دوطرفه ران و کتف
۲۳- تست شیلینگ ^۱ (تزریق ویتامین ب _{۱۲})	۶٪ دفع
رادیوآکتیو)	
۲۴- تعیین چربی در مدفوع	۳۰g چربی در ۷۲ ساعت
۲۵- تست تحمل گلوکز (دهانی)	منحنی low flat
۲۶- تعیین مقدار دفعی متیل مالئات در ادرار	طبیعی
۲۷- تست تحمل د- گریلوز	۱g از ۲۵g در ۵ ساعت دفع شد
۲۸- سیستوسکوپی	طبیعی
۲۹- رادیوگرافی روده کوچک	تورم در قسمت فوقانی روده کوچک وجود دارد و سگمانته است
۳۰- باریم انما	طبیعی

سؤال چهارم:

حالا باید کدامیک از آزمایشهای زیر را درخواست نمایید؟

اطلاعات قابل دستیابی	اطلاعات و نتایج فرضی
۳۱- بیوپسی کبد	طبیعی
۳۲- بیوپسی کلیه	طبیعی
۳۳- معاینه مغز استخوان	هیپرلازی ارتروئید، تخلیه ذخیره آهن
۳۴- بیوپسی روده کوچک	صاف شدن مزکهای اپی تلیال
۳۵- لاپاروتومی تشخیصی	طبیعی

سؤال پنجم:

با کدامیک از داروهای زیر به درمان می‌پردازید؟

اطلاعات قابل دستیابی	اطلاعات و نتایج فرضی
۳۶- سولفات آهن از طریق دهان	انجام شد
۳۷- انتقال خون کامل	داده شد
۳۸- کلسیم و ویتامین د	داده شد، Ca سرم بالا رفت
۳۹- رژیم غذایی خالص	داده شد
۴۰- کلرید آمونیم	داده شد
۴۱- مخلوط سترات سدیم و اسیداسکوریک	داده شد
۴۲- تزریق ویتامین ب ۱۲	انجام شد
۴۳- اسید فولیک خوراکی	داده شد
۴۴- اسیداسکوریک در دوز زیاد	داده شد

طراحی

بهترین و ساده‌ترین راه برای طرح چنین سؤالاتی آن است که سؤال خود را براساس یک مورد بیمار - که خود با آن برخورد داشته‌اید - طرح کنید. البته این امر در صورتی امکان‌پذیر است که شما خود به عنوان پزشک مشغول کار باشید. اگر تمام وقت شما صرف تدریس می‌گردد بهتر است با یک پزشک مشورت کنید. مرحله بعد تقسیم تدبیر انجام شده برای بیمار به چند بخش است. مثلاً: اول چه اتفاقی می‌افتد. بعد چه تصمیمی باید گرفت؟ در مرحله بعد باید تصمیم بگیرید چه اطلاعاتی به دانشجو بدهید و چه سؤالاتی از وی بنمایید. تمام پاسخهایی را که دانشجویان ممکن است ارائه دهند (چه صحیح، چه غلط) بنویسید و بدین ترتیب امتحان را آماده کنید.

در طراحی سؤالات به نکات زیر توجه نمایید:

- ۱- با تعریف اهداف، هدف مسأله را به وضوح مشخص سازید.
- ۲- مسأله را طوری طرح کنید که واقعاً همانند آنچه بین پزشک و بیمار اتفاق می‌افتد باشد.
- ۳- سعی کنید تصمیمات متوالی و مربوط به هم را برای حل مسأله به صورت پی در پی آورده شود.

۴- به دانشجو فرصت دهید اطلاعات واقعی در مورد هر سؤال و نتایج و کاربرد آن را در

تصمیمات بعدی پیدا کند.

۵- امتحان رابه صورتی برگزار نمایید که دانشجو تصمیمات بعدی را براساس تصمیمات قبلی گرفته، پس از انتخاب یک گزینه، قادر به برگشت از تصمیم خود نباشد. برای اینکه آزمون چنین ویژگیهایی داشته باشد لازم است یک توضیح کوتاه یا یک فیلم رنگی آماده سازید که همان موقعیت دیدار بین پزشک و بیمار را شبیه سازی کند. به این فیلم، نما^۱ یا شجره نامه می گویند که بر اساس آن دانشجو تصمیم می گیرد چگونه مسأله را ادامه دهد. شما به دانشجو فرصت انتخاب و دیدن نتیجه قضاوت خود را می دهید. سپس او به طرف تصمیم بعدی هدایت می شود و این روند تا آخر ادامه می یابد.

در شجره نامه باید موارد زیر را بگنجانید:

۱- وضعیت بدنی بیمار؛

۲- نقش دانشجو به عنوان کارورز یا دستیار؛

۳- وظیفه دانشجو؛

۴- تسهیلاتی که دانشجو در اختیار دارد؛

۵- هر گونه اطلاعاتی که لازم است دانشجو برای اولین اقدام در اختیار داشته باشد.

همچنین جوابها را طوری تنظیم کنید که به صورت جملات توضیحی باشد نه تفسیری. مثلاً: اگر دانشجو سفارش عکس برداری از قفسه سینه می دهد آن را ضمیمه سؤال کرده، از به کار بردن توضیحاتی مثل «در حدود طبیعی» یا «عوارض غیر طبیعی یافت نشد» کمتر استفاده نمایید.

۲- پاسخها را متناسب با منبع سؤال انتخاب کنید. مثلاً: اگر مورد سؤال کودک است، جوابها را نیز متناسب با سن کودک به کار برید. یا اگر سؤال، گزارش آزمایشگاه است، پاسخها را با همان اصطلاحات انتخاب کنید.

انواع آزمونهای اداره بیمار

این آزمونها دو نوع است:

۱- آزمون اداره بیمار خطی که تمام گزینهها روی وضعیت مسأله اثر می گذارد مانند نمونه ای که قبلاً ذکر گردید.

۲- آزمون اداره بیمار شاخه دار آزاد که پس از دیباچه، دانشجو به مراحل بعدی هدایت

می‌شود. هتلم:

- بیمار را معاینه فیزیکی نمایید.
با پزشک متخصص ملاقات کنید.
- به بخش N مراجعه کنید.
به بخش H مراجعه کنید.

انواع گزینه‌های به کار برده شده

پاسخهای استفاده شده در این آزمون سه دسته است که در این قسمت بیان می‌شود.

۱- کمک‌کننده: گزینه‌هایی است که برای حل مسأله ضروری است و یا آن را تسهیل می‌کند.

- ۲- خشی: که به حل مسأله کمک نمی‌کند و از طرفی مانعی برای آن به حساب نمی‌آید.
- ۳- زبان‌آور: که از حل مسأله جلوگیری می‌کند یا لازم است از انجام آنها پرهیز شود.

روش نمره‌گذاری

در این امتحانات، امتحان‌دهنده اول گزینه‌ها را انتخاب می‌کند و سپس جواب آنها را با قلم مخصوص پاک می‌کند؛ به عبارت دیگر وقتی امتحان‌دهنده تصمیمی در حل مسأله می‌گیرد، جوابش را پاک می‌کند.

اگر امتحان‌دهنده جوابی را پاک کند، پاسخ می‌تواند صحیح یا غلط باشد و اگر گزینه‌ای را پاک نکند ممکن است جواب صحیح را که برای بیمار مفید و ضروری است انتخاب نکرده باشد و یا اینکه از انتخاب پاسخ غلط پرهیز کرده است.

پس در انتخاب کردن یا نکردن هر گزینه چهار احتمال وجود دارد:

در دو حال به امتحان‌دهنده امتیاز داده می‌شود:

۱- پاک کردن جواب صحیح:

۲- پاک نکردن جواب غلط.

و در دو حال هم فرد جریمه می‌شود:

۱- پاک کردن جواب غلط:

۲- پاک نکردن جواب صحیح.

نمره‌کل امتحان مجموع جبری نمرات مثبت و منفی است.

گاهی تصمیم نادرست دانشجو یا اتخاذ نکردن تصمیم مناسب در شرایط واقعی، منجر به حوادث ناگواری می‌شود. بنابراین در امتحان شبیه‌سازی در مورد گزینه‌هایی که

انتخاب آنها نشانه بی‌اطلاعی یا بی‌توجهی دانشجو است یک نمره منفی اختصاص داده، به تصمیماتی که در جهت کمک به بهبود بیمار گرفته می‌شود یک نمره مثبت بدهید. نمره صفر را برای گزینه‌هایی که انتخاب کردن یا نکردن آنها تفاوتی ندارد، در نظر بگیرید.

گاهی دانشجو تصمیم دارد میان بر زده، یک سبک از خود ابداع کند. مثلاً: قسمت جمع‌آوری اطلاعات را حذف کند. ولی این کار برای دانشجوی بی‌تجربه که سابقه کار ندارد، مناسب نیست. بدین منظور به این قسمت و قسمت‌های مشابه نمره بیشتری اختصاص داده، جلوی این اقدام دانشجو را بگیرید.

راه دیگر نمره‌گذاری در امتحانات مذکور آن است که به گزینه‌ها بر حسب اهمیت آنها ارزش یک تا پنج نمره مثبت و نیز یک تا پنج نمره منفی داده شود. در این روش به گزینه‌هایی که انتخاب آنها برای حل مسأله لازم و اساسی است ارزش پنج نمره مثبت و برای انتخاب‌هایی که لازم ولی اساسی نیست، به ترتیب اهمیت، نمره کمتری داده می‌شود. بر عکس برای انتخاب‌هایی که لازم نبوده، ضرر دارند ارزش منفی پنج و انتخاب‌هایی که لازم نیستند ولی بی‌ضرر هستند به ترتیب اهمیت، نمرات منفی در نظر می‌گیرند. بدیهی است که برای انتخاب بی‌تفاوت یعنی تصمیماتی که نه مفید است و نه مضر، ارزش برابر با صفر قائل می‌شوند.

تعیین حدنصاب قبولی

برای تعیین حدنصاب قبولی استفاده از روش زیر پیشنهاد می‌گردد:

- ۱- ابتدا کلیه گزینه‌هایی را که دارای نمره مثبت است بررسی کنید.
- ۲- گزینه‌هایی را که دانشجو با حداقل تسلط قابل قبول، حتماً باید انتخاب نماید جمع نمایید. این گزینه‌ها به نام مثبت‌های صرف نظر نکردنی^۱ خوانده می‌شود.
- ۳- کلیه گزینه‌هایی را که نمره منفی دارد مورد بررسی قرار دهید.
- ۴- مقابل گزینه‌هایی که دانشجوی لب مرز باید قادر باشد از مرئی کردن پاسخ آنها خودداری نماید علامتی قرار دهید. این گزینه‌ها به نام منفی‌های نابخشودنی خوانده می‌شود.
- ۵- بقیه گزینه‌های منفی در P.M.P منفی‌های بخشودنی^۲ است. یعنی انتظار نمی‌رود که دانشجوی با حداقل شایستگی از آنها اطلاع داشته باشد.

۶- با جمع جبری مثبت‌های صرف‌نظر نکردنی و منفی‌های بخشودنی حد نصاب قبولی تعیین می‌شود.

مزایا

- ۱- استفاده از این روش پایایی امتحان شما را برای سنجش توانایی تصمیم‌گیری بالینی بالا می‌برد.
- ۲- با این روش می‌توانید عکس‌العمل‌های عملی دانشجویان را بسنجید.
- ۳- این نوع آزمون به حد کافی روایی دارد؛ چون همان چیزی را از دانشجویان می‌خواهد که برای انجام آن تعلیم دیده‌اند.
- ۴- محدوده ضعف‌های بالینی دانشجویان برای شما مشخص می‌شود.
- ۵- با این روش می‌توانید زمینه‌های تشخیص، پیگیری و درمان را بسنجید.
- ۶- اگر طرح نمره‌دهی را پس از امتحان به دانشجویان بدهید، آنها قادر خواهند بود جواب درست را بیاموزند.
- ۷- این روش را هنگامی به کار ببرید که تعداد زیادی داوطلب در یک زمان و مکان دارید.

معایب

- ۱- این امتحان تا حدودی شانسی است.
- ۲- اطلاعاتی را که در هر مرحله می‌توانید بسنجید، محدود است.
- ۳- داوطلبان در تنظیم وقت برای قسمت‌های مختلف امتحان مشکل دارند.
- ۴- اگر دانشجویگزینه‌ای را اشتباه انتخاب کرد، آزمون‌گیرنده کاملاً آنها را می‌فهمد؛ چون دانشجو نمی‌تواند اشتباهات را بپوشاند. اما دانشجو فرصت تغییر جهت خواهد داشت هر چند نمره از او کسر می‌شود (بر خلاف حالت واقعی در بیمارستان).

منابع

- ۱- بهشتی، اطهره. شبیه‌سازی، تهران: انتشارات مرکز توسعه آموزش پزشکی، ۱۳۷۰.
- ۲- گیلبرت. راهنمای آموزش برای بهداشت‌کاران، ترجمه کیومرث ناصری، تهران: انتشارات دانشگاه تهران، ۱۳۶۳.
- 3- Abbatt FR. *Teaching For better learning*. 2nd ver.ed. Geneva: World Health Organization, 1992.
- 4- Beck DE. *Evaluating student performance in the experiential setting with confidence*. Am J Pharm Educ, 1995 fall; 59(2): 236-47.
- 5- Fabb WE, Marshall JR. *Assessment of clinical competence general family practice*. Edinburgh: MTP Press, 1983.
- 6- Nuefeld VR, Noramn GR. *Assessing clinical competence*. NewYork: Spriger Publishing Company, 1985.

فصل پنجم

آزمون‌های شفاهی

آزمونهای شفاهی

هر نوع آزمون را که در آن ارتباط مستقیم کلامی بین فرد آزمون‌گیرنده و دانشجو وجود داشته باشد، آزمون شفاهی گویند. در این نوع امتحان که یک یا دو آزمون‌گیرنده دانشجو را مورد مصاحبه قرار می‌دهند، از دانشجو در مورد دانسته‌هایش راجع به چند عنوان خاص سؤال شده، یا از وی خواسته می‌شود واکنشهای خویش را در برخی موقعیتهای خاصی که در جریان حرفه آینده او پیش خواهد آمد بیان کند. سپس دانشجو باید اطلاعات لازم را به خاطر آورده، معمولاً در وضعیتی اضطراب‌آور پاسخ دهد.

روش معمول آزمون شفاهی آن است که ابتدا آزمون‌گیرنده یک سؤال کلی مطرح می‌کند، سپس دانشجو باید به آن پاسخ دهد. این پاسخ ممکن است در وهله اول کامل بوده، پذیرفته شود؛ اما بیشتر اوقات آزمون‌گیرنده با توجه به پاسخ دانشجو سؤالات متعدد دیگری در زمینه همان سؤال نخست مطرح می‌کند تا جنبه‌های مختلف دانش وی را آزموده، سیر پاسخ دانشجو را به همان مسیر مورد نظر خود هدایت کند و متوجه شود که دانشجو تا چه حد قادر است به نتیجه‌گیریها و جمع‌بندیهای مورد نظر او برسد. در نهایت آزمون‌گیرنده راجع به کیفیت پاسخ دانشجو قضاوت کرده، نمره می‌دهد.

مزیت اصلی این نوع آزمون، نظارت مستقیم ممتحن بر آن است؛ یعنی او می‌تواند برحسب موقعیت زمانی و مکانی پدید آمده در هر لحظه، مسیر بحث را تعیین کند. بنابراین می‌تواند جزئی‌ترین دانسته‌های دانشجو را مورد سؤال قرار داده، با تجزیه و تحلیل خود، میزان اطلاعات وی را تخمین بزند.

به طور کلی امتحان شفاهی روش چندان مناسبی برای بخش حیطة دانشی افراد نیست؛ چون آزمون‌گیرنده هر چه سعی کند برخورد دوستانه‌ای داشته باشد باز هم دانشجو در حضور وی بشدت مضطرب می‌شود. این اضطراب هیچ استفاده‌ای برای دانشجو ندارد چون هرگز قرار نیست وی در شغل آینده خود با چنین نوعی از اضطراب و نگرانی روبرو شود. بلکه دانشجوی گروه پزشکی در آینده به جنبه‌هایی از امور درمان بیماران پرداخته، همه نگرانی او درباره سلامت و بهبودی بیمار خواهد بود و قرار نیست به صورت مداوم با امتحان

شفاهی مواجه گردد. در نتیجه، این نوع آزمون، در این موارد روایی لازم را که به مفهوم سنجش مهارتهای واقعی^۱ است، در برنخواهد داشت.

معمولاً نتایجی که دانشجویان در خلال آزمون شفاهی کسب می‌کنند بسیار بدتر از آن چیزی است که واقعاً استحقاق آن را دارند. به دلیل آنکه معمولاً به اهداف و محدودیتها و شیوه صحیح برگزاری آزمونهای شفاهی توجه کافی نمی‌گردد، در اغلب موارد استفاده از این روش بسیار وقت‌گیر بوده، پایایی چندانی هم ندارد؛ همچنین نه تنها مهارتهای مهم را ارزیابی نکرده، کمک چندانی به یادگیری دانشجویان نیز نمی‌کند بنابراین، سعی کنید آزمون شفاهی را به کار نگیرید مگر اینکه هدفی خاص داشته باشید که صرفاً از این راه قابل دستیابی است.

انواع آزمونهای شفاهی

آزمونهای شفاهی بر چهار نوع است که در اینجا مورد بحث قرار می‌گیرد.

۱- آزمون شفاهی میان دوره‌ای

این آزمون یک نوع آزمون میان دوره ساده است که در آن برای تک تک دانشجویان سؤالاتی مطرح می‌شود، بدین منظور که میزان پیشرفت و اشکالات آنها در طی دوره آموزش مشخص گردد. می‌توان سؤالات آزمون را در هر یک از سطوح شش‌گانه یادگیری طرح نمود.

۲- آزمون ارائه یک مورد بیماری^۲

شرح یک مورد به دانشجوی دوره بالینی ارائه شده، از او درخواست می‌شود آن را مطالعه نماید. سپس استاد راجع به آن مورد از دانشجو به طور شفاهی پرسش می‌کند.

۳- شبیه‌سازی شفاهی^۳

از دانشجو درخواست می‌شود با بیماری فرضی^۴، مصاحبه کرده، از او تاریخچه و شرح حال بگیرد و سپس اقدامات لازم جهت درمان بیمار را شرح دهد. استاد نیز بر این گفتگو نظارت می‌نماید. در این روش بیمار فرضی برای ارائه اطلاعات صحیح، قبلاً آموزش دیده است.

۴- حل مسائل بالینی

در این شیوه حداقل اطلاعات لازم راجع به یک بیمار واقعی به دانشجو داده می‌شود.

۱- مهارتهای واقعی یا مهارتهای مهم یعنی: آنچه دانشجو در جریان شغل آینده‌اش با آن روبرو می‌شود.

2- Case Presentation

3- Simulation

۴- بیمار فرضی (Simulator) فردی است که برای اجرای نقش بیمار آموزش دیده و بیمار واقعی نیست.

سپس دانشجو اقدام به طرح پرسشهای متعددی می‌نماید و سؤالاتی به طور متقابل بین استاد، دانشجو و بیمار طرح می‌شود؛ بدین معنی که هم استاد از دانشجو سؤال می‌کند تا او را به سیر منطقی هدایت کند و هم دانشجو راجع به نکاتی که به نظر او مبهم و در تشخیص نهایی مؤثر است پرسش می‌کند تا با کسب اطلاعات کافی به نتیجه نهایی برسد.

نوع دوم و چهارم تقریباً بر یک مبناست، با این تفاوت که در مورد دوم استاد یک مورد بیماری را برای دانشجو مطرح کرده، از او سؤالاتی راجع به هر جزء آن می‌پرسد و در نهایت از دانشجو تشخیص می‌خواهد. اما در مورد چهارم، دانشجو به یک موقعیت واقعی بر بالین بیمار موردنظر برده می‌شود تا با گرفتن شرح حال به تشخیص موردنظر برسد و استاد با نظارت بر این گفتگو و نیز با طرح سؤالات به موقع، مانع از به خطا رفتن دانشجو شده، او را به مسیر اصلی باز می‌گرداند. در مورد سوم همان اعمال مورد چهارم اتفاق می‌افتد با این تفاوت که در اینجا موقعیت واقعی نیست، یعنی فرد بیمار ایفای نقش کرده، کاملاً تعلیم دیده است که چه پاسخهایی را به دانشجو ارائه دهد.

قواعد طراحی آزمون شفاهی

برای برگزاری آزمون شفاهی باید به دنبال یک شیوه اخلاقی و منطقی بود. برگزاری این نوع آزمون بدون برنامه‌ریزی به نتیجه نرسیده، استاد و دانشجو را دچار سردرگمی عجیبی می‌کند، پس سعی کنید نکات زیر را رعایت نمایید:

۱- سؤالات امتحان شفاهی را قبل از شروع آزمون تهیه کنید؛ اگر بخواهید حین برگزاری امتحان با یک کتاب بزرگ سر جلسه نشسته و برای یافتن سؤال پیوسته آن را ورق بزنید، دانشجو احساس بدی خواهد کرد و تصور می‌کند این امتحان به حدی بی‌ارزش بوده که استاد حاضر نشده است برای طرح سؤالاتی آن وقت صرف کند. در عین حال بدین صورت از روایی و پایایی آزمون کاسته می‌شود؛ چون احتمالاً فی‌البداهه قادر نخواهید بود همه اهداف را به خاطر آورده، در مورد آنها سؤال طرح کنید.

۲- چک لیستی راجع به پاسخهای ضروری آماده کرده، سعی کنید فهرستی را از قبل در مورد عناوین مورد بحث در امتحان شفاهی تهیه نمایید و گرنه پس از برگزاری امتحان این احساس ناخوشایند به شما دست خواهد داد که هیچ کاری پیش نبرده، اطلاعاتی که کسب کرده‌اید آنقدر کم و ناچیز است که می‌توانید آنها را روی یک سر سوزن جای دهید!

۳- ضروری است که اطلاعاتی راجع به تواناییهای فرد در بخشهای دیگر امتحان نداشته باشید؛ یعنی اگر آزمون شفاهی بخشی از یک امتحان جامع است که در آن دانشجویان در یک آزمون چند مرحله‌ای مختلف شرکت می‌کنند از نمرات سایر مراحل بی‌اطلاع بمانید؛ چون نتایج آزمون شما را تحت تأثیر قرار می‌دهد.

۴- دانشجو باید احساس راحتی داشته باشد، زیرا در وضعیت اضطراب‌آور به خوبی نمی‌تواند پاسخ گوید؛ بنابراین از ایجاد چنین جوی پرهیزید. البته با مهارت و تجربه می‌توانید او را در آغاز آزمون شفاهی آرام کرده، کاری کنید که احساس مطلوبی به او دست دهد.

معمولاً چند لحظه بعد از شروع آزمون، فرد مورد امتحان در یک فاز شبه مانیک قرار گرفته، دچار یک طغیان فکری می‌گردد. بنابراین لازم است شما جریان ناگهانی کلماتی را که بر ذهن و زبان او جاری می‌شود کم کرده، دوباره بحث را به گرمی و آرامی آغاز کنید.

۵- برای شروع بحث در آزمون شفاهی پزشکی، معمولاً جنبه‌های مختلف بیماریها مورد سؤال قرار می‌گیرد. بنابراین دقت کنید که از انواع عجیب و غریب و نادر نمونه‌ها استفاده نکنید چون برخی از آنها واقعاً غیرقابل تشخیص است.

از طرفی اگر آزمون راجع به تصاویر رادیوگرافیک است باید از ابتدا منبع امتحان، ثابت و مشخص بوده، افراد مورد آزمون از چگونگی آماده شدن برای امتحان آگاه شوند و مقدمات لازم برای شرکت در چنین امتحانی را بدانند. در شیوه‌ای که فرد مورد آزمون یک عکس رادیوگرافی را برای شما می‌خواند اطلاعات بسیار جالبی راجع به دانسته‌های خود ارائه می‌دهد. بدین طریق حدسها و گمانهایی که فرد آنها را از جایی الهام گرفته براحتی آشکار خواهد شد.

۶- موضوعات مورد بحث را در آزمون شفاهی متنوع انتخاب کنید تا کمتر خسته و کسل شوید.

۷- از آزمون خود، به طریق ویدئویی فیلم‌برداری کنید. این کار به شما امکان می‌دهد تا ارزیابی خود را مورد بازبینی قرار دهید.

۸- هر چه تعداد افراد آزمون‌گیرنده و زمان اختصاص یافته به هر آزمون بیشتر باشد امتحان شما با ارزشتر خواهد بود؛ یعنی سعی کنید چند نفر با هم به طرح و اجرای آزمون بپردازید. البته وقتی چند آزمون‌گیرنده به کار می‌برید، آنها دریافته‌ها و آگاهیهای

- متفاوت دارند که باید پیش از شروع آزمون با هم به توافق برسند. همچنین عملکرد آزمون‌گیرنده‌ها می‌تواند رفتار دانشجو را متأثر کند.
- ۹- برای دانشجویان مختلف از آزمون‌گیرنده‌های متفاوت و سؤالات گوناگون استفاده نکنید، چون همه افراد مورد امتحان باید به سؤالات یکسان پاسخ گویند، در عین حال توسط آزمون‌گیرنده‌های یکسانی مورد پرسش قرار گیرند تا پایایی آزمون افزایش یابد.
- ۱۰- انتخاب آزمون‌گیرنده بسیار مهم است، این فرد باید کاملاً در زمینه موضوع امتحان مطلع و مجرب باشد.
- ۱۱- وقتی چند آزمون‌گیرنده در اجرای آزمون با هم همکاری می‌کنند به همه آنها نحوه نمره دادن یکسان را آموزش داده، نکات مهم و ضروری بحث را برای همه مشخص نمایید.
- ۱۲- اگر آزمون شفاهی به شیوه شبیه‌سازی است به فرد شبیه‌ساز (بیمار فرضی) نیز به دقت آموزش داده تا همان رفتارها و پاسخ‌هایی را ارائه دهد که مورد نظر شماست و دانشجو را به اشتباه نیندازد.
- ۱۳- پایایی امتحانات شفاهی پایین است اما با افزایش تعداد سؤال آزمون و برقراری نوعی توافق بین آزمون‌گیرنده‌ها در مورد کیفیت و استانداردهای امتحان (از طریق آموزش آنها) می‌توانید آن را افزایش دهید. روایی این آزمون‌ها از نظر در برگرفتن کل محتوای آزمون کم است، اما در عین حال برای درک میزان تواناییهای فردی از حیث برآورد عمق اطلاعات نظری فرد مناسب می‌باشد.
- ۱۴- با استفاده از شبیه‌سازی، روایی بسیار بالا خواهد رفت، چون رفتار دانشجو مشابه آن چیزی است که در واقعیت، حین برخورد با بیمار اتفاق می‌افتد.
- ۱۵- همزمان با برگزاری آزمون شفاهی باید گزارشی تهیه کنید که حاوی دلایل مستند برای نمره‌ای باشد که در مورد آن به توافق رسیده‌اید.
- ۱۶- کاملاً منطقی و اصولی است که اگر می‌خواهید مثلاً نحوه کاربرد «کاتتر» را از دانشجو سؤال کنید یک کاتتر در اختیار او گذاشته، از او بخواهید عملاً این فرآیند را به شما نشان دهد؛ زیرا این کاری است که او بعدها مجبور است بارها انجام دهد. پس بهتر است در این موارد به طراحی آزمونهای سنجش عملکرد اقدام نمایید.
- ۱۷- توصیه می‌کنیم همراه با چند همکار دیگر آزمون را برگزار کنید اما اگر همکار شما پیوسته پرسشهایی تکراری درباره موضوع یا موضوعات محدود طرح کند، خاطره‌ای ناخوشایند پدید می‌آورد. گر چه بدتر از همه آن است که مجبور باشید در کنار همکاری

برای طرح آزمون باشید که سؤالاتی را طرح کند که شما خودتان از فرد مورد آزمون قبلی پرسیده‌اید (یعنی سؤالات شما را تقلید می‌کند)!

۱۸- در آزمون شفاهی سعی نکنید با ذکر چند سؤال نظری و حفظی امتحان را هر چه سریعتر خاتمه بدهید وگرنه از هر لحاظ به استاندارد بودن امتحان خود لطمه وارد نموده‌اید.

مزایای آزمون شفاهی

امتحانات شفاهی در واقع دریچه‌ای رو به دانسته‌ها و توانایی‌های فرد به ویژه در سطوح تجزیه و تحلیل، ترکیب و ارزشیابی است و به منظور تشویق دانشجو به تفکر عمیق در پاسخگویی به یک مشکل خاص به کار می‌رود. در این نوع امتحان، امکان حدس پاسخ بسیار ناچیز بوده، معمولاً شرایط آنقدر توسط یک آزمون‌گیرنده مجرب کنترل می‌شود که توان گریز از پاسخ‌دهی وجود ندارد چون وی بلافاصله حیلۀ دانشجو را نقش بر آب می‌کند. فایده دیگر این نوع آزمون؛ سهولت طرح سؤال است. به علاوه آزمون‌گیرنده کنترل کلی و کاملی بر سراسر روند آزمون دارد.

معایب آزمونهای شفاهی:

- ۱- یکی از اشکالات سؤالات شفاهی طولانی بودن زمانی است که برای امتحان از هر دانشجو لازم است. معمولاً مدت زمان مناسب برای امتحان از هر فرد سی دقیقه است.
- ۲- محدودیت زمانی باعث محدود شدن تعداد سؤالاتی می‌شود که از هر دانشجو پرسیده می‌شود؛ در نتیجه عناوین قابل سؤال را نیز محدود می‌کند. این محدودیت روایی این آزمونها را زیر سؤال می‌برد.
- ۳- برخلاف امتحانات کتبی، پاسخ دانشجو بلافاصله از دست می‌رود، مگر اینکه روی نوار ویدئویی ضبط گردد.
- ۴- خصوصیات شخصی دانشجو مثل لحن صدا، خصوصیات بدنی، فردی و اعتماد به نفس، نتایج آزمون را به شکل یک اثر هاله‌ای متأثر می‌کند.
- ۵- اضطراب باعث می‌شود که دانشجو به طور ضعیف پاسخ داده، نمره پایین‌تری نسبت به آنچه واقعاً استحقاق دارد، بگیرد.
- ۶- امکان تفسیر ناصحیح پاسخ دانشجو از سوی استاد وجود دارد. در ضمن ممکن است استاد نتواند دقیقاً بیان کند چه چیزی از دانشجو انتظار دارد؛ یعنی هر کدام از استاد و

دانشجو ممکن است منظور خود را خوب ارائه نکنند.

۷- به دلیل اختلاف نظر آزمون‌گیرندگان و امکان انتقال نیافتن صحیح مفاهیم، در اغلب موارد پایایی این آزمون‌ها نیز به حد کافی نیست. البته می‌توان با استفاده از فهرستهای بازبینی، استفاده از پرسشهای سازمان یافته و آموزش پرسشگران تا حدی بر این محدودیت غلبه کرد.

چند نکته در برگزاری آزمونهای شفاهی

- آزمون‌گیرنده ماهر و خوب بودن یک هنر بوده، تا حدودی به استعداد انسانها برمی‌گردد. اما هرگز نباید نقش تجربه را در این زمینه فراموش کرد. همیشه اجرای یک آزمون مثل دانشنامه تخصصی پر از هول و هراس است و آزمون شونده‌گان از شدت اضطراب نمی‌دانند چه بکنند چون سرنوشت و آینده آنها را همین امتحان روشن می‌کند.
- ممکن است به آسانی فراموش نمایید که روزی خودتان در چنین وضعیتی اضطراب‌آوری بوده‌اید و آزمون‌شونده را نادیده بگیرید، اما اگر به عدالت رفتار کنید باید وضعیتی فرد مورد آزمون را هم مد نظر قرار دهید. به خصوص در آزمونهای دانشنامه تخصصی، شکست در آزمون باعث از دست رفتن آسایش و اعتماد به نفس می‌شود.
- آزمون‌گیرنده خوب باید مؤدب، باهوش و نکته‌سنج باشد. برای فردی که در آزمون دانشنامه تخصصی شرکت می‌کند بزرگترین بداقبالی آن است که گرفتار آزمون‌گیرنده‌ای شود که هرگز حاضر نیست خودش را در موقعیت آزمون شونده قرار دهد! و از اینکه وی را مثل یک متهم مورد سؤال قرار دهد لذت می‌برد. پس سعی کنید ابتدا محیط آزمون را به یک محیط صمیمانه و بدون هراس تبدیل نمایید.
- چون بیماران حوصله و توان آن را ندارند که با شما در برگزاری یک آزمون دانشنامه تخصصی همکاری کنند، اغلب اوقات از بیماران فرضی استفاده می‌شود. معمولاً از کارکنان سازمان متبوع به عنوان بیمار برای شبیه‌سازی استفاده می‌شود. پس ابتدا آنها را به خوبی آموزش داده، در پایان کار از آنها تشکر کنید.
- بهتر است مسئول برگزاری، آزمون دانشنامه را در یک شهر دیگر غیر از شهر خود برگزار کند، چون هم خیلی راحت‌تر و دوست‌داشتنی‌تر است و هم از مسؤولیتهای شفلی عادی خود رها شده، از امکانات میهمانی هم استفاده خواهد کرد.
- حتماً پیش از شروع آزمون به محل رفته، مدتی با بیمار یا بیماران فرضی گفتگو کنید. در

این صورت کلیه مطالب راجع به بیماریهای طرح شده تحت کنترل شما بوده، به راحتی می‌توانید قابلیت کاری فرد مورد آزمون را اندازه بگیرید.

● ممکن است مطالبی به بیمار فرضی آموخته شده باشد که شما بشدت با آن مخالف هستید. در چنین حالتی سعی کنید خودتان را کنترل کرده، تا حد امکان این مطالب را وارد بحث خود نکنید.

● پرداختن یک روز کامل به برگزاری امتحان باعث خستگی می‌شود؛ پس بهتر است صبح یا عصر در آزمون شرکت کرده، بقیه روز را استراحت کنید. سعی نمایید آزمون را برای یک روز عادی برنامه‌ریزی کنید.

● وقتی سؤال را طرح می‌کنید به آزمون شونده توضیح دهید که پاسخش باید چه اجزایی را در برداشته باشد.

حتماً یک الگوی پاسخ برای بحث خود با آزمون‌شونده تهیه کنید؛ چون بعداً باید بتوانید با سایر آزمون‌گیرنده‌ها راجع به آن بحث کنید. قبل از شروع آزمون شفاهی دانشنامه تخصصی، کلیه برگه‌های کتبی را تصحیح کرده، علت نمره داده شده را ذکر نمایید؛ بخصوص این امر به افراد متوسط کمک می‌کند تا در آزمون شفاهی با تکرار این خطاها، باعث شکست خود نشوند. پس از پایان آزمون، تمام آزمون‌گیرنده‌ها باید با هم به مشورت نشست، نمره کل را ارائه کنند.

● یادداشتهایی که در خلال برگزاری آزمون نوشته‌اید در قضاوت شما بسیار مؤثر است؛ بخصوص ممکن است در میان آنها نکات مثبتی برای کمک به نمره افراد بیابید. اگر فردی را در آزمون رد می‌کنید دلیل رد نمودن او را برایش توضیح دهید.

● در چنین آزمونی مشخص کردن نمره قبولی بسیار مشکل بوده، احتیاج به مشورت همه اعضای برگزارکننده آزمون دارد؛ پس در انتها از تک‌تک افراد همکار بخواهید گزارش کتبی از کار خود ارائه کرده، موشکافانه به بررسی وضعیت اجرای آزمون بپردازند؛ این کار به برگزاری آزمونهای دوره بعد نیز کمک خواهد کرد.

منابع:

- ۱- حوزه معاونت آموزشی توسعه آموزش پزشکی مرکز اصفهان. *آزمون شفاهی*. اصفهان: ۱۳۶۹.
- 2- Abbatt FR. *Teaching for better learning*. 2nd ver.ed. Geneva: World Health Organization, 1992.
- 3- Allison R. *Audit of oral examination in psychiatry*. Med Teach 1992; 14 (4): 383-390.
- 4- Anastakins DJ, Conen R, Reznick RK. *The structured oral examination as a method for assessing surgical residents*. Am J Surg 1991 Jul; 162(1): 67-70.
- 5- Burchord KW, Rowland-morin PA, Coe NPW, Carb JL. *A surgery oral examination: Intrater agreement and influence of rater characteristics*. Acad Med 1995 11; 70 (11): 1044-6.
- 6- Goodman LJ, Bruescnke EE, Bone RC, Rose WH, Williams EJ, Paul HA. *An experiment in medical education. A critical analysis using traditional criteria*. JAMA 1991 May 8; 265(18): 2376-6.
- 7- Jayawickramarajah PT. *Oral examination in medical education*. Med Educ. 1985; 19: 290-293.
- 8- Moor R. *The oral examination*. Practitioner. 1992 Mar 3; 236.

فصل ششم

امتحانات میان دوره

امتحانات میان دوره ۱

امتحان میان دوره به نوع خاصی از سؤال محدود نمی‌شود بلکه شامل کلیه آزمون‌هایی است که در حین تدریس یک دوره آموزشی برگزار می‌گردد. پس این نوع آزمون در واقع نوعی امتحان تکوینی (آزمون در خلال دوره آموزش) است.

هدف اصلی از برگزاری آزمون میان دوره، ارائه نوعی بازخورد به دانشجویان تا وضعیت فعلی او را روشن کند و وی را راهنمایی نماید که چگونه نواقص خود را برطرف سازد. دانشجویان دارای سطوح متفاوتی از اطلاعات هستند و هرگز نمی‌توان انتظار داشت که میزان یادگیری دانشجویان مختلف از یک موضوع درسی واحد به یک اندازه باشد. از آنجا که شما استاد همه دانشجویان هستید نه فقط استاد دانشجویان با استعداد و باهوش، عدالت حکم می‌کند سؤالاتی را در آزمون‌های نهایی دوره طرح کنید که دانشجویان متوسط و ضعیف هم توان پاسخگویی به آنها را داشته باشند.

همچنین اگر مایل هستید توانایی‌های عالی دانشجویان را در سطوح بالای یادگیری صرفاً به منظور اطلاع از توانایی بالقوه دانشجویان موردسنجش قرار دهید ابزار مناسب، آزمون‌های میان دوره می‌باشد.

هدف اصلی از برگزاری آزمون‌های میان دوره به صدا در آوردن زنگ خطر برای دانشجویان کم تلاش است. از نتایج این آزمون‌ها به عنوان تنها ملاک تصمیم‌گیری نهایی راجع به دانشجویان استفاده نکنید، چون هدف از برگزاری آن را زیر سؤال می‌برد. اگر می‌خواهید نمرات آزمون میان دوره را به عنوان بخشی از نمره نهایی دانشجویان در نظر بگیرید این امتحان در واقع بخشی از آزمون پایان دوره شماست که فقط از نظر زمانی با آن فاصله دارد. مثل این است که آزمون پایان دوره را به قطعاتی تقسیم کرده، هر بخش را در زمانی اجرا کنید. در این صورت این آزمون را دیگر نمی‌توان «میان دوره» نامید.

البته دانشجویان باید قبلاً از قصد شما برای برگزاری آزمون میان دوره مطلع بوده، زمان دقیق آن را بدانند. در غیر این صورت این آزمون به عنوان وسیله‌ای تهدیدآمیز در ذهن دانشجویان

تداعی می‌شود که استاد هر از چندگاه برای آزار او و قدرت‌نمایی، به آن متوسل می‌شود. از سوی دیگر اکثراً دانشجویانی که در امتحانات پایان نیمسال مردود می‌شوند معتقدند که درس را خوب فرا گرفته‌اند و از شکست خود اظهار تعجب می‌کنند. اما آیا واقعاً این دانشجویان راست می‌گویند یا اینکه برای توجیه خود چنین استدلالی می‌نمایند؟ یکی از فواید آزمونهای میان‌دوره آن است که از طریق نتایج آن می‌توانید به دانشجویان نشان دهید آیا شکست نهایی او به حق بوده است یا خیر. زیرا دانشجویی که در تمام آزمونهای میان‌دوره نمره ضعیفی کسب کرده و در امتحان نهایی نیز با شکست مواجه شده است طبعاً نمی‌تواند ادعا کند که در همه آنها با بدشانسی روبرو بوده است. برای برگزاری آزمون میان‌دوره می‌توانید از هر دو شیوه کتبی و مشاهده‌ای (عملی) استفاده کنید.

الف) شیوه مشاهده‌ای

در طی یک دوره آموزش، دانشجویان ممکن است زمانی را به کار در مراکز بهداشتی یا بیمارستانها صرف کنند. در این مکانها به فراگیری مهارتهای برقراری ارتباط و مهارتهای روانی - رفتاری مورد نیاز در شغل آینده خود می‌پردازند. از این دوره زمانی می‌توانید هم برای تدریس و هم ارزیابی‌های میان‌دوره‌ای که از طریق مشاهده رفتار دانشجو در مراکز درمانی انجام می‌شود، استفاده کنید. در این شیوه افراد متفاوتی می‌توانند دانشجو را مورد ارزیابی قرار داده، راجع به نحوه عملکرد به او توضیحاتی بدهند. چنین کاری اگر همراه با چکلیست انجام شود بسیار مؤثرتر خواهد بود.

ب) شیوه کتبی

شیوه‌ای که در برگزاری این آزمون انتخاب می‌شود ممکن است یکی از روشهای آزمون کتبی باشد، مانند: آزمون چندگزینه‌ای، آزمون صحیح - غلط، تشریحی و شیوه‌ای که برمی‌گزینید به وقت شما و هدفی که دنبال می‌کنید بستگی دارد، ولی اکثر استادان برای راحتی کار به تعداد محدودی سؤال چندگزینه‌ای اکتفا می‌کنند. برای بررسی نحوه تأثیر آزمونهای میان‌دوره بر دانشجویان در مطالعات درسی بد نیست به تحقیق زیر اشاره کنیم.

ینکین^۱ در یک پژوهش بسیار ساده تصمیم گرفت نشان دهد که آگاهی دانشجویان از میزان دانسته‌هایشان پیش از برگزاری امتحان پایان نیمسال چه تأثیری بر نتیجه نهایی آنها خواهد داشت.

برای گروهی از دانشجویان که دوره زنان و زایمان^۲ را می‌گذراندند امتحان چندگزینه‌ای معمول برگزار شد. در این آزمون شصت دانشجو در یک امتحان یک ساعته حاوی ۲۵ سؤال چندگزینه‌ای راجع به مبحث مذکور شرکت کردند. آنها از بین دانشجویان سال آخر پزشکی برگزیده شدند که در دوره شش هفته‌ای زنان شرکت کرده بودند.

وی پس از برگزاری امتحان با تک تک دانشجویان به طور جداگانه مصاحبه‌ای ترتیب داد و راجع به نحوه پاسخ گفتن به سؤالات، اطلاعاتی در اختیار آنها گذاشت. پس از آن به هر فرد گفت که هیأت آزمون‌گیرنده او را در کدامیک از زیرشاخه‌های رشته زنان و زایمان دارای ضعف علمی می‌داند. به افرادی که نمره پایین گرفته بودند توصیه کرد موضوع خاصی را بیشتر بخوانند و گرنه باید خطر شکست در امتحان پایان دوره را بپذیرند. همچنین افرادی که نمره خوبی گرفته بودند تشویق به ادامه کار خود شدند. در نتیجه کار هیچکدام از افرادی که در آزمون موفق گردیدند در آزمون نهایی دوره شکست نخورده، حتی جزء نمره‌های مشروط هم نشدند. بنابراین می‌توان گفت آزمون میان دوره روایی پیش‌بین^۳ خوبی دارد.

از طرفی نیمی از افرادی که در آزمون میان دوره نمرات ۴۵ درصد تا ۵۵ درصد را اخذ کرده بودند در آزمون نهایی نمرات بالاتر از ۵۵ درصد گرفتند. کلیه افرادی که در آزمون نهایی شکست خوردند کسانی بودند که در آزمون میان دوره نمرات زیر ۲۰ درصد گرفته بودند.

از مطالب مذکور می‌توان نتیجه گرفت حتی در صورتی که یک آزمون چندگزینه‌ای نه چندان خوب و مطلوب در طول دوره برگزار کنیم اما بعد براساس نتایج آن به دانشجویان توصیه‌هایی ارائه کنیم و در مورد نحوه مطالعه آنها را راهنمایی نماییم، این کار تأثیر خوبی بر جای خواهد گذاشت و به برخی افراد کمک بالقوه‌ای خواهد کرد.

با این مطالعه مشخص گردید افراد ضعیف شناخته شده در این امتحان کسانی بودند که در تمام طول تحصیل شاگردانی ضعیف و کم‌کار محسوب می‌شوند. بنابراین

1- David A Jenking

2- Obstetrics & Gynecology

۳- Predictive Validity: به قابلیت آزمون برای پیش‌بینی عملکرد خاصی اطلاق می‌گردد.

شاید نتیجه این آزمون شناسایی دانشجویان کاملاً ضعیف بوده که در این صورت شاهکاری نکرده‌ایم؛ چون آنها خود انگشت‌نما و شناخته شده بوده‌اند، احتیاج به برگزاری آزمون برای شناسایی آنها نیست. از طرفی، فقط اکتفا کردن به چند توصیه هم معمولاً برای چنین دانشجویانی بیهوده است و شاید لازم باشد توصیه را با کمی جبر و زور همراه کرد. اما در هر حال نیمی از افرادی که به آنها توصیه‌های هشداردهنده داده شد، در امتحان نهایی قبول شدند. پس این آزمون چندان هم به خطا نرفته است و حداقل با این روش ساده، استاد می‌تواند ۵۰ درصد دانشجویان در معرض خطر را نجات دهد، به خصوص برگزاری چنین آزمونهایی در ارتقای دانشجویان متوسط بسیار مؤثر است.

با برگزاری آزمونهایی میان دوره بسیار ساده که جلسات مشورتی به دنبال داشته باشد، می‌توانید دانشجویان ضعیف را شناسایی کرده، ۵۰ درصد از میزان مردود شدن^۱ کلاس خود را کم کنید که هم برای دانشجو خیلی مفید است و هم باعث امیدواری و علاقه‌مندی خود شما خواهد شد؛ گرچه این توصیه نمی‌تواند به همه کمک کند.

از سوی دیگر در مطالعاتی که در دانشکده‌های مختلف به عمل آمده، بیشتر دانشجویان معتقدند امتحانهایی که به طور معمول در خلال دوره آموزشی برگزار می‌شود، برای یادگیری آنها بسیار مفید است. در مطالعه‌ای که توسط کراوز^۲ بین دانشجویان داروسازی (دیدگاههای ۱۸۶ دانشجو) راجع به امتحانهای معمول، ارزش آنها در برگزاری آزمون پایان دوره و تناوب آنها انجام شد اکثر آنها معتقد بودند آزمونهایی کلاسی باید به طور مکرر برگزار شود. حتی بسیاری از آنها براین باور بودند که باید بخشی از نمره نهایی، به نمرات مربوط به آزمونهایی میان دوره اختصاص یابد.

معمولاً نمرات آزمونهایی پایان دوره، چه نمرات عملی و چه نظری بالاتر از نمرات آزمونهایی میان دوره است. ولی اغلب می‌توان بین نمرات دو دسته آزمون پایان دوره و میان دوره همبستگی پیدا کرد. اما در صورتی که آزمون پایان دوره نیز به صورت نظری باشد همبستگی قویتر خواهد بود. این ضریب همبستگی برای آزمونهایی پایان دوره نظری حدود $r = 0/93$ و برای آزمونهایی پایان دوره عملی حدود $r = 0/71$ به دست آمده است که همبستگی قوی و قابل تأملی را بین آزمونهایی میان دوره و پایان دوره نشان می‌دهد.

منابع:

- 1- Eadra D. *Weightage of formative examinations in certifying examination of pharmacology: An opinion poll and relative performance of learners.* Med Teach 1992; 14(213): 197-200.
- 2- Hill DA. *Formative assessment: A student spective.* Med Educ 1994; 28: 394-399.
- 3- Jenkins DM. *Dose incourse assessment and counselling improve exit performance?* Med Teach 1992; 14(2/3): 241-2.
- 4- Mehrens WA. *Measurement and evaluation in education and psychology.* 3rd rev. ed. New York: Holt Saunders, 1984.
- 5- Robins I. *The effect of pass/fail grading and weekly quizzes on first- year student's performance and satisfaction.* Acad Med 1995; 70 (4): 327-29.

فصل هفتم

خودآزمایی

خود آزمایی^۱

بیشتر انسانها علاقه دارند میزان دانش خود را پنهان از دیگران محک زده، نقصهای آن را دریابند بدون آنکه خود را در معرض ارزیابی افراد دیگر قرار گیرند. اگر قرار باشد تنها وسیله سنجش اطلاعات و آموخته‌های روزانه افراد آزمونهایی باشد که مسئولان دانشکده یا وزارتخانه هر چند وقت یک بار برگزار می‌کنند، شکی نیست که باید از طیف وسیعی از دانسته‌های خود چشم پوشی کرده، فقط مطالبی بیاموزیم که در هر آزمون خاص مورد سؤال قرار می‌گیرد. در عین حال باید به طور مداوم در آزمونهایی متعدد شرکت کنیم که جز سردرگمی چیزی به بار نخواهد آورد. خود آزمایی کلید حل این معما هستند، البته به شرط آنکه به صورت استاندارد و منطقی اجرا شود.

این رویکرد، آزمونی است که هر فرد برای سنجش وضعیت و پیشینه معلومات خود به کار می‌برد. شکل ساده آن سؤالاتی است که در پایان فصلهای کتابهای درسی به صورت گزیده پاسخ یا باز پاسخ وجود دارد. هنگامی که حین خواندن یک مطلب، سؤالاتی در ذهن خود طرح کرده، به آنها پاسخ می‌گویید و یا جمله ساده «پس اینجا چه می‌خواست بگوید؟» را بیان کرده، نکات مهم متن را برای خود تعریف می‌کنید در حقیقت درگیر نوعی خود آزمایی هستید.

در نوع مرسوم خود آزمایی طراح سؤال استاد درس است اما دانشجو در پاسخ دادن به آن مختار است. این شکل خود آزمایی را می‌توان به صورتهای مختلف و برای پیگیری اهداف متفاوتی به کار برد.

در خود آزمایی استاد بخشی از وقت آموزش را به ارزیابی اختصاص نمی‌دهد بلکه دانشجو باید از وقت خود استفاده کند. در ضمن ملزم به تصحیح ورقه‌ها نیز نخواهید بود، پس بسیار با صرفه است.

انواع آزمونهایی خود آزمایی و کاربردهای آن

شیوه‌ای که برای خود آزمایی به کار می‌رود مثل هر آزمون دیگر به هدف اجرای

آزمون بستگی دارد. شیوه‌های عمده و اساسی خودآزمایی عبارتند از:

۱- ارائه جمله هدف درسی به صورت پرسشی؛

۲- طرح سؤال آزمون در مورد تک تک اهداف.

۱- ارائه جمله هدف درس به صورت پرسشی

در این شیوه، از اهداف طراحی شده برای تدریس استفاده می‌شود. البته به شرط آنکه اهداف، با دقت و طبق اصول طرح شده باشد. اگر طراح هدف، فقط مسائل کلی را انتخاب کرده، دقت چندانی صرف نوشتن سایر اهداف نکرده باشد، باید از این روش چشم‌پوشی کرد.

اکثر افراد، از ارزش یک هدف‌نویسی اساسی در منسجم کردن فرآیند تدریس اطلاع ندارند. به زعم آنان، هدف، صرفاً جملاتی است که آنها را باید روی یک کاغذ یا تابلو شفاف^۱ نوشته، در جلسه اول تدریس که تعداد انگشت‌شماری دانشجو سرکلاس حاضر می‌شوند و همین تعداد اندک هم حوصله تدریس استاد را ندارند، آن را به دانشجویان نشان داد.

در چنین اوضاعی دانشجویان به دلیل ناآشنایی با درس، معنی جملات استاد را متوجه نمی‌شوند و جلسه اول کلاس با بهت و تعجب آنان تمام خواهد شد. البته چنین استادانی به فکر برگزاری خودآزمایی برای دانشجویان نخواهند افتاد.

بنابراین آنچه برای ارزیابی لازم است اهدافی دقیق، ملموس، منطقی، دست‌یافتنی و قابل سنجش است. این اهداف باید در تمام طول تدریس به نحوی کاربرد داشته باشد که هم دانشجو از جمله هدف، دقیقاً برداشت کند که پس از پایان تدریس چه عملی را به طور عینی از او انتظار دارند و هم استاد بداند چگونه می‌توان راجع به آن، سؤال طرح نموده، به راحتی نحوه ارزیابی هر هدف خاص را حدس زد.

بدیهی است که جملات تخیلی و توصیفی، مردم پسندتر است، ولی هرگز قابل دستیابی و ارزیابی نیست. پس لازمه هر آزمون، اهداف عینی است.

هر دوره درسی دارای اهداف جزئی بی‌شماری است که هرگز امکان ندارد کسی بتواند از همه آنها در یک آزمون، سؤال طرح کند، اما در صورتی که این اهداف، رؤیایی نباشد، حتماً قابل اندازه‌گیری خواهد بود. معمولاً، برای ارزیابی چنین اهدافی از خودآزمایی استفاده می‌شود.

در اجرای این شیوه، ابتدا بهتر است بار دیگر اهداف تعیین شده در ابتدای دوره درسی بازبینی شود؛ زیرا ممکن است در روند تدریس و در موقعیت‌هایی پیش‌بینی نشده،

استاد مجبور به تدریس مطالبی شده باشد که قبلاً انتظار آن را نداشته است به عنوان مثال دانشجویان سؤالاتی طرح کرده‌اند که شما را ملزم به تدریس مطالب خاصی نموده است. باید این مطالب را به فهرست اهداف خود اضافه نموده، بعد از اطمینان از وجود همه اهداف در فهرست مزبور به طرح سؤال پردازید. در اینجا صورت جمله هدف به صورت سؤال مطرح شده، پنج مقیاس (مانند کاملاً می‌توانم یا کاملاً می‌دانم، می‌توانم، مطمئن نیستم، نمی‌توانم، اصلاً نمی‌توانم) در مقابل آن قرار می‌گیرد. به مثال زیر توجه کنید:

هدف: دانشجو باید بتواند اجزای سیستم تحریکی - هدایتی قلب را طبق کتاب فیزیولوژی گایتون نام ببرد.

سؤال آزمون خودآزمایی: آیا می‌توانید اجزای سیستم تحریکی - هدایتی قلب را طبق کتاب گایتون نام ببرید؟

کاملاً می‌توانم □ می‌توانم □ مطمئن نیستم □ نمی‌توانم □ اصلاً نمی‌توانم □

پس از اینکه به همین طریق از تک تک اهداف سؤال طرح نمودید سؤالات را برای پاسخ‌گویی در اختیار دانشجویان قرار دهید.

در ابتدای برگه سؤالات جملاتی به کار ببرید که دانشجو را در پاسخ دادن به سؤالات ترغیب کند. هرگز در ابتدای برگه، جملات امری بکار نبرید، چون این امتحان، یک آزمون اختیاری است و اگر دانشجو را امر به پاسخ‌گویی نمایید مطمئن باشید سؤالات شما بی‌جواب خواهد ماند.

از دانشجویان بخواهید که در مورد هر سؤال با دقت و حوصله تصمیم بگیرند؛ به آنها تذکر بدهید که این فهرست در بردارنده کلیه اهداف شما از تدریس است و در صورتی دانشجو مبحث را کاملاً فرا گرفته که بتواند به کلیه آنها پاسخ گوید (در واقع به کلیه این اهداف رسیده باشد).

برای دانشجو روشن نمایید که اگر گزینه‌های «مطمئن نیستم، نمی‌توانم یا اصلاً نمی‌توانم» را انتخاب کرده، باید نسبت به یادگیری مجدد بخش خاصی از درس که در حیطه آن سؤال است اقدام کرده، راجع به دانسته‌های خود تجدید نظر نماید.

این شیوه علاوه بر آنکه نوعی بازخورد در اختیار دانشجو قرار می‌دهد - یعنی کلیه ضعفهای اطلاعاتی دانشجو را در یک مقوله خاص برای او روشن می‌کند تا با برطرف کردن آنها مطلب را به طور کامل فرا گیرد - وسیله‌ای برای ارزیابی کار استاد توسط خود او نیز به حساب می‌آید. بدین معنی که اگر اکثر دانشجویان یک یا تعدادی سؤال خاص مربوط به هم

را در مورد یکی از عناوین درس «نمی‌توانم» یا «اصلاً نمی‌توانم» پاسخ داده‌اند، به احتمال زیاد اشکال در تدریس استاد بوده است. بنابراین، مدرس می‌تواند آن مبحث خاص را تکرار کند، یا علت نفهمیدن دانشجویان را ریشه‌یابی کرده، آن را برطرف نماید.

البته به این شیوه چندان نمی‌توان اعتماد کرد، چون معمولاً زمان برگزاری چنین آزمونهایی پس از پایان دوره تحصیلی و برگزاری آزمونهای پایان نیمسال است. اغلب، این سؤالات توسط استاد در پایان جلسه امتحان به دانشجو داده می‌شود، که نتیجه کار از قبل معلوم است. به دانشجویی که به زحمت امتحان را گذرانده، از سؤالات خلاصی یافته است حق بدهید که پس از خروج از جلسه امتحان، برگه‌های سؤالات خودآزمایی را در راهروهای دانشکده رها کرده، یا از آنها موشک کاغذی بسازد، شاید هم لابلای کتاب و جزوات درس موردنظر به فراموشی سپرده شود.

اما اگر تصمیم بگیرید که دانشجو، سؤالات را سر جلسه امتحان پاسخ گوید بدتر خواهد شد. چون بیشتر دانشجویان، برای خروج هر چه سریعتر از جلسه بدون خواندن متن سؤالات جوابها را به صورت احتمالی و براساس شانس، تکمیل خواهند کرد.

شاید برگزاری برنامه خودآزمایی در خلال دوره درسی، در مورد اهدافی که تا آن زمان تدریس شده است شیوه بهتری باشد، در این شیوه مباحثی را که اکثر دانشجویان در یادگیری آن با اشکال روبرو شده‌اند باید مجدداً تدریس کنید.

ضمناً باید عواقب این کار را هم در نظر بگیرید؛ چون اگر دانشجویان زیرک از هدف شما آگاه شوند همه جوابها یا اکثر آنها را «نمی‌توانم» انتخاب کرده، شما را به تنگنا می‌اندازند و اگر حوصله تدریس شما را نداشته باشند، همه با هم متحد شده، تمام جوابها را «کاملاً می‌توانم» انتخاب می‌کنند.

«خودآزمایی» را باید تنها به عنوان وسیله‌ای برای تهیه بازخورد به دانشجو در نظر گرفت و اگر در پی استفاده از نتایج آن در برنامه‌ریزی تدریس خود هستید، باید مطمئن شوید که دانشجویان شما در قضاوت، عادل بوده، به موضوع درس علاقه‌مندند.

البته گاهی با ایجاد موقعیتهای خاص و کاربرد یکی، دو ترفند می‌توان از این شیوه به عنوان روشی بی‌مانند برای ارزیابی تدریس استفاده کرد. به تجربه و کیاست شما بستگی دارد که چگونه همکاری دانشجویان زیرک را جلب نمایید.

۲- طرح سؤال آزمون در مورد تک تک اهداف

بهتر است در این مورد از سؤالات چند گزینه‌ای استفاده شود. این کار پیچیده‌تر از

روش اول خواهد بود؛ چون باید سوالات استاندارد و بر طبق قاعده طرح سؤال باشد و اجرای این مقصود در مورد تک تک اهداف ریز و درشت کمی وقت گیر است.
مثال:

هدف: دانشجو باید بتواند اجزای سیستم تحریکی - هدایتی قلب را طبق کتاب فیزیولوژی گایتون نام ببرد.

سؤال آزمون خودآزمایی: همه اجزای زیر جزء سیستم تحریکی - هدایتی قلب هستند بجز:

الف) گره سینوسی دهلیزی ب) ماهیچه پاپیلاری

ج) گره دهلیزی بطنی د) دسته هیس

در این شیوه نیز به دانشجو بازخورد داده می شود؛ یعنی مجدداً باید برای دانشجو روشن کنید که تنها وقتی وی به تمام سوالات پاسخ صحیح داده باشد، به اهداف مورد نظر درس دست یافته است.

به ایشان متذکر شوید که پاسخ ندادن به هر سؤال، دلیل یک خلاء^۱ اطلاعاتی است؛ حتی اگر او آزمون پایان دوره را با نمره خوبی گذرانده شده باشد.

این شیوه ها فقط در مورد دانشجویان فعال که واقعاً برای دانسته های خود ارزش قائلند سودمند بوده، دانشجویان ضعیفی که با زحمت بسیار، موفق به کسب نمره ای در حد مرز قبولی شده اند هرگز در صدد رفع کمبودهای خود بر نمی آیند.

خودآزمایی و برآورد دانشجویان

آنچه امروز بیش از پیش توجه صاحب نظران را به خود جلب کرده، اهمیت خودآزمایی دقیق در کسب مهارت های بالینی است. دانشجو به عنوان پزشک آینده باید بتواند آنچه را بر بالین بیمار انجام می دهد ارزیابی کند؛ اگر او در تشخیص خود بیش از حد از حدس و گمان استفاده کند باعث تشخیص اشتباه و بسته شدن زود هنگام پرونده بیمار می شود و اگر توان حدس و تخمین چندانی نداشته و صرفاً متکی به آزمایش های تشخیصی باشد به حس شک و ترغیب و همچنین ارجاع های بی مورد در سیستم خدمات بهداشتی، درمانی منجر خواهد شد. پس دانشجوی پزشکی باید توانایی سنجش خود را داشته باشد.

روش های معمول خودآزمایی، روش هایی است که دانشجو را با سایر سنجشها مثل

آزمون توسط استاد یا آزمونهای ارزیابی همتا^۱ مرتبط می‌کند.

نکته‌ای که همیشه آزمونهای خودآزمایی را زیر سؤال می‌برد آن است که دانشجو تا چه حد در خودآزمایی خود، به حق رفتار کرده، چه اندازه در این کار قابلیت دارد. تحقیقاتی که درباره خودآزمایی از سال ۱۹۷۰ به بعد انجام شد بیشتر مؤید آن بود که اغلب دانشجویان تمایل دارند خود را در رده‌های پایین‌تری نسبت به آنچه واقعاً استحقاق دارند، قرار دهند. البته گاهی اوقات نیز نتیجه برعکس است. هتلا^۲ استوارت^۳ در سال ۱۹۸۰ نشان داد وقتی از دستیاران دانشکده پزشکی خواسته می‌شود از روی نوارهای ویدئویی، مصاحبه خود را با بیماران ارزیابی کنند، نمرات خیلی پایین‌تری نسبت به نمره واقعی (که توسط استاد به آنها داده شده) به خود اختصاص می‌دهند، اما در تحقیقی مشابه، دانشجویان سال سوم پزشکی نمراتی به خود می‌دادند که به طرز جالب توجهی بالاتر از میزانی بود که استحقاق داشتند (پیتر^۳ ۱۹۹۰).

نه تنها گروههای متفاوت از نظر رده تحصیلی، به طرز متفاوتی خودآزمایی می‌کنند بلکه در یک گروه همسان نیز نحوه خودآزمایی تحت تأثیر زمان و مکان قرار می‌گیرد. عوامل دخیل در نحوه خودآزمایی دانشجو از این قرار است:

۱- عوامل درونی، هتلا: تمایل دانشجو به روند خودآزمایی.

۲- عوامل بیرونی، هتلا: مشخصات یک موقعیت مهارتی خاص.

کیفر^۴ متوجه شد دانشجویانی که خودآزمایی دقیقتری انجام می‌دهند نسبت به دانشجویان با خودآزمایی‌های بی‌دقت، نمرات درسی بالاتری دارند، به علاوه در دستیابی به هدف نهایی خویش نیز موفق‌تر عمل می‌کنند. او نشان داد دانشجویان از تصورات شخصی و تجربیات خودآزمایی‌های قبلی، اثر می‌پذیرند و نوعی تمایل درونی در آنها بوجود می‌آید که در خودآزمایی‌های بعدی برای همیشه به طور دقیق یا بی‌دقت عمل کنند. در هر موقعیت جدید آنچه رفتارهای خودآزمایی دانشجو را متأثر می‌کند محیط، متون درسی و مشخصات آزمون است که در مقابله با تمایلات درونی دانشجو قرار می‌گیرد. همچنین آشنایی دانشجو با حوزه مهارتی مورد ارزیابی و پیچیدگی و تخصصی بودن آن در دقت خودآزمایی او تأثیر دارد. البته میزان دانسته‌ها و توانایی‌های دانشجو، نحوه خودآزمایی را چندان متأثر نخواهد کرد؛ یعنی این طور نیست که همیشه دانشجویی

1- Peer Assessment

2- Stewart

3- Peter

4- Keefer

که بیشتر و بهتر می‌داند خود آزمایی دقیقتری هم از خود داشته باشد.

فرای^۱ و همکارانش تحقیق جالبی را در دانشگاه دریک‌فارسست در امریکا انجام دادند. آنها از دانشجویان خواستند مهارت خود را در چهار امتحان برگزار شده در دو سال اول دانشکده پزشکی پیش‌بینی کنند (نوعی خود آزمایی از آزمون اجرا شده). آنها می‌خواستند بدانند به مرور زمان چه تغییری در میزان دقت دانشجو در محک زدن خود پیش می‌آید.

در این مطالعه سعی بر آن بود که عوامل خارجی مثل نوع امتحان، نوع مهارت مورد امتحان و پیچیدگی آن ثابت نگه داشته شود و آنچه واقعاً در دقت خود آزمایی اثر می‌گذارد فقط تمایل درونی دانشجو و مهارت او با گذشت زمان باشد. برای اجرای این مقصود ۲۲ دانشجوی پزشکی که شیوه آموزش آنها مبتنی بر روش یادگیری بر مبنای حل مسأله^۲ بود، انتخاب شدند. این دانشجویان در سال اول دانشکده در سه امتحان تشریحی شش ساعته با فاصله ده هفته و در سال دوم پس از گذشت پانزده هفته از درس، در امتحان چهارم شرکت کردند. در هر جلسه امتحان یک برگه خود آزمایی همراه سؤالات به آنها داده می‌شد که حاوی نمره اختصاص یافته به هر سؤال بود و دانشجو باید بعد از اتمام پاسخ‌گویی، نمره خود را در آن سؤال، بر مبنای حدس و گمان علامت می‌زد. پس از پایان هر امتحان نمره فرد و جواب پیشنهادی برای هر پرسش به او ارائه می‌شد، اما هیچ اطلاعی راجع به دقت خود آزمایی به او نمی‌دادند. سپس نمراتی که مصحح با تجربه، به دانشجو داده و نمره‌ای که دانشجو برای خود در نظر گرفته بود به صورت درصد تبدیل و اختلاف آنها محاسبه شد (مثلاً: نمره ۳ از ۴ به ۷۵ درصد تبدیل می‌شد). نمره اصلی - نمره خود آزمایی = a

همانگونه که مشخص است، اگر a عددی مثبت باشد، یعنی دانشجو توانایی خود را بالاتر از آن چیزی در نظر گرفته که واقعاً استحقاق آن را داشته و اگر عددی منفی باشد یعنی دانشجو خودش را دست کم گرفته است. سپس میانگین موارد تخمین زیاد و تخمین کم دانشجو بدست آمد:

$$\begin{aligned} \text{مجموع مقادیر مثبت } a &= \frac{\text{میانگین تخمین زیاد}}{n} \\ \text{مجموع مقادیر منفی } b &= \frac{\text{میانگین تخمین کم}}{n} \end{aligned} \quad n = \text{تعداد کل سؤالات}$$

در حقیقت مقادیر زیاد عددی میانگین تخمین نشان‌دهنده درصد مواردی است که دانشجو تصور می‌کند به آن مطلب خاص رسیده، اما در حقیقت نرسیده است و مقادیر کم

میانگین تخمین نشان دهنده درصد مواردی است که دانشجو تصور نمی‌کند به مطلب رسیده باشد. مواردی که تخمین بالا یا پایین نیست در واقع تخمین دقیق دانشجویست. ($a=0$). با توجه به آنکه یک نفر، مسؤلیت تصحیح تمام سؤالات را بر عهده داشت، مشکلی در پایایی نمرده‌هی وجود نداشت. همچنین همبستگی بین نمرات چهار امتحان با آلفای کرونباخ محاسبه شد که همبستگی خوبی را نشان داد (از $0/79$ تا $0/85$).

نتایج:

۱- تخمین دقیق ($a=0$) از امتحان اول تا چهارم روند صعودی داشت.
 ۲- موارد تخمین کم ($a < 0$) در امتحان اول تا سوم نسبتاً ثابت و در امتحان چهارم افزایش ناگهانی داشت.

۳- کلاً بین نمرات خودآزمایی و نمرات واقعی، همبستگی چندانی وجود نداشت.
 ۴- خودآزمایی‌ها در امتحانات مجاور هم (مثل امتحان اول و دوم) بسیار همبستگی داشت، اما در امتحانات غیر مجاور (مثل اول و سوم) همبستگی چندانی نداشت.
 آنچه از این مطالعه و مطالعات مشابه بر می‌آید آن است که به طور کلی بیشتر دانشجویان مهارت‌های خود را کمتر از حد واقعی تخمین می‌زنند.

با توجه به این که عوامل خارجی ثابت نگه داشته شده بود، احتمالاً تغییرات مشاهده شده در امتحانات متوالی، به علت تغییر تمایلات درونی دانشجو و اثر آن بر روند خود آزمایی است؛ یعنی دانشجو به تدریج با افزایش دانشها و آگاهیها، بیشتر به نواقص خود در حوزه مهارتی مورد نظر پی می‌برد.

در آزمون اول، دانشجویان تصور می‌کردند که کارآموزی و تجربه، آنان را به اندازه کافی ماهر کرده بنابراین، نمرات بالاتری برای خود در نظر گرفتند. اما بعد از دریافت بازخوردهای لازم و نمره واقعی آزمون، به میزان اطلاعات خود در حوزه مورد نظر پی برده، به کاستیهای خود بیشتر واقف شدند.

از آنجا که همبستگی چندانی بین نمرات مصحح و نمرات خودآزمایی دانشجویان وجود نداشت می‌توان نتیجه گرفت آگاهی دانشجو از ناتوانیهایش بسیار بیشتر از اطلاعات واقعی او، خودآزمایی را متأثر می‌کند. زیرا آنچه مصحح می‌سنجد اطلاعات واقعی دانشجویست.

تمایل دانشجو- که خود را کمتر از آن چیزی که واقعاً هست بداند- به تدریج افزایش می‌یابد. اما چرا تفاوت چشمگیر بین امتحان سوم و چهارم وجود دارد؟ (منظور تفاوت

فاحشی است که در موارد تخمین پایین به چشم می‌خورد.) شاید به دلیل فاصله زمانی طولانی بین آنهاست که بیشتر از سایر آزمونها بوده، در نتیجه دانشجو تجارب قبلی خود را بدست فراموشی سپرده است. پس شاید بتوان با کاهش فاصله بین خودآزمایی‌ها به دانشجو کمک کرد تا تصحیح مناسب انجام داده، خطاهای قبلی را تکرار نکند.

چون خود آزمایی دقیق احتیاج به تمرین دارد، اگر در پی کاربرد نتایج خودآزمایی در برنامه‌ریزی‌های آموزشی هستید، باید از تأخیر طولانی و فواصل نامنظم بین امتحانات جلوگیری کنید.

اگر بتوان افت دقت بین امتحان سوم و چهارم را به افزایش آگاهیهای دانشجو و اثر وقوف او بر ناتوانی‌هایش نسبت داد، باید به او آموخت که به جای پاسخگویی به حس اضطراب درونی، به خود و دانسته‌های خویش اعتماد کند و توانایی خود را دست کم نگیرد. البته با توجه به محدود بودن تعداد نمونه در این مطالعه و شیوه تدریس که یادگیری بر مبنای حل مسأله بوده است، شاید نتوان نتایج این تحقیق را تعمیم داد و این تعمیم احتیاج به تحقیقات وسیع‌تر دارد.

پیش از انجام برنامه‌های خودآزمایی ابتدا باید به دانشجویان آموخت که چگونه توانایی‌های خود را سنجیده آنها را پرورش دهند. این همان چیزی است که در حرفه آینده آنان نیز بسیار حیاتی بوده، آنها را از هر نوع افراط و تفریط در مورد آنچه می‌دانند، برحذر خواهد داشت.

جنبه‌های کاربردی خودآزمایی

یکی از کاربردهای بسیار چشمگیر خودآزمایی استفاده از نتایج آن برای برنامه‌ریزی دوره‌های آموزش مداوم (آموزش ضمن خدمت) پزشکان است. این شیوه نخستین بار توسط NBME^۱ اجرا شد.

دوره‌های آموزش مداوم در بسیاری از کشورها بدون هیچ مطالعه قبلی برنامه‌ریزی می‌شود. یعنی مسئولان دانشکده‌ها بدون در نظر گرفتن نیازهای آموزشی پزشکان در کارهای حرفه‌ای اقدام به برگزاری دوره‌های آموزش ضمن خدمت می‌کنند. مثلاً: چون مسئول برگزاری این دوره‌ها به تازگی چیزهایی از یادگیری بر مبنای حل مسأله شنیده که برایش جالب بوده، یک دوره راجع به موضوعی نه چندان مهم با استفاده از این روش برگزار

کرده، زمانی دیگر اقدام به برگزاری یک دوره آموزش رایانه یا روش تدریس برای پزشکان می‌کند. چنین روندی که هیچ نظم خاصی را دنبال نمی‌کند فقط هر چند وقت یکبار پزشکان را وادار می‌کند که وقت آنها را هدر می‌دهد.

وظیفه بخش آموزش ضمن خدمت هر دانشکده بسیار مهمتر و پیچیده‌تر است. طرح‌ریزی دوره‌های بازآموزی باید به صورتی باشد که اگر همه این دوره‌ها در کنار هم قرار بگیرد مانند اجزای یک ماشین با هم هماهنگ بوده، بتوان از مجموعه این دوره‌ها یک شکل مطلوب و دلپذیر ساخت، همه آنها به سمت یک هدف واحد جهت‌گیری شده باشد. اگر هر دوره برای خود مستقل بوده، با دوره‌های دیگر هیچ تناسبی نداشته باشد، شاید در کوتاه مدت نتایجی در بر داشته باشد اما در دراز مدت بی‌اثر بوده، جز صرف بی‌نتیجه هزینه و وقت، هیچ فایده دیگری ندارد. برای رفع چنین معضلی NBME دست به کار شد.

چنانکه قبلاً بیان گردید آزمونهای چندگزینه‌ای بسیار کارآمد بوده، به کمک آن می‌توان طیف وسیعی از دانسته‌ها را مورد سنجش قرار داد. سوالات آن همچون دارویی شفابخش هر روز کاربرد جدیدی پیدا می‌کند. در سال ۱۹۶۶ مجمع پزشکی آمریکا^۱ مبتکر کاربرد این شیوه آزمون، در راه جدیدی به نام «خودآزمایی» شد.

هدف این مجمع آن بود که بفهمد پزشکان آمریکایی تا چه حد به نیازهای آموزشی خود واقفند و چقدر علاقه‌مند هستند اطلاعات خود را روزآمد^۲ کنند.

برای اجرای این منظور شیوه ساده‌ای را انتخاب کردند: ابتدا پرسشنامه‌ای تهیه و آن را بین پزشکان توزیع نموده، از آنها درخواست کردند که بی هیچ اغمازی حیطه‌های دانشی را که در آنها احساس کمبود می‌کنند، بیان کنند.

واضح است که پاسخ چنین سؤالی اعتبار چندانی نداشته، اصل این سؤال نیز کمی ناجوانمردانه می‌نمود؛ چون برای یک پزشک بسیار مشکل است که به طور دقیق آنچه را که نمی‌داند مشخص کند. در عین حال اقرار به ندانستن، خود طبع بلند مایه‌ای می‌خواهد. ایراد اصلی آن بود که اگر پزشکی راجع به موضوعی در علم طب، اطلاعی نداشته، اصلاً نسبت به وجود آن آگاه نباشد چگونه می‌تواند این امر را نشان دهد. به عنوان مثال اگر برای برگزاری یک دوره روش تدریس، از همکارانتان در مورد روشهای مختلف تدریس سؤال کنید، هر فرد فقط می‌تواند راجع به شیوه‌هایی نظر دهد که در مورد آنها اطلاعی ولو جزئی داشته، در آن مقوله احساس ضعف می‌کند. مثلاً: ممکن است چنین جمله‌ای را دریافت کنید که: «راجع به

شیوه کار در گروه کوچک^۱ اطلاع کمی دارم» اما اگر دوست شما حتی یکبار هم راجع به PBL (یادگیری بر مبنای حل مسأله) چیزی نشنیده باشد، بدیهی است که در پاسخ به این پرسش راجع به آن چیزی نمی نویسد و به طور قطع شما جوابی این چنین دریافت نخواهید کرد که: «من راجع به تدریس به شیوه PBL چیزی نمی دانم چون تاکنون اسم آن را نشنیده‌ام».

در سال ۱۹۶۷ به دنبال رفع این نقیصه، کالج پزشکان آمریکا^۲ شیوه جدیدی اتخاذ کرد. در این شیوه تعدادی سؤال چند گزینه‌ای در اختیار پزشکان قرار گرفته، از آنها درخواست شد پاسخ آن را برای کالج بفرستند. مدت زمانی که پزشکان صرف پاسخگویی به سؤالات می‌کردند اهمیتی نداشت (یعنی طول زمان پاسخ‌گویی محدود نبود). آنها در پی آن بودند که موقعیتی در اختیار پزشک قرار دهند تا بتواند هر نوع کمبود در دانسته‌های خود را تشخیص داده بر طبق آن، مورد آموزش قرار گیرد. این شیوه را «خودآزمایی دانش پزشکی»^۳ نامیدند.

برای طرح پرسشهای این آزمون، نه گروه شش نفری از متخصصان داخلی همکاری کردند. از هر گروه درخواست شد که در زمینه حیطه فوق تخصصی خود هشتاد سؤال چند گزینه‌ای طرح کند. اعضای هر گروه که در یکی از زیر شاخه‌های رشته داخلی تبحر داشتند شامل گروههای زیر بودند:

گروه ۱: هماتولوژی (بیماریهای خون)

گروه ۲: متابولیسم و اندوکرینولوژی

گروه ۳: بیماریهای ریوی

گروه ۴: بیماریهای کلیوی و اختلالات آب و الکترولیت

گروه ۵: بیماریهای مفاصل و بافت همبند

گروه ۶: بیماریهای عفونی

گروه ۷: گوارش

گروه ۸: قلب

گروه ۹: آلرژی و بیماریهای ایمنی

سؤالات طرح شده توسط هر گروه در بردارنده کلیه اهداف رشته موردنظر بود، به طوری که پاسخ فرد به همه آنها وجود اطلاعات کافی را در آن زمینه خاص نشان می‌داد.

1- Small Group Discussion

2- (A.C.P) American College of Physicians

3- Medical Knowledge Self Assessment

طبیعی است که گنجاندن اهداف یک رشته تخصصی در قالب هشتاد سؤال کار بسیار مشکلی است. ولی در این آزمون تکیه اصلی سوالات بر مسائل روز پزشکی بود و توجه خاصی به پیشرفتهای جدید این علم داشت. هر گروه از طراحان سؤال، موظف بودند برای تک تک سوالات خود مرجع معرفی نمایند، چون هدف اصلی از برگزاری این آزمون مراجعه شرکت کنندگان به منابع مطالعاتی و جبران کمبودهای اطلاعاتی بود.

سوالات پس از تأیید توسط NBME، به صورت یک پرسشنامه ۷۲۰ سؤالی راجع به رؤوس مطالبی که هر گروه، مهم تشخیص داده بود، در اختیار ۹ هزار پزشک داوطلب آمریکایی قرار گرفت. حضور تعداد زیاد شرکت کنندگان، نشان دهنده حساسیتی است که بسیاری از پزشکان در مورد دانسته هایشان به خرج می دهند.

اکثر پزشکان معتقد بودند که این آزمون مهمترین فعالیت پس از تحصیل از سوی مسئولان آموزشی دانشکده ها بود؛ اما خیلی ناچیز به نظر می رسید. زیرا مراکز قانونی و رسمی و مسئولان آموزشی نتیجه خاصی بدست نیاورده، صرفاً نمرات را به طور محرمانه به هر پزشک ارائه کردند. البته بی فایده هم نبود، چون حداقل توانستند به پزشکان اطمینان دهند که این آزمون صرفاً برای اطلاع آنان از وضعیت خویش بوده است و هیچ مرجعی نمی خواهد با استفاده از این سوالات آنها را در تنگنا گذاشته، معلومات ایشان را زیر سؤال ببرد.

پس از آنکه کالج پزشکان توانست اعتماد شرکت کنندگان در آزمون را جلب کند تصمیم گرفت این بار خیلی جدی تر گام برداشته، برنامه را دنبال کند.

برای برگزاری مجدد این آزمون آکادمی پزشکان اطفال آمریکا^۱، انجمن روانپزشکان آمریکا^۲ و انجمن پزشکان بیهوشی آمریکا^۳ با کالج پزشکان همکاری کرده، به مجموع رشته های مورد امتحان، زنان و زایمان هم اضافه شد.

همراه با سوالات، پرسشنامه ای نیز برای شرکت کنندگان ارسال شده، از آنها درخواست گردید به سوالات زیر پاسخ گویند:

- ۱- رشته تخصصی و فوق تخصصی؛
- ۲- سال فراغت از تحصیل؛
- ۳- زمینه هایی که به آن علاقه مند بوده، راجع به آن مطالعات جنبی دارند. مثلاً: فرد

1- American Academy of Pediatrics
3- American Association of Anesthesiologists

2- American Association of Psychiatrists

هماتولوژیستی که به علت علائق شخصی راجع به ریه مطالعه می‌کند.

۴- بیان این نکته که آیا در پاسخ به سؤالات از کمک و یاری همکاران یا کتابهای مرجع استفاده کرده‌اند یا خیر.

مجمع در همان پرسشنامه به افراد اطمینان داد آنچه در پی آن است تهیه آمار و اطلاعات کلی بوده، نمرات افراد از هم قابل تفکیک خواهد بود.

این بار پزشکان به خاطر تجربه شرکت در آزمون قبلی، تشویشی نداشتند که شاید این آزمون راهی پیچیده و مخفیانه برای کسب اطلاعات از افراد باشد و خیلی راحت به آن پاسخ گفتند.

کالج پزشکان آمریکا با توجه به داده‌ها و نمرات شرکت‌کنندگان به طبعه‌بندی افراد و یافتن خط‌مشی برای دوره‌های آموزش مداوم پرداخت. پیش از آن در کلاسهای بازآموزی چیزهایی تدریس می‌شد که مدرس آنها را مهم تلقی می‌نمود اما پس از برگزاری آزمونهای خودآزمایی، سعی بر تدریس مطالبی شد که اغلب شرکت‌کنندگان در آن زمینه نقص اطلاعاتی داشتند.

در ضمن با ارائه اطلاعات کلی راجع به میانگین نمره هر گروه، هر فرد می‌توانست خود را با سایرین مقایسه کند. مثلاً: با ارائه گستره^۱ نمره کسانی که از کتب مرجع برای پاسخ‌گویی به سؤالات استفاده کرده بودند یا محدوده نمره افراد متخصص در هر رشته، سایر افراد می‌توانستند خود را با آنها مقایسه کنند.

پس از اعلام نتایج به صورت محرمانه، به هر فرد پیشنهاد می‌شد که با توجه به نمرات کسب شده در هر زمینه خاص، بهتر است در کدامیک از دوره‌های بازآموزی شرکت کند.

نمرات این آزمون گرچه پایایی چندانی نداشت، آثار چشمگیر آن به هر فرد این امکان را می‌داد که نقائص اطلاعاتی خود را شناخته، برای رفع آن از طریق مطالعه کتب مرجع یا شرکت در دوره‌های بازآموزی اقدام کند و به مسؤولان نیز برای جهت‌دهی به دوره‌های آموزش مداوم کمک می‌کرد.

از آنجا که هر فرد برای موفقیت باید کاستی‌های دانسته‌های خود را بشناسد خودآزمایی، وسیله‌ای عینی برای شناسایی این نقایص است. همچنین هر جامعه‌ای علاقه‌مند است مشخص کند پزشکی که سالها قبل فردی حاذق و لایق شناخته شده آیا هنوز

قابلیتهای خود را دارد. بدین منظور با کمک خودآزمایی می‌توان مدارک پزشکان را دوباره تأیید کرده، توانایی آنها را مشخص نمود.

در آمریکا NBME تلاش می‌کند تا بدین وسیله ارزیابی واقع‌بینانه‌ای از دانش بالینی و مهارت پزشکان انجام داده، آنها را به مجامع تخصصی معرفی کند. البته این شیوه برای کسانی که علاقه‌ای به شرکت در برنامه‌های خودآزمایی ندارند بی‌فایده است و شاید تنها راهی که در مورد این افراد باقی می‌ماند استفاده از الزام قانونی باشد.

منابع:

- 1- Crawford MW, Kiger AM. *Development through self assessment: Strategies used during clinical nursing placements.* J Adv Nurs 1998 Jan; 27(1): 157-64.
- 2- Hay JA. *Investigating the development of self evaluation skills in a problem based tutorial course.* Acad Med 1995; 70(9): 633-5.
- 3- Krause E. *A group interaction peer/self assessment process in a pharmacy practice courses.* Am J Pharm Educ 1996 Sum; 60(2): 136-1246.
- 4- Tracey JM, Arroll B, Richmond DE, Barnam PM. *The validity of general practitioners self assessment of knowledg: Cross sectional study.* BMJ 1997 Nov 29; 315 (7120): 1426-8.

فصل هشتم

سنجش حیطة‌های فیر شناختی

● سنجش عملکرد

● سنجش نگرش

سنجش عملکرد

شاغلین پزشکی در عرصه ارائه خدمات بهداشتی - درمانی، مسؤول سلامت مردم و در پی انجام اموری هستند که منجر به حفظ و ارتقای سلامت و بهداشت عمومی خواهد شد. این امر به ویژه برای پزشکان، دندانپزشکان و داروسازان - که گاهی مسؤولیت دارند بر کار سایر اعضای تیم مراقبت‌های بهداشتی نظارت کنند - بسیار خطیر و حساس است. در راستای این ایفای وظیفه، علاوه بر داشتن دانش کافی، وجود مهارت‌های حرفه‌ای نیز امری الزامی می‌نماید.

اما باید توجه داشت، وجود دانش و مهارت کافی و مناسب هم، تضمین کننده عملکرد مناسب در عمل نمی‌باشد. به عنوان مثال برای یک پزشک کفایت بالینی^۱ شامل مواردی است که هر پزشک باید در حد قابل قبولی بداند و انجام دهد و از آن جمله می‌توان مهارت‌های زیر را نام برد:

آگاهی از اطلاعات و علوم پایه، توانایی به کار بردن اطلاعات علمی، توانایی کسب اطلاعات جدید، مهارت در دریافت اطلاعات از بیمار (شامل: گرفتن شرح حال، معاینه فیزیکی و آزمایش‌های پاراکلینیکی)، توانایی استنتاج اطلاعات بالینی، (یعنی: ترکیب اطلاعات و ارزیابی فرضیات و حل مشکل بیمار)، توانایی طراحی اداره و تدبیر^۲ بیمار و اجرای طرح‌های درمانی (یعنی تشخیص، درمان، آموزش، مشاوره و پیگیری بیمار) و بالاخره ویژگی‌های شخصیتی (شامل عملکرد و عادات حرفه‌ای، شخصیت اجتماعی و مسؤولیت‌پذیری و نگرش پزشک).

اما عملکرد بالینی^۳ آن چیزی است که پزشک یا دانشجو واقعاً در کارهای روزمره در مطب و بالین بیمار، یعنی در برخورد با بیماران انجام می‌دهد. واضح است که در حالت ایده‌آل، عملکرد بالینی باید با کفایت بالینی هماهنگ و همسان باشد، اما متأسفانه در عمل، به دلایل زیادی این دو، با یکدیگر ارتباط قوی ندارند. ریتنس^۴ و همکارانش (۹۱-۱۹۹۰) نشان

1- Clinical Competence
3- Clinical Performance

2- Management
4- Retance

دادند بین نتایج حاصل از برخورد پزشکان با بیماران تقلیدی و بیماران واقعی در مطب و عملکرد روزمره، اختلاف وجود دارد. پزشکان در نظام نمره‌دهی در برخورد با بیماران تقلیدی، نمرات بهتری کسب کرده‌اند؛ بدین معنا که بین کفایت و عملکرد پزشکان تفاوت وجود دارد، استفاده از بیماران تقلیدی برای ارزیابی کفایت پزشکان روش مناسبی است اما برای بررسی عملکرد واقعی پزشکان روش مطلوبی به حساب نمی‌آید.

بدیهی است آزمونهای مرسوم که به آزمونهای کاغذ - مدادی یا آزمونهای کتبی مشهور هستند، بیشتر آن چیزی را می‌سنجند که اندازه‌گیری آن آسان است و کمتر در پی سنجش عملکرد فرد در موقعیت انجام کار یا حل مسأله هستند. بنابراین با توجه به اهمیت نقش عملکرد در ایفای مسؤولیت حرفه‌ای، آزمونهای عملکرد که به آنها آزمونهای واقعی یا اصیل^۱ نیز می‌گویند، امروزه جایگاه ویژه‌ای در سنجش دانشجویان یافته است. آزمونهای عملکرد روشهایی هستند که سعی دارند «توان انجام» فرد را اندازه‌گیری کنند. عملکرد بر رفتار کلی کارکنان بهداشتی متمرکز است، به عنوان مثال: قدرت سازمان‌دهی، یادداری^۲، استفاده از دانش اختصاصی و ارتباط با عامه مردم. این واژه، گستره کاملی از دانش، مهارت و نگرش موردنیاز که در عمل، سازمان‌دهی و تلفیق شده است را در بر می‌گیرد. البته باید توجه داشت که به دلیل وجود عوامل مختلف مؤثر در ایجاد رفتار انسانی و پیچیدگی عناصر مرتبط با آن، در بسیاری از موارد، ایجاد تمایز دلخواهانه بین دانش، مهارتها و نگرش‌ها چندان سودمند نیست.

سنجش عملکرد^۳ در پی اندازه‌گیری توانایی فرد در انجام یک وظیفه^۴ ویژه بوده، به جنبه‌های سازماندهی و سازگاری نتیجه عملکرد به صورت یک کل، توجه اساسی دارد؛ زیرا یک عملکرد مناسب در کل، معادل جمع قطعات جزء به جزء آن نیست. نتایج این آزمونها بر تعمیم مشاهده انجام یک فعالیت مراقبت بهداشتی، توسط یک فرد به سایر توانمندیهای او مبتنی است؛ که این نتایج مانند هر ارزیابی دیگر می‌تواند در بازنگری برنامه‌های آموزشی، انتخاب، قبول و ارتقای افراد موردنظر و تشخیص ضعفها و فعالیتهایی که نیاز به تکرار دارند، مورد استفاده قرار گیرد.

آزمونهای عملکرد در پی اندازه‌گیری فرایند^۵ و یا فراورده^۶ است. فرایند به مراحل که یک وظیفه انجام می‌شود باز می‌گردد و معمولاً به وسیله افعالی مانند بیرون کشیدن،

1- Authentic

3- Preformance Assessment

5- Procedure

2- Retention

4- Task

6- Product

اصلاح کردن، گرفتن، ساختن و ... مشخص می‌شود. فرآورده یا محصول به نتیجه پایانی باز می‌گردد و معمولاً با کاربرد اسامی ویژه‌ای مانند شرح حال، معاینه و بالاخره سلامت همراه است.

در حالاتی که انجام عملکرد به یک فرآورده ختم نشود، اندازه‌گیری فرایند اهمیت ویژه‌ای می‌یابد و در صورتی که عملکرد محصول عینی داشته باشد به فرآورده توجه خاص مبذول می‌شود؛ اما در اغلب موارد اندازه‌گیری فرایند و فرآورده به صورت همزمان مورد توجه قرار می‌گیرد. این امر به خصوص در علوم تندرستی که به جسم و جان آدمی مربوط بوده، به ابعاد انسان توجه دارد، بیشتر ملموس است.

اگر چه فرآورده و فرایند ارتباط درونی دارد، معمولاً آنقدر جد است که می‌توان آنها را مشاهده کرد، اندازه گرفت و مستقلاً ارزشیابی نمود. نمره‌دهی فرایند در خلال مراحل اول یادگیری و فرآورده در مراحل نهایی اهمیت می‌یابد. به عنوان مثال: در مراحل اولیه از دانشجوی بالینی پزشکی انتظار داریم به درستی فشارخون را اندازه‌گیری نماید. اما درمان صحیح بیمار و کنترل فشارخون در حد طبیعی، فرآورده نهایی مورد انتظار است. گاهی اوقات ممکن است با وجود داشتن یک فرآورده مناسب، فرایند دچار نقصان و کاستی باشد، بنابراین، اندازه‌گیری هر دو مدنظر قرار می‌گیرد. مثلاً ممکن است با وجود درمان صحیح بیمار، ارتباط پزشک و بیمار روابط صحیح حاکم نبوده پزشک نتواند ارزشهای اخلاق پزشکی را در حدی که از اجزای عملکرد بالینی اوست به نحو مطلوب به اجرا گذارد. ذکر این نکته ضروری است که چون دانشجوی می‌آموزد و نحوه آن به صورت آشکاری تحت تأثیر نحوه سنجش اوست؛ بنابراین هماهنگی و ارتباط شیوه‌های سنجش با نیازهای آموزشی اهمیت بسزایی خواهد داشت، در ارزیابی دانشجوی باید به بررسی عملکرد در ارائه خدمات و مراقبتهای بهداشتی - درمانی توجه ویژه نمود.

سطوح سنجش عملکرد

سطح اول

دانش بالینی: شامل آزمونهای کتبی (تشریحی، MEQ، کوتاه پاسخ، چندگزینه‌ای، اداره بیمار با استفاده از کاغذ یا کامپیوتر) می‌باشد که مبتنی بر شبیه‌سازی کلامی از واقعیتهای بالینی است. از آزمون تشریحی می‌توان برای اندازه‌گیری توانایی دانشجوی در

سازمان‌دهی^۱، تلفیق^۲ یا سنتز^۳ اطلاعات استفاده کرد. این آزمونها نسبت به آزمونهای کوتاه پاسخ و چند گزینه‌ای قابلیت بیشتری در اندازه‌گیری رویکردهای ابتکاری و نوآوری فردی در حل مشکلات، از برخوردار است. قابلیت‌های آزمونهای مزبور به شکل کامل در فصول مربوط ذکر گردید.

استفاده از دانش بالینی برای بررسی عملکرد بالینی مناسب و ضروری است اما با آنچه در واقعیت رخ می‌دهد کاملاً مساوی نیست. چنانکه ذکر شده «دانش» معادل «انجام» نیست و هیچ روش گفتاری را اعم از شفاهی یا کتبی نمی‌توان آزمون عملکرد بالینی دانست چرا که خوب گفتن یک روش، معادل انجام دادن خوب آن نیست. بدین لحاظ آزمونهای مزبور برای سنجش عملکرد میدانی فرد، روایی کافی ندارند.

به عنوان مثال: برای اندازه‌گیری توانایی یک دانشجوی پزشکی، در گرفتن شرح حال از یک زن باردار در نیمه دوم حاملگی، نمی‌توان یک دسته سؤال چندگزینه‌ای تنظیم کرد و از وی خواست آنها را جواب دهد، بلکه استاد یا ناظر باید یک فهرست بازبینی (چکلیستی) تهیه کرده، از دانشجو بخواهد از بیمار موردنظر شرح حال تهیه کند و عملکرد دانشجوی موردنظر را، در حالت مواجهه با بیمار، مورد بررسی قرار دهد.

سطح دوم

آزمونهای شناسایی^۴: در این آزمونها مشخص می‌شود که آیا آزمون شونده می‌تواند ویژگیهای اساسی فراورده یا عملکرد را شناسایی کند یا خیر. در این سطح اجزای مهارتهای عملکرد بالینی می‌تواند بصورت وظایف تجزیه شده، جداگانه مورد اندازه‌گیری قرار گیرد. به عنوان مثال می‌توان با پخش چند صدای قلب از دانشجویان خواست صدایی را که حاوی سوفل می‌باشد مورد شناسایی قرار دهند. فعالیت‌های لازم در هر مرحله را می‌توان به صورت دسته‌بندی مراحل مشخص کرد. از نمونه‌های این روش ارزیابی، آزمونهای آسکی (OSCE)^۵ می‌باشد. این چهارچوب برای سنجش مهارت در روشها و فنون و نیز آزمون میزان فراگیری مهارتهای بالینی مناسب است و انجام بحث و تبادل نظر قبلی اساتید در مورد ملاکها و انتظارات به ویژه وقتی مهارتهای فردی بررسی می‌شود ارزشمند خواهد بود. به هر حال این آزمونها در اندازه‌گیری مهارتهای بالینی از دقت و کارایی کمی برخوردار است.

1- Organization

2- Integration

3- Synthesis

4- Recognition Test

۵- آزمونهای OSPE یا OSCE (Objective Structured Clinical Examination)

سطح سوم

آزمون در شرایط شبیه‌سازی^۱: در این آزمون‌ها سعی بر آن است که از آزمون‌شونده بخواهند عملکرد خود را در یک موقعیت شبیه‌سازی شده نشان دهد. هدف این سطح، سنجش کفایت بالینی^۲ دانشجوی در مواجهه با یک بیمار فرضی است. واحد مطالعه به طور اعم، بیمار و بیماری است. در این آزمون‌ها دانشجویان باید بدانند چه پرسند و به دنبال چه چیزهایی بگردند تا بتوانند اطلاعات مناسب، دقیق و کاملی را جمع‌آوری کنند و در نهایت، براساس شواهد جمع‌آوری شده تفسیر کنند و شواهد را به تشخیص ملحق سازند. وقتی دانشجوی یا یک بیمار برخورد می‌کند، وظایف وی ثابت و شناخته شده نیست. فرایند بالینی دربرگیرنده برخورد فعال و پویای دانشجوی و بیمار است. لازم به یادآوری است که استفاده از موقعیتهای شبیه‌سازی شده، به‌ویژه برای آموزش مهارتهای مختلف نیز سودمند است.

سطح چهارم

آزمونهای نمونه کار^۳: در این شیوه از آزمون‌شونده درخواست می‌شود تا یک دسته اعمال را که معرف عملکرد واقعی اوست به انجام رساند. این اعمال، شامل رفتارهایی عادت‌گونه همراه با پیچیدگی مسؤولیت‌های عملکرد بالینی می‌گردد. گستره مطالعه در مورد عملکرد بالینی یک پزشک از یک بیمار تا مراقبت از بیماران در یک بخش، بیماران مراجعه‌کننده به اورژانس، ارتباط با سایر اعضای مراقبتهای بهداشتی، ارجاع به سایر پزشکان، مراقبتهای طولانی مدت از بیماران با بیماریهای مزمن و مانند آن تنوع دارد. مهارتهایی مانند مدیریت یا اداره بیمار، قدرت برقراری ارتباط و کارایی، در کیفیت مراقبت ارائه شده کاملاً مؤثر می‌باشند؛ و در نتیجه، در این سطح مورد توجه قرار می‌گیرد. به دلیل آنکه وضعیت حاکم، کاملاً واقعی است و شخص دقیقاً کاری را که از او انتظار می‌رود انجام می‌دهد، این آزمون‌ها از روایی بالایی برخوردار است. برای رسیدن به پایایی این آزمون‌ها، باید در طراحی و برگزاری آنها دقت کافی مبذول داشت. مثلاً علاوه بر استفاده از ابزارهای صحیح مانند فهرستهای بازبینی دقیق، لازم است از وجود چند مشاهده‌گر استفاده نمود.

در راستای سنجش عملکرد در موقعیتهای شبیه‌سازی شده یا نمونه کار، علاوه بر مشاهده رفتار آزمون‌شونده، می‌توان از مصاحبه با افرادی بهره گرفت که وی با آنها ارتباط دارد (مانند بیمار یا اعضای تیم مراقبتهای بهداشتی - درمانی).

مراحل طراحی آزمونهای عملکرد

تهیه آزمونهای عملکرد نیز تابع قوانین کلی تهیه سایر آزمونها می باشد، اما طراحی، اجرا و نمره گذاری آنها تجربه و دقت بیشتری را طلب می کند. دستورالعملهای زیر به منظور ارائه راهکارهای مدون برای ایجاد سهولت بیشتر در این امر بیان می گردد.

۱- تجزیه و تحلیل وظایف شغلی^۱

اولین اقدام، شناسایی وظایف شغلی یک عضو تیم مراقبتهای بهداشتی - درمانی در عرصه ارائه خدمات می باشد. هر عملکرد کلی از چند نقش^۲ تشکیل شده است؛ هر نقش یک یا چند وظیفه را در بر می گیرد و هر وظیفه شغلی شامل اجزای وظیفه و کفایتهای لازم است، که این اجزا نیز قابل تجزیه می باشد. تجزیه شغل به اجزایش، تجزیه و تحلیل وظایف شغلی نامیده می شود. این امر به شناسایی آنچه یک دانشجو در آینده باید بتواند انجام دهد، کمک می کند. در انجام تجزیه و تحلیل وظایف شغلی می توان از روشهای مختلف بهره جست از آن جمله:

۱- مشاهده شاغلان به حرفه در عرصه ارائه خدمات یعنی همانهایی که آن وظیفه شغلی را انجام می دهند، به ویژه آن جنبههایی که به صورت قاطع ملاک رد یا قبول فرد می باشد.
 ۲- مصاحبه با شاغلان که از آنها درخواست می شود نحوه انجام یک وظیفه شغلی را بیان کنند. همچنین از دفاتر خاطرات روزانه، بررسی پرونده بیماران، تجزیه و تحلیل پیگیری بیماران و شبیه سازی موقعیت برای شاغلان نیز می توان استفاده نمود.

اما بهترین راه حل آن است که خود آزمون گیرندگان، کار را بیاموزند یا انجام دهند. هنگامی که آزمون گیرندگان درگیر شغل باشند، آزمون بهتری برگزار خواهند کرد.

در تجزیه و تحلیل وظایف شغلی باید به موارد زیر توجه داشت:

۱- تجزیه و تحلیل وظایف شغلی باید خلاصه ای از عملکرد واقعی شغل باشد، لذا فقط اهداف رفتاری را در یک برنامه درسی در بر می گیرد.

۲- طبیعت هر وظیفه شغلی بسته به موقعیت زمانی و مکانی، متفاوت می باشد.

۳- سنجش عملکرد در سطوح مختلف از کفایتهای طراحی می گردد و از هر فرد بسته به درجه و مرحله آموزشی، یکی از این سلسله مراتب مهارتی انتظار می رود.

۴- در سنجش عملکرد ممکن است، جنبه های مختلف آن از جمله دقت انجام، هتل اندازه گیری دقیق فشارخون با دقت ۱۰ mmHg، توالی درست مراحل هتل برخی کارهای

آزمایشگاهی، سرعت انجام هانند لوله‌گذاری تراشه در یازده ثانیه، مهارت در انجام هتل گچ گرفتن یک پای شکسته با استفاده از مواد و وسایل موردنیاز با مهارت کافی مدنظر قرار گیرند؛ البته ممکن است برای یک عملکرد، جنبه‌های مختلف انجام آن اندازه‌گیری گردد. این امر تابع طبیعت عملکرد و سطح مورد نظر اندازه‌گیری است. به عنوان مثال: سرعت انجام در مراحل نهایی آموزش از اهمیت ویژه‌ای برخوردار خواهد بود. در برخی موارد باید اعمال در رفتارها به صورت کاملاً متوالی انجام شود و این توالی، خود مدنظر است؛ اما در برخی موارد این توالی نه تنها سودی ندارد بلکه مضر است. هانند گرفتن شرح حال، که تکیه بر یک روش یکسان کار صحیحی نبوده، بسته به مورد، تصمیم‌گیری صورت می‌گیرد.

۲- انتخاب وظایف، مهارتها و تواناییهای مهم مرتبط با شغل

چنانکه ذکر شد، چند وظیفه نمایان‌کننده یک مهارت می‌باشد. بنابراین پس از تصمیم‌گیری راجع به اندازه‌گیری یک توانایی باید تصمیم گرفت که آیا عملکرد آن وظیفه، در طی فرایند مورد اندازه‌گیری قرار گیرد یا فرآورده آن عملکرد مورد اندازه‌گیری واقع شود و یا ترکیبی از هر دو روش.

تصمیم‌گیری راجع به آنچه «مهم است» کار ساده‌ای نیست. یک دیدگاه بر اساس ملاحظات اجرایی بر برون‌ده^۱ تأکید می‌کند. دیدگاه حرفه‌ای دیگر بر اطمینان از کفایت رفتار اخلاقی پا می‌فشارد و یک دیدگاه اقتصادی به دنبال کارآیی، تأثیر و بهره‌وری است و نظریه مشتری - محور، ظاهر ارتباط بین افراد و ارائه خدمات مناسب را مدنظر قرار می‌دهد. چه سؤالی را در مورد عملکرد می‌توان طرح کرد؟ شاید سؤالات زیر به ما کمک نماید:

الف- آیا او درست انجام می‌دهد؟ یعنی فرایند تشخیص که برابر است با طی یک فرایند منطقی و مؤثر برای جمع‌آوری اطلاعات دقیق بالینی.

ب- آیا او دانش صحیح دارد؟ یعنی دانش بالینی که معادل است با استفاده از دانش مناسب برای بررسی و کشف این مشکل ویژه.

ج- آیا او درست فهمیده است؟ یعنی تشخیص دقیق و کامل که معادل است با تقسیم، ارزش‌گذاری و تلفیق یافته‌های بالینی.

د- آیا طراحی صحیحی برای رسیدگی به بیمار دارد؟ یعنی طرح رسیدگی به بیمار.

ه- آیا بیمار راضی است؟ یعنی رضایت بیمار که برابر است با اطمینان از ارتباط مؤثر برای

ارائه راه‌حلهای قابل قبول.

و- آیا او کارایی دارد؟ یعنی کارایی که برابر است با استفاده از حداقل منابع، آزمایشها و صرف وقت.

ز- آیا رفتار درستی دارد؟ یعنی رفتار که معادل است با به دقت گوش دادن، ادای احترام، داشتن نزاکت و ادب.

ح- آیا وی قضاوت صحیحی دارد؟ یعنی قضاوت بالینی که برابر است با ایجاد تعادل بین تبادل ارزشهای متضاد و برون‌ده‌ها.

ط- آیا او ارجاعها را خوب تنظیم می‌کند؟ یعنی مسؤولیت‌پذیری گروهی که معادل است با انتخاب مناسب و انتقال صحیح سازمان‌دهی مراقبت.

ی- چقدر سوددهی دارد؟ یعنی اثربخشی که برابر است با رسیدن به بهترین نتایج با حداقل عوارض.

۳- تهیه فرمهای درجه‌بندی مشاهده‌ای^۱

این فرمها باید شامل موارد مشاهده و شیوه ثبت اطلاعات باشد. مثلاً آیا کیفیت عملکرد یا سرعت در آن مهم است؟ انواع مقیاسهای قابل استفاده در این فرمها در همین فصل مورد بررسی قرار خواهد گرفت.

۴- تهیه برخی از انواع طرحهای نمونه‌گیری از وظایف

باید جنبه‌های نسبتاً مهم یک وظیفه شغلی مورد انتخاب قرار گیرد. از آنجا که لازم است ارزیاب در زمینه جنبه‌های مختلف یک شغل، به صورت یک کل قضاوت کند در مورد انتخاب جنبه‌های مهم باید به سؤالات زیر پاسخ داد:

آیا وظیفه موردنظر یک جنبه، مهم و حیاتی شغل است؟

در چه شرایطی این عمل دارای برتری است؟

آیا نقصان عملکرد موجب اثر عمده‌ای بر بیماران یا تأثیری بر عملکرد کلی کارکنان نظام ارائه خدمات می‌شود؟

آیا باید تمام شاغلان به حرفه در این زمینه تسلط داشته باشند؟

۵- تدوین چارچوب اجرایی

باید محدودیتهای زمانی، امکانات لازم و آنچه موجود است، و شیوه نمره‌دهی توسط مربیان و هر آنچه موردنیاز است مشخص گردد.

روشهای ارزیابی انتخاب شده باید عملی بوده، حداکثر اطلاعات را در حداقل زمان و با حداقل هزینه ارائه نماید. یک ارزیابی باید در وضعیت ویژه قابلیت اجرا داشته، هزینه آن قابل پرداخت باشد. به عنوان مثال: آیا برای سنجش عملکرد مهارتی ۱۵۰ دانشجوی بالینی پزشکی در زمینه احیای ریوی- قلبی امکان استفاده از بیماران واقعی وجود دارد؟!

۶- بازبینی مجدد

سوالات آزمون باید قبل از اجرای نهایی مورد بررسی مجدد قرار گیرد. عملکردهایی که مورد سنجش قرار می‌گیرد، باید نمونه‌ای از گستره کامل شایستگیها برای اعمال و وظایف مورد انتظار آن شغل باشد.

ابزارها و روشهای مشاهده‌ای

بهرتر است مشاهدات نظام‌دار، در موقعیتهای حقیقی و در عرصه ارائه خدمات، و یا در شرایط شبیه‌سازی شده، به منظور اندازه‌گیری عملکرد دانشجویان و افراد شاغل به حرفه‌های بهداشتی - درمانی مورد استفاده قرار گیرد. این مشاهدات دارای محاسنی است، از جمله:

۱- مشاهده عملکرد و روش کار دانشجو به صورت منظم، میزان پیشرفت وی را مشخص می‌سازد. استاد نیز با شناسایی خطاها و مشکلات او، در اسرع وقت روشهای صحیح را به وی می‌آموزد.

۲- علاوه بر آن که مستلزم صرف وقت چندانی نیست، برخلاف آزمونهای مرسوم، دانشجویان نیز از آن کمتر واهمه دارند.

۳- اطلاعات ارزشمندی را که از سایر راهها قابل دسترسی نیست، به استاد ارائه می‌کند. از آنجا که استادان همواره مشاهده می‌کنند و براساس این مشاهدات به قضاوت و تصمیم‌گیری می‌پردازند، توجه به این نکته ضروری است که در این روش هم، مانند آزمونهای شناختی و کتبی باید ویژگیهای یک اندازه‌گیری خوب (داشتن روایی و پایایی، عینی و عملی بودن) مدنظر قرار گیرد. در این راستا پیشنهادهای زیر ارائه می‌گردد:

الف) یک فهرست راهنما یا فرم برای اهداف و ساختار مشاهده تهیه کنید. مشخص کنید چه اهدافی باید ارزشیابی شود و چه تصمیم یا تصمیماتی، احتمالاً براساس داده‌های جمع‌آوری شده اتخاذ خواهد شد. همچنین رفتار ویژه‌ای را که باید مشاهده گردد،

مشخص نمایید.

نخست باید اهداف دوره روشن گردد. سپس به دو سؤال زیر پاسخ داده شود:

۱- هدف مشاهده چیست؟

۲- چه خصوصیتی باید مشاهده شوند؟

باید توجه داشت برخلاف حیطه روانی - حرکتی، مشاهده آنچه در حیطه عاطفی رخ می‌دهد بسیار مشکل است. در این حیطه حرکاتی مانند صحبت‌های بیجا، لحن‌های خشن و تند و یا ظاهر آرام و متبسم می‌تواند جلب توجه نماید.

ابزارهای مشاهده می‌تواند اطلاعات زیر را عرضه کند:

اول) بیان‌کننده تفاوت افراد در حیطه مورد نظر است.

دوم) اجازه می‌دهد راجع به رفتارهای بعدی پیش‌بینی درستی صورت پذیرد.

سوم) در مورد اعمالی که منجر به اصلاح یا بهبود رفتارهای نادرست می‌گردد اطلاعاتی عرضه می‌کنند.

بنابراین باید دانست چه چیزی و به چه علتی مشاهده می‌شود.

ب) فقط به یک یا دو رفتار ویژه توجه نمایید. در طی یک جلسه درس یک کلاس سی نفری، با توجه به محدودیت زمان نمی‌توان تمامی رفتارهای متفاوت را در نظر گرفت؛ چرا که تمرکز مشاهده یک یا دو رفتار در هر ساعت، نتایج پایاتری را به بار خواهد آورد. رفتارهای مهمتر باید مشخص گردد.

ج) از اصطلاحات واضح و روشن استفاده کنید. اصطلاحات ذهنی و تعریف نشده که هر فرد، از آن به نوعی برداشت خواهد نمود یا معرف رفتار ویژه‌ای نیست، به روایی و پایایی اطلاعات به دست آمده خدشه وارد خواهد ساخت.

د) در صورت استفاده از چند مشاهده‌کننده، خطاهای نمونه‌گیری را به آنها آموزش دهید. برای به حداقل رساندن خطاهای نمونه‌گیری از نمونه‌گیری زمانی^۱ استفاده کنید. در این روش مشاهدات در مراحل و فواصل مختلف زمانی (روز، هفته، نیمسال) انجام می‌گیرد تا بدین ترتیب مشاهدات، نمونه‌ای از رفتار فرد موردنظر باشد. با افزایش دفعات و طول زمان هر مشاهده، روایی و پایایی نتایج افزایش می‌یابد.

ه) مشاهدات را براساس اهداف آموزشی تنظیم نمایید. در مشاهده نظام‌دار، هماهنگی آموزش داده شده و مشاهده ضروری است. بنابراین طرح مشاهده جزئی از فرایند طرح

درس خواهد بود. نبود هماهنگی بین مشاهده و اهداف آموزشی به نتایج ناروا^۱ منجر خواهد شد.

و) پس از رخداد واقعه، نتایج مشاهده را به دقت و فوری ثبت نمایید. فاصله زمانی بین وقوع رفتار و ثبت اطلاعات، احتمال تورش^۲ را افزوده، پایایی نتایج را کاهش می‌دهد.

ز) سعی کنید تا رسیدن موعد مقرر هیچ تفسیری از رفتارهای مشاهده شده ارائه ندهید. در اغلب اوقات هرگونه تمایل به تفسیر رفتار دانشجو در هنگام مشاهده یک رفتار ویژه، با عینی بودن جمع‌آوری اطلاعات حاصل از مشاهده تناقض دارد.

ح) از یک نظام دسته‌بندی و کدبندی مناسب استفاده کنید. این امر به مشاهده آسان و ثبت دقیق اطلاعات کمک شایانی می‌کند. هر چه میزان استنباط شخصی مشاهده کننده در ثبت و ضبط اطلاعات کاهش یابد به روایی داده‌ها کمک بیشتری می‌کند.

ط) مشاهده را به نحوی طراحی کنید که آزمون‌شوندگان احساس مزاحمت نکنند. اغلب مردم اگر بدانند در حال مشاهده شدن هستند، تمایلی به بروز رفتار معمولشان نخواهند داشت، بنابراین، مشاهدات مخفیانه و در طول دوره‌های آموزشی از اهمیت بسزایی برخوردار است.

انواع ابزارهای مشاهده

۱- فهرستهای بازبینی (وارسی)

فهرست بازبینی در برگیرنده فهرستی از مراحل، فعالیتها یا رفتارهایی است که یک مشاهده‌گر در هنگام انجام عمل ثبت می‌نماید. فهرستهای بازبینی مقیاسهای درجه‌بندی بوده برخی آن را در همین گروه، دسته‌بندی نموده‌اند. به هر حال باید توجه داشت که فهرست بازبینی به مشاهده‌کننده کمک می‌کند تا در مورد رخداد یا اتفاق نیفتادن یک فعل، یا حضور یا فقدان یک خصیصه (ولو آنکه بسیار سریع اتفاق افتاده باشد) اظهار نظر نماید ولی در پی ثبت درجه کیفی یک خصیصه و تواتر و فراوانی وقوع یک رفتار ویژه نیست. اگر چنین اطلاعاتی موردنیاز باشد از سایر مقیاسها مانند مقیاسهای درجه‌بندی یا واقعه‌نگاری استفاده می‌شود. فهرستهای بازبینی برای اندازه‌گیری فرایندهایی قابل استفاده است که به اجزایی از فعالیتها واضح، جداگانه و مشخص قابل تقسیم باشد؛ هانند به‌کارگیری یک میکروسکوپ. در استفاده از فهرستهای بازبینی به پیشنهادهای زیر توجه نمایید:

جدول ۸-۱. فهرست بازبینی مربوط به معاینه بیماران (اقتباس شده از طرح تحقیقاتی بررسی عملکرد و آگاهی پزشکان عمومی اصفهان - ۱۳۷۳)

خیر	بلی	
<input type="checkbox"/>	<input type="checkbox"/>	۱- معاینه انجام شد.
<input type="checkbox"/>	<input type="checkbox"/>	۲- فشارخون گرفته شد.
<input type="checkbox"/>	<input type="checkbox"/>	۳- جواب فشار خون به بیمار گفته شد.
<input type="checkbox"/>	<input type="checkbox"/>	۴- قلب سمع شد.
<input type="checkbox"/>	<input type="checkbox"/>	۵- افتالموسکپی انجام شد.
<input type="checkbox"/>	<input type="checkbox"/>	۶- وزن بیمار اندازه گیری شد.
<input type="checkbox"/>	<input type="checkbox"/>	۷- تندرns سوپراپوبیک بررسی شد.
<input type="checkbox"/>	<input type="checkbox"/>	۸- تندرns زاویه کوستوورتمبرال بررسی شد.
<input type="checkbox"/>	<input type="checkbox"/>	۹- تب اندازه گیری شد.
<input type="checkbox"/>	<input type="checkbox"/>	۱۰- معاینه ناحیه ژنتیال انجام شد.
<input type="checkbox"/>	<input type="checkbox"/>	۱۱- معاینات اضافی دیگر انجام شد - ذکر شود.

- ۱- تنها وقتی از فهرست بازبینی استفاده کنید که بخواهید از وجود یا فقدان یک خصیصه اطلاع پیدا کند.
- ۲- جنبه قابل مشاهده آن خصیصه را به صورت کامل و اختصاصی مشخص نمایید. بنابراین باید هر مهارت رفتاری مانند معاینه شکم، اندازه گیری فشارخون، اندازه گیری درجه حرارت، تعویض پانسمان، گرفتن شرح حال و ... به اجزای تشکیل دهنده آن - که قابل مشاهده باشد - تقسیم گردد.
- ۳- هر فرد را در یک زمان مشاهده کنید و مشاهده را به نکات مشخص شده در فهرست بازبینی محدود و منحصر نمایید.
- ۴- برای هر فرد یک فهرست بازبینی جداگانه داشته باشید.
- ۵- بهتر است استادان بدانند که چگونه و چه چیزی را مشاهده نمایند و چگونه رفتار مشاهده شده را ثبت کنند. مسیر مشاهده کننده باید کاملاً مشخص و واضح باشد. برای رفتارهایی که در مورد آنها اطلاعات کافی برای قضاوت صحیح وجود ندارد، خانه های مربوط را تکمیل ننمایید. جدول ۸-۱ بخشی از یک فهرست بازبینی مربوط به معاینه

بیماران (اقتباس شده از طرح تحقیقاتی بررسی عملکرد و آگاهی پزشکان عمومی اصفهان - ۱۳۷۳) را ارائه می‌نماید.

۶- با افزودن خطاهای معمول به ویژه خطاهای غیرقابل اغماض و در نظر گرفتن نمره منفی برای چنین رفتارهایی، فهرست خود را به یک ابزار تشخیصی تبدیل خواهید نمود که می‌توان به وسیله آن، اشکالات موجود در عملکرد یک فرد را مورد شناسایی قرار داد.

۲- مقیاسهای درجه‌بندی^۱

در بسیاری از موارد، فهرستهای بازبینی از کارایی لازم برخوردار نیست. مثلاً در اندازه‌گیری برقراری ارتباط یک عضو تیم مراقبتهای بهداشتی - درمانی با بیماران و سایر اعضای تیم که درجاتی گوناگون از رفتار قابل مشاهده است، نمی‌توان از آنها استفاده نمود. در چنین مواردی که به جای ثبت حضور یا فقدان یک خصیصه، در پی درجه‌بندی وضعیت یا کیفیت چیزی هستیم از مقیاسهای درجه‌بندی استفاده می‌کنیم.

با استفاده از این روش می‌توان اهداف، فرایندها، فرآورده‌ها (بروندها)ی مهم و موردنظر را اندازه گرفت. اما کاربرد آن محدود به اندازه‌گیری حیطه غیرشناختی است. فرایندهایی مانند اندازه‌گیری فشارخون، بخیه محل خونریزی، آندسکوپی، کار با کروماتوگرافی گاز- مایع، پر کردن دندان پوسیده، برقراری ارتباط با بیمار و ... با آزمونهای کتبی قابل اندازه‌گیری نیست و در این موارد از مقیاسهای درجه‌بندی استفاده می‌گردد.

انواع مقیاسهای درجه‌بندی

۱- **مقیاس درجه‌بندی عددی^۲**: این نوع مقیاس ساده‌ترین نوع مقیاسهای درجه‌بندی است. نمره‌دهنده یک عدد را مشخص می‌سازد؛ که مشخص‌کننده درجه حضور ویژگی یا خصیصه موردنظر است. صفت یا خصیصه به صورت یک عبارت بیان می‌شود و مثلاً از یک تا پنج ارزش‌گذاری می‌گردد. گستره^۳ نمره‌دهی به دلخواه قابل تغییر است اما توصیه می‌شود که حداکثر به عدد ده ختم گردد؛ زیرا افتراق قائل شدن بیش از آن مشکل خواهد بود. معمولاً از یک راهنمای نوشتاری برای توصیف و شرح اعداد استفاده می‌شود.

1- Rating Scales
3- Range

2- Numerical Rating Scale

مثال:

غیر قابل قبول (ضعیف) = ۱ زیر حد متوسط = ۲ متوسط = ۳
 بیش از حد متوسط = ۴ برجسته (عالی) = ۵

لصفاً عملکرد دانشجو در هر یک از فعالیتهای زیر را نمره‌دهی نمایید:

۱- نحوه فعالیت دانشجو در پروژه‌های کلاسی

۱ ۲ ۳ ۴ ۵

۲- میزان ارتباط مناسب دانشجو با همکلاسهایش

۱ ۲ ۳ ۴ ۵

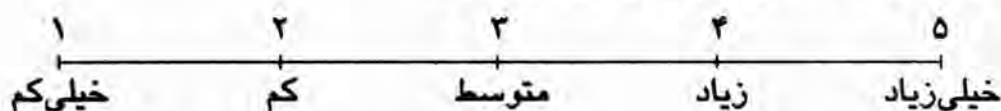
۳- میزان مشارکت دانشجو در بخشها

۱ ۲ ۳ ۴ ۵

۲- مقیاس درجه‌بندی ترسیمی^۱

در این روش شبیه مقیاس درجه‌بندی عددی، نمره‌دهنده به هر خصیصه نمره‌ای اختصاص می‌دهد، اما به جای استفاده از اعداد از قبل تعیین شده، نمره‌دهی با استفاده از یک شکل ترسیمی (نموداری) صورت می‌پذیرد و این شکل ویژه، عبارت از یک خط ممتد است که هر نقطه آن قابل انتخاب باشد. انتخاب نقاط بر این خط محدودیتی ندارد. اما باید روی خط توسط اعداد یا عبارات به نحوی مشخص گردد تا تمام نمره‌دهندگان برداشت یکسانی از نقاط داشته باشند.

میزان مشارکت دانشجو در فعالیتهای فوق برنامه



۳- رتبه‌بندی^۲

در این روش به جای توصیف نوشتاری و یا نمره‌دادن به هر خصیصه، افراد بر حسب وجود آن خصیصه براساس اهمیت موردنظر از زیاد تا کم رتبه‌بندی می‌گردند، البته این کار در صورتی که تعداد دانشجویان یا خصیصه‌ها زیاد شد، مشکل خواهد بود.

پیشنهاد شده که تعداد خصیصه‌های مورد رتبه‌بندی حداکثر به هفت مورد محدود گردد. همچنین تعداد افراد مورد رتبه‌بندی باید محدود گردد، زیرا تعداد زیاد خصیصه‌ها یا افراد، پایایی و روایی اندازه‌گیری را تحت تأثیر قرار می‌دهد. رتبه‌بندی معمولاً وقتی مورد استفاده قرار می‌گیرد که فراورده‌های حاصل از عملکرد دانشجویان بررسی و ارزشیابی شود. البته به عقیدهٔ مِرِن^۱ ارزشیابی محصولات دانشجویان در رده‌های عالی، متوسط و ضعیف بر رتبه‌بندی ارجحیت دارد.

۴- مقیاس دوتایی^۲

در این روش دانشجویان دو تا دو تا مقایسه می‌شوند و ارزشیابی در هر مقایسه تعیین می‌کند کدامیک از آن دو از لحاظ خصیصهٔ موردنظر در رتبهٔ بالاتری قرار دارند تا سرانجام همه دانشجویان پشت سرهم قرار گیرند. این روش نسبت به رتبه‌بندی، نتایج پایاتر و دقیقتری را به بار خواهد آورد؛ بنابراین در مواردی کاربرد بیشتری دارد که نتایج بسیار دقیقی موردنیاز است.

منابع خطا در مقیاسهای درجه‌بندی

معمولاً در مقیاسهای درجه‌بندی چند منبع خطا وجود دارد: خطا ممکن است ناشی از ابهام در خود مقیاس مورد استفاده، خطای شخصی ارزشیاب (نمره‌دهنده)، طبیعت خصیصهٔ مورد ارزیابی و یا فرصتی که ارزشیاب برای رتبه‌بندی در اختیار دارد، باشد. ابهام در مقیاس به عبارات‌بندی و مفاهیم خصیصه موردنظر برمی‌گردد، به خصوص در مواردی که نمره‌دهندگان از کلمات مورد استفاده برداشت یکسانی نداشته و در مصادیق اختلاف‌نظر داشته باشند، نمره‌دهی پایایی و روایی کافی نخواهد داشت. وقتی رفتار قابل قبول یا عالی در برخورد با بیمار تعریف نشده باشد استفاده از چنین مقیاسی مشکل‌آفرین خواهد بود. به راستی عالی چیست؟ خوب کدام است؟ اصلاً آیا میزان عالی و خوب بودن با حد انتظار از دانشجویان در رده‌های تحصیلی مختلف متفاوت نیست؟ به همین جهت کیفیت باید در طبیعت خصیصه مورد نظر تفسیر گردد، سن و توانایی مدنظر واقع شود و مصادیق این مفاهیم مورد توافق نمره‌دهندگان قرار گیرد. احتمالاً مشکل ابهام در خصیصهٔ موردنظر، با تجزیهٔ رفتار به خصیصه‌های منفرد قابل مشاهده و قابل اندازه‌گیری، توصیف قبلی نحوهٔ بروز هر خصیصه و ارائهٔ مثالهایی از

انواع متفاوت رفتار در ارتباط با سطوح مختلف خصیصه کاهش می‌یابد.
 خطاهایی که مربوط به ارزشیاب است شامل خطاهایی مانند اثر هاله‌ای^۱، تمایل به میانگین^۲، تورش فردی، خطای قیاسی^۳ و مانند آن است.
 یک ارزشیاب خوب و ایده‌ال باید توانمند، با کفایت، عینی‌نگر و آموزش دیده در امر استفاده از مقیاس باشد. به هر حال اگر نمره‌های پایا مطلوب است باید تورش فردی کنترل گردد.

راههای بهبود داده‌های ناشی از مقیاسهای درجه‌بندی

برای کاهش خطاهای تصادفی، باید مقیاسها، ارزشیابها و روشها را بهبود بخشید. پیشنهادهای زیر می‌تواند به تحقق این امر کمک نماید:

راههای بهبود مقیاس درجه‌بندی

۱- حیطه رفتارها و خصیصه‌های ویژه‌ای که باید اندازه‌گیری شود باید دقیقاً مورد شناسایی قرار گرفته، تعریف شود. این خصیصه‌ها می‌تواند شامل مهارتهای روانی-حرکتی یا برخی وظایف عملکردی^۴ باشند. مطمئن شوید که آنچه در پی اندازه‌گیری آن هستید از نظر آموزشی مهم است و در دوره آموزشی یا حین انجام آزمون به نحوی رخ می‌دهد که استاد بتواند آنها را مشاهده کند.

۲- خصیصه مورد نظر (مثلاً شستن دست قبل از عمل جراحی) و نقاط مقیاس مورد استفاده را (مثلاً: عالی، خوب، متوسط، ضعیف) به صورت واضح و روشن تعریف کنید. کلمات مورد استفاده باید برای همه نمره‌دهندگان تداعی‌کننده یک معنی واحد باشد. برای رسیدن به چنین وضعیتی بهتر است از تجزیه صفت مورد نظر به موارد کوچکتر تا حد رسیدن به اعمال محدود و اختصاصی استفاده نمایید.

۳- نمونه‌گیری دقیق از رفتارها در قلمرو خصیصه مورد نظر، تعمیم‌پذیر ساختن نتایج را امکان‌پذیر می‌سازد. چرا که نمی‌توان تمام رفتارهای جزئی را برای انجام یک عملکرد کلی فهرست‌بندی و اندازه‌گیری نمود.

۴- خصیصه را به جای استفاده از جملات و عبارات خبری، به صورت سؤالی بیان کنید. برای این کار می‌توانید از کلماتی مانند نحوه، چگونگی، شیوه، تا چه اندازه و ... که

1- Halo Effect
 3- Logical Error

2- Central Tendency Error
 4- Performance Task

نمایان‌کننده کیفیت یا رتبه‌بندی است استفاده نمایید.

۵- برای مقیاسهای نموداری باید امتداد خط، بلافاصله بعد از سؤال (خصیصه‌ای که باید نمره‌دهی شود) باشد و یا زیر سؤال قرار گیرد.

۶- ضوابط مربوط به افتراق بین نمره‌ها و تقسیم خط را قبلاً مشخص سازید. بهتر است خط به سه تا هفت فاصله تقسیم شود؛ اگر چه برخی تقسیم نه‌تایی را کاراتر می‌دانند.

راههای بهبود عملکرد ارزشیابان (نمره‌دهندگان)

۱- یک جلسه آموزشی ترتیب دهید. در این جلسه بر ارزش دقت و صداقت در کار تأکید نمایید. انواع خطاهای احتمالی مربوط به نمره‌دهندگان و راههای جلوگیری یا به حداقل رساندن آنها را شرح دهید.

۲- انگیزه را بین نمره‌دهندگان برای انجام دقیق وظایفشان ایجاد و تقویت نمایید.

۳- افرادی را که می‌توانند نمره‌هایی به دور از تورش و عینی بدهند انتخاب کرده، از انتخاب افراد خیلی مشتاق یا خیلی عیب‌جو بپرهیزید.

راههای بهبود روش نمره‌دهی

روایی نمرات به دست آمده با مراقبت در ساخت مقیاس، مراقبت در انتخاب و آموزش نمره‌دهندگان و استفاده از بیش از یک نمره‌دهنده افزایش می‌یابد. موارد زیر در افزایش روایی آزمون مؤثرند، گرچه به نمره‌دهندگان و مقیاس ارتباطی ندارد:

۱- از میانگین دیدگاههای چند نمره‌دهنده استفاده نمایید. تعدد دفعات نمره‌دهی مستقل و تعدد نمره‌دهندگان، نتایج پایاتری به دست می‌دهد.

۲- هر خصیصه یا ویژگی را در یک زمان نمره‌دهی کنید. مثلاً اگر قرار است پانزده خصیصه درسی دانشجو مورد بررسی قرار گیرد، نخست خصیصه اول را برای همه آنها نمره داده، بعد به خصیصه دوم بپردازید. این امر به نمره‌دهنده امکان بیشتری برای توجه و قضاوت در مورد خصیصه می‌دهد و نتایج اندازه‌گیری را پایاتر می‌کند.

۳- فقط وقتی از مقیاس درجه‌بندی استفاده کنید که ویژگی موردنظر کاملاً به اجزای کوچکتر دسته‌بندی شده، نمره‌دهندگان کاملاً درباره نوع رفتاری توافق نظر داشته باشند که مثلاً نمایان‌کننده رفتار برجسته (عالی) یا متوسط است در غیر این صورت اختلاف نظرهای معنی‌دار در نحوه تقسیم ارزش مقیاسها می‌تواند به یک سری نتایج ناروا و یا با روایی کم منجر می‌شود.

۴- برای کاهش اثر هاله‌ای و جوابهای کلیشه‌ای، جهت مقیاس را تغییر دهید. مثلاً در سؤال

- اول؛ مکان اول، صفت مطلوب و در بعدی صفت نامطلوب باشد.
- ۵- اثر متغیرهای فرعی خارجی را کنترل کنید. به عنوان مثال شرایط زمانی و مکانی را مدنظر قرار دهید.
- ۶- نمره‌دهی را سریعاً پس از مشاهده انجام دهید. تأخیر در نمره‌دهی رفتارهای واقعی، منجر به ثبت ذهنی نمرهٔ آزمون‌شونده (به جای ثبت رفتارهای واقعی) می‌گردد.

آزمونهای سنجش نگرش^۱

این آزمونها به جای آنکه در پی کشف میزان دانسته‌ها یا مهارتهای ایشان باشد توصیف‌کنندهٔ احساس افراد یا نحوهٔ رفتار عادی آنها می‌باشد. نگرشها نشان دهندهٔ آمادگی برای پاسخ‌دهی آشکار و واضح به مباحث اجتماعی است، مستقیماً قابل مشاهده نیست و از رفتار واضح فردی (گفتاری یا غیرگفتاری) استنتاج می‌شوند. حقیقت دیده نمی‌شود، تعصب قابل مشاهده نیست بلکه معمولاً رفتاری که نشان‌دهندهٔ تعصب است مشاهده می‌گردد. بنابراین براساس الگوی رفتاری فرد می‌توان فهمید یک فرد کدام نگرش را دارد. نگرشها به ویژه در بزرگسالان نسبتاً پایدار و ثابت است.

به هر حال نگرشها و اهداف آموزشی در حوزهٔ عاطفی معمولاً بخش لاینفک از برنامه درسی را به خصوص در علوم بهداشتی تشکیل می‌دهد. چنانکه ذکر شد رفتار آدمی براساس ارزشها و نگرشهایی که دارد هدایت می‌شود و از آنجا که اعضای تیم مراقبتهای بهداشتی-درمانی در ارتباط مستقیم با بیمار و یکدیگر هستند و با او به عنوان یک انسان مواجه دارند، حتی بخش عمده‌ای از مراقبتهای اولیه، ثانویه و ثالثیه در گرو این ارتباط صحیح انسانی است؛ بنابراین ارزشیابی نگرشها به ویژه در بخش تربیت نیروی انسانی بهداشت و درمان اهمیت ویژه‌ای می‌یابد.

برای سنجش نگرش دو رویکرد وجود دارد:

- ۱- مشاهده موضوع در یک موقعیت واقعی (یا شبیه‌سازی شده)؛
- ۲- استفاده از مقیاسها و دستنوشته‌های^۲ گزارش شخصی^۳.

انواع مقیاسهای نگرش گزارش شخصی

مقیاسهای نگرش مانند علایق‌سنجها و آزمونهای شخصیت در پی کشف خصیصه‌های مربوط به رفتار آدمی هستند که به مفاهیم انتزاعی این خصیصه‌های مربوط به رفتار آدمی سازه^۱ گویند. بنابراین مقیاسهای نگرش شاخص^۲ و اندازه‌گیری‌کننده سازه^۳ نگرش هستند. سه شیوه برای تهیه این نوع مقیاسها وجود دارد:

۱- مقیاسهای درجه‌بندی تراکمی^۳ یا نوع لیکرت^۴؛

۲- مقیاسهای فاصله‌ای^۵ یا نوع ترستون مانند مقایسه‌های ترستون و ریمرز^۶؛

۳- مقیاسهای انباشته^۷ یا نوع گاتمن^۸.

اگر چه روش تفکیک معنایی^۹ یک نوع از مقیاس نیست، یک روش برای اندازه‌گیری نگرش است.

روش گاتمن بسیار پیچیده و فراتر از اهداف این کتاب است؛ استفاده از روش ترستون هم مشکلات زیادی دارد پس فقط به شرح دو روش لیکرت و تفکیک معنایی بسنده می‌کنیم.

ساخت مقیاس نگرش به روش لیکرت

روش لیکرت معمولترین روش سنجش نگرش است. مقیاس لیکرت به آسانی تهیه شده، آسانتر از مقیاسهای نوع ترستون و ریمرز نمره‌گذاری می‌شود. سایر مزایای آن عبارتست از:

۱- مقایسه‌هایی با یکنواختی مناسب ایجاد می‌کند؛

۲- با کمک آن، درجه یا شدت احساس مشخص می‌شود؛

۳- واریانس نتایج آن بیشتر است، بنابراین به راحتی تفاوتها قابل سنجش است.

برای به دست آوردن یک گروه از جملات با مفاهیم مثبت و منفی می‌توان از دانشجویان خواست تا نظر خود را راجع به یک موضوع، اشخاص کلاس یا شیوه‌های آموزشی در قالب جملاتی بنویسند که درجات متفاوتی از احساس مثبت یا منفی داشته

1- Construct

3- Minnesota Scale Summated Rating Scales

5- Interval Scales

7- Cumulative

9- Semantic Differential Technique

2- Indicator

4- Likert Type

6- Thurstone & Remmers Scales

8- Guttman Type

باشد. به عنوان مثال در مورد نگرش دانشجویان در مورد فیزیولوژی از چنین جملاتی می‌توان استفاده نمود:

فیزیولوژی کسل‌کننده است.

فیزیولوژی یک درس جالب و خواندنی است.

مطالعه فیزیولوژی تلف‌کننده وقت است.

پرسشنامه اولیه باید دارای تعدادی سؤال خنثی باشد، همچنین تعدادی سؤال در انتهای گستره دیدگاهی داشته باشد. سؤالات با مفاهیم مثبت و منفی با هم مخلوط شوند. سپس پرسشنامه اولیه به افراد زیادی داده می‌شود تا در مقیاس پنج‌تایی (از یک تا پنج) به هر سؤال نمره دهند مانند: «قویاً موافقم»، «موافقم»، «نظری ندارم» (یا «نمی‌توانم تصمیم بگیرم»): «مخالقم»، «قویاً مخالفم». سؤالات از یک تا پنج نمره می‌گیرند و نمره کل به دست می‌آید. نمره‌گذاری به صورت اظهارنظر در مورد گزاره‌های با مفاهیم منفی، به عکس حالت فوق انجام می‌گیرد. بدین گونه که مثلاً پاسخ «قویاً موافقم» نمره یک خواهد گرفت.

برای تفسیر نمره هر فرد، می‌توان مجموع نمرات وی را با میانگین مقیاس، مقایسه نمود. اگر نمره آن فرد بالاتر از میانگین باشد، نگرش وی نسبت به موضوع «مثبت» تلقی می‌گردد و در غیر این صورت وی دارای نگرش «منفی» نسبت به موضوع مورد بررسی است. میانگین مقیاس نگرش سنج n درجه‌ای لیکرت با استفاده از رابطه زیر قابل محاسبه است:

$M =$ میانگین مقیاس

$$M = \frac{n(K+1)}{2}$$

$n =$ تعداد درجات مقیاس

$K =$ تعداد اظهارنظرها

برای تجزیه و تحلیل اطلاعات مربوط به نمونه آماری نیز می‌توان از این روش سود جست.

وقتی دامنه دیدگاهها گسترده باشد، تعدد دسته‌ها، اثری بر ثبات درونی نخواهد داشت، اما اگر دامنه نظر افراد به دسته‌های مختلف تقسیم نشود، تعداد دسته‌ها می‌تواند منجر به کاهش پایایی گردد.

پس از تعیین همبستگی^۱ هر سؤال و نمره کل، فقط سؤالاتی که با نمره کل ارتباط بیشتری داشته باشد، در پرسشنامه نهایی باقی می‌ماند. با این روش، پرسشنامه نهایی تهیه

می‌شود که دارای پایایی و ثبات درونی بیشتری خواهد بود.

روش تفکیک معنایی آزگود^۱

این روش، به‌خودی‌خود، یک روش ساختن مقیاس نگرش نیست؛ اما راهی برای اندازه‌گیری نگرش محسوب می‌گردد. برخلاف روش لیکرت یا ترستون که در آن، برای سنجش نگرش از تعدادی جمله در مورد یک موضوع استفاده می‌گردد، در این روش موضوع به تنهایی مطرح شده، با استفاده از مقیاسهای منفک دوتایی (مثل دلسوز- بی‌تفاوت) نمره‌دهی می‌شود. از جمله صفت‌های متضادی که به کار می‌رود می‌توان خوب - بد / شیرین - ترش / زیاد - کم / خوش‌عزه - بدمزه / عادل - ظالم / تلخ - شیرین / خالی - پر / روشن - تیره / نو - کهنه / تمیز - کثیف / خوشحال - غمگین / زشت - زیبا و منصف - بی‌انصاف را نام برد.

در انتخاب این لغات باید حتماً به سن و وضعیت پاسخ‌دهندگان توجه نمود و باید آنها معانی این کلمات را بدانند.

مثال:

سقط جنین

بد: _____ : _____ : _____ : _____ : _____ : _____ : _____ : _____ : _____ : _____
 سعادت: _____ : _____ : _____ : _____ : _____ : _____ : _____ : _____ : _____ : _____
 تنفر: _____ : _____ : _____ : _____ : _____ : _____ : _____ : _____ : _____ : _____

برای اصلاح و بهبود یک روش تفکیک معنایی به نکات زیر توجه کنید:

- ۱- مفهومی را که باید نمره‌دهی شود مشخص کنید. اگر مفاهیم بر طبق اهداف مختلف است هر مفهوم به یکی از اهداف مربوط باشد.
 - ۲- مقیاس مناسب دوتایی را انتخاب کنید، چرا که ما به موضوعات ذهنی بیشتر علاقه داریم تا مفاهیم توصیفی.
 - ۳- یک برگه پاسخ برای نمره‌دهی به هر مفهوم (مثلاً سقط جنین) تهیه کنید به نحوی که مقیاسها در زیر آن قرار گرفته باشد.
- الف) هر مفهوم را در یک صفحه بیاورید.

ب) مقیاسهای مطلوب را جابه‌جا کنید به نحوی که مسیر مثبت و منفی، هر بار متفاوت باشد.

ج) اگر چه نقاط مقیاس می‌تواند بین سه تا نه باشد، مقیاس هفت نقطه‌ای کارآیی بیشتری دارد.

د) برای هر مقیاس، ارزش عددی تعیین کنید، سپس براساس نقاط انتخاب شده، نمره کل را برای هر مفهوم به دست آورید.

رهنمودهای کلی برای ساخت مقیاسهای نگرش

در تهیه مقیاسهای نگرش توجه به نکات زیر می‌تواند راهگشا باشد:

۱- به جای استفاده از زمان گذشته از جملاتی با زمان حال استفاده کنید. توجه کنید که شما معمولاً نگرش را نسبت به چیزی می‌سنجید که هم‌اکنون وجود دارد نه چیزی که ارزش تاریخی دارد.

۲- جملات را واضح و با زبان ساده بنویسید. از جملات ساده به جای جملات پیچیده استفاده کنید.

۳- تا حد امکان از لغاتی مانند «فقط»، «حالا»، «انحصاراً» و «تنها» استفاده نکنید.

۴- هر جمله را خلاصه و ترجیحاً در کمتر از قابلیت کلمه بنویسید.

۵- از ذکر جملات مبهم که به طرق مختلف تعبیر می‌شود پرهیزید.

۶- از اشاره به یک جواب ویژه اجتناب کنید. توجه داشته باشید که جامعه پژوهش همواره سعی در راضی کردن محققان دارد. دانشجویان شما هم اگر نظر قطعی شما را بدانند سعی می‌کنند خود را موافق شما نشان دهند.

۷- هر جمله باید یک بعد و یک جنبه مربوط به یک مفهوم را در برداشته باشد. از ذکر دو جمله پیاپی مانند: اگر ... آنگاه ... دروی کنید.

۸- از جملاتی که همگان تأیید یا رد می‌کنند پرهیز کنید. به خاطر داشته باشید شما سعی می‌کنید تفاوت نگرش دانشجویان را بسنجید. جملاتی را بنویسید که حدس می‌زنید نیمی از افراد با آن موافق یا مخالفند. بنابراین جمله «من دوست دارم با مردمان نیک‌سرشت ارتباط داشته باشم». یک سؤال خوب نگرش‌سنج نیست.

۹- سعی کنید تعداد سؤالات با مفاهیم مثبت و منفی برابر باشد. این کار جواب دادن قالبی یا راضی‌کننده را به حداقل می‌رساند.

- ۱۰- جملات را به صورت تصادفی در کل پرسشنامه پخش کنید، به نحوی که بیش از چهار تا پنج جمله با مفاهیم مثبت یا منفی (له یا علیه موضوع نگرش) پشت سرهم نباشند.
- ۱۱- سوالات حساس باید در وسط پرسشنامه قرار گیرند.
- ۱۲- خودتان را به ۳ انتخاب محدود نکنید، هر چند بهتر است بیش از ۷ نقطه نداشته باشید.
- ۱۳- خودتان را به مقیاس «موافق- مخالف» محدود نکنید. مثلاً در مورد خصیصه‌ای مانند «ارزش دوره» بهتر است از مقیاس «ارزشمند - بی‌ارزش» استفاده کنید. احساسات دلپذیر، جالب و ارزشمند دو وجهی برای گستره مقیاسهای انتخاب مناسبند هانند: «راضی - ناراضی»؛ «مهم - بی‌اهمیت».

مراحل طراحی یک پرسشنامه ایده‌آل

در پایان این فصل، به مراحل طراحی یک پرسشنامه ایده‌آل با استفاده از مقیاسهای ذکر شده، اشاره می‌گردد. انجام مراحل ذیل به شما کمک می‌کند، تا نتایج قابل اعتمادی بدست آورید:

۱- ارائه یک تعریف محدود و واضح از سازه مورد نظر. در این مرحله با استفاده از مرور متون و نظر کارشناسان در مورد جلوه‌های رفتاری یا نگرشی سازه مورد نظر اطمینان لازم را کسب کنید.

۲- اجرای یک پرسشنامه باز پاسخ: با استفاده از پرسشنامه‌های بازپاسخ در نمونه‌ای محدود از جمعیت مورد مطالعه می‌توان نظر قبلی در مورد تعریف محدوده و حیطه‌های مورد اندازه‌گیری را تکمیل نمایید. این امر به شما کمک می‌کند تا آنچه را به علت وضعیت متفاوت زیستی، فرهنگی، اجتماعی، اقتصادی، مذهبی از نظر شما به دور مانده است شناسایی نمایید. علاوه بر آنکه می‌تواند گزاره‌هایی را به شما ارائه کند که می‌توان مستقیماً در مقیاس لیکرت از آنها بهره گرفت.

۳- انتخاب مقیاس و نوشتن گزاره‌ها: مقیاس مناسب را با توجه به محدودیت‌های اجرایی و آنچه در پی آن هستید و شیوه‌هایی که اطلاعات به دست آمده به شما در تصمیم‌گیری یاری خواهد رساند، انتخاب نمایید. مثلاً اگر قرار است از نتایج برای اصلاح کاستیهای موجود در یک ساختمان یا سازمان یا سازمان یا فرد استفاده شود احتمالاً فهرستهای بازبینی مناسب خواهد بود. به هر حال پس از انتخاب مقیاس از حیطه‌های مورد نظر که همان جلوه‌های رفتاری یا نگرشی مورد توافق است نمونه‌گیری کرده، ابعاد مهم را در

- پرسشنامه ذکر نمایید. چنانکه ذکر شد برای مقیاس نگرش سنج لیکرت مانند «قویاً موافقم» تا «قویاً مخالفم» می‌توان عیناً از جملات جامعه مورد پژوهش استفاده نمود.
- ۴- تأیید روایی محتوا و روایی ظاهری: با استفاده از مطالب گفته شده در فصلهای آینده روایی پرسشنامه را از لحاظ ظاهر و محتوا مورد تأیید قرار دهید.
- ۵- اجرای مقدماتی: پرسشنامه تهیه شده را حداقل در ده درصد حجم نمونه موردنظر به اجرا گذارید. با این کار علاوه بر آنکه اشکالات ظاهری پرسشنامه و نکات مبهم آن از دید جامعه مورد پژوهش مشخص خواهد شد، می‌توانید به محاسبه پایایی پرسشنامه نیز اقدام کنید. سؤالاتی را که از همبستگی کافی با کل پرسشنامه برخوردار نیستند یا قدرت تمیز کافی ندارند (اغلب پاسخ‌دهندگان با آن موافق یا مخالفند) یا وجودشان باعث کاهش پایایی پرسشنامه می‌شوند با توجه به روایی محتوا حذف یا اصلاح کنید.
- ۶- اجرای نهایی: پس از اطمینان از روایی و پایایی مناسب پرسشنامه، آن را در نمونه موردنظر اجرا کنید.

منابع:

- ۱- جمشیدی، حمیدرضا. بررسی عملکرد حرفه‌ای و آگاهی علمی پزشکان عمومی شهر اصفهان. طرح پژوهشی ۷۳۰۱۲ معاونت پژوهشی دانشگاه علوم پزشکی و خدمات بهداشتی درمانی استان اصفهان، ۱۳۷۴.
- 2- Abbatt F, McMahon R. *Teaching Health-care Workers*. Hong Kong: Macmillan Education, 1988.
- 3- Bowd SB. *Teaching in the health-related professions*. Durham: Carolina Academic Press, 1997.
- 4- Brow G, Bull J, Penlebury M. *Assessing student learning in higher education*. London: Routledge, 1997.
- 5- Cox KR, Ewan CE. *The medical teacher*. 2nd ver.ed. London: Churchill Livingstone. 1988.
- 6- Gronlund NE. *How to construct achievement tests*. 4th ver.ed. Newjersey: Prentice-hall, 1988.
- 7- Henerson ME, Morris LL, Fitz-Gibbon CT. *How to measure attitudes*. New York: SAGE Publications, 1987.
- 8- Mehrans WA, Lehman IJ. *Measurment and evaluation in education and psychology*. 3rd ver.ed. New York: Holt-Saunders: 1984.
- 9- Neufeld VR, Norman GR. *Assessing clinical competence*. New York: Springe Publishing Company, 1985.
- 10- Newble D, Cannon R, Kapelis Z. *A handbook for medical teachers*. 2nd ver.ed. Lancaster: MTP Press, 1987.

بخش سوم

فراوری آزمون

فصل نهم

پایایی و روایی

- پایایی
- روایی

- یک آزمون خوب، به عنوان وسیله‌ای قابل اعتماد برای نیل به اهداف از پیش تعیین شده، باید دارای ویژگیهای زیر باشد:
- ۱- پایایی: بیان‌کننده یکسانی نتایج آزمون در اجراهای متعدد است.
 - ۲- روایی^۱: نشان‌دهنده قابلیت آزمون در اندازه‌گیری همان متغیری است که هدف واقعی از طراحی آزمون بوده است.
 - ۳- عینی بودن^۲: نشان‌دهنده میزان هماهنگی امتحان‌گیرندگان مختلف در تعیین پاسخ صحیح یک سؤال است.
 - ۴- عملی بودن^۳: بیان‌کننده قابلیت اجرای آزمون در مراحل تهیه سؤالات، اجرا، تصحیح و تفسیر نتایج است.
- در این فصل دو مبحث بسیار مهم پایایی و روایی آزمونها به تفصیل مورد توجه قرار می‌گیرد.

پایایی آزمون

- زمانی یک آزمون از پایایی برخوردار است که نمرات آن در شرایط مختلف زمانی و مکانی، ثبات و تشابه داشته باشد. یعنی:
- الف) مصححین مختلف نمره یکسانی به یک برگه واحد اختصاص بدهند. بخصوص، تفاوت بیش از ۱۰ درصد در نمره‌گذاری آزمون تشریحی گسترده پاسخ^۴، نشان‌دهنده اختلاف نظر آزمون‌گیرندگان مختلف است. در این حالت برگزاری جلسات بحث و گفتگو برای رسیدن به نتیجه یکسان ضروری است.
- ب) یک آزمون‌گیرنده در زمانهای مختلف، نمره یکسانی، به یک برگه امتحانی اختصاص دهد.
- ج) آزمون شونده در زمانهای مختلف، از یک آزمون واحد نمره یکسانی بگیرد و موقعیت

1- Validity
3- Practibility

2- Objectivity
4- Long Essay Question(LEQ)

فرد نسبت به سایر افراد شرکت‌کننده در آزمون ثابت بماند. برای محاسبه پایایی آزمون از شاخصی به نام ضریب پایایی (ضریب اعتبار) استفاده می‌شود. از آنجا که برای چنین محاسباتی به طیف وسیعی از نمرات نیاز است، محاسبه پایایی در آزمونهای ملاک نسبی، کاربرد بیشتری دارد؛ چرا که پراکندگی نمرات در این آزمونها بیشتر است.

برای محاسبه ضریب پایایی ابتدا باید ضریب همبستگی را تعیین نمود. بنابراین قبل از بیان روشهای محاسبه پایایی، ضریب همبستگی توصیف می‌گردد.

ضریب همبستگی^۱

همبستگی مفهومی مرتبط با پایایی است؛ زیرا با توجه به میزان همبستگی می‌توان به میزان پایایی یک آزمون پی برد. ضریب همبستگی، میزان ارتباط بین دو اندازه‌گیری مختلف در یک گروه واحد از افراد است. مثلاً: ارتباط بین نتایج دو آزمون بافت‌شناسی و آسیب‌شناسی یک گروه از دانشجویان و یا ارتباط نتایج آزمون بافت‌شناسی در دو زمان مختلف.

روش محاسبه ضریب همبستگی

رایجترین روش برای تعیین ضریب همبستگی، استفاده از فرمول همبستگی گشتاوری پیرسون است که مطلب را با یک مثال بیان می‌کنیم:

فرض کنید نمرات بافت‌شناسی یک گروه پنج نفری از دانشجویان را با X و نمرات آسیب‌شناسی آنها را با Y نشان دهیم. برای محاسبه ضریب همبستگی، می‌توانید ابتدا میانگین نمرات دانشجویان را در هر دو درس محاسبه کرده (\bar{Y}, \bar{X}) سپس فاصله هر نمره را از میانگین مشخص کنید. حاصل این محاسبه، نمره انحراف نامیده می‌شود. (نمره انحراف بافت‌شناسی = x و نمره انحراف آسیب‌شناسی = y). حال نمرات انحراف را مجذور کرده (x^2, y^2) ، میانگین آنها را بدست آورید. همچنین حاصلضرب نمرات انحراف را محاسبه کرده (xy) ، با استفاده از فرمول زیر ضریب همبستگی را به دست آورید:

$$r_{xy} = \frac{\overline{xy}}{\sqrt{\overline{x^2} \overline{y^2}}}$$

دانشجو	X	Y	x	y	x ²	y ²	xy
A	۸	۶	۱۸-۸=۱۰	۱۵-۶=۹	۱۰۰	۸۱	۹۰
B	۲۲	۱۸	۱۸-۲۲=-۴	۱۵-۱۸=-۳	۱۶	۹	۱۲
C	۲۰	۲۲	۱۸-۲۰=-۲	۱۵-۲۲=-۷	۴	۴۹	۱۲
D	۲۲	۱۵	۱۸-۲۲=-۴	۱۵-۱۵=۰	۱۶	۰	۰
E	۱۸	۱۴	۱۸-۱۸=۰	۱۵-۱۴=۱	۰	۱	۰
میانگین	۱۸	۱۵			۲۷/۲	۲۸	۲۳/۲

$$r_{xy} = \frac{23/2}{\sqrt{27/2 \times 28}} = 0/84$$

علاوه بر این روش، ضریب همبستگی را می‌توان با استفاده از نمرات خام به وسیله فرمول محاسبه ضریب همبستگی گشتاوری پیرسون بیان کرد:

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

که در این رابطه:

N = تعداد دانشجویان؛

$\sum X$ = مجموع نمرات خام گروه X (نمرات دانشجویان در درس بافت‌شناسی)؛

$\sum Y$ = مجموع نمرات خام گروه Y (نمرات دانشجویان در درس آسیب‌شناسی)؛

$\sum XY$ = مجموع حاصلضربهای Y و X؛

$\sum Y^2$ = مجموع مجذورات Y؛

$(\sum Y)^2$ = مجذور حاصل جمع Yها؛

$\sum X^2$ = مجموع مجذورات X؛

$(\sum X)^2$ = مجذور حاصل جمع Xها؛

مثال:

دانشجو	X	Y	XY	X ²	Y ²
A	۸	۶	۴۸	۶۴	۳۶
B	۲۲	۱۸	۳۹۶	۴۸۴	۳۲۴
C	۲۰	۲۲	۴۴۰	۴۰۰	۴۸۴
D	۲۲	۱۵	۳۳۰	۴۸۴	۲۲۵
E	۱۸	۱۴	۲۵۲	۳۲۴	۱۹۶
Σ	۹۰	۷۵	۱۴۶۶	۱۷۵۶	۱۲۶۵

$$r_{xy} = \frac{5 \times 1466 - (90 \times 75)}{\sqrt{[(5 \times 1756) - 8100][(5 \times 1265) - 5625]}} = \frac{7330 - 6750}{\sqrt{680 \times 700}} = 0/84$$

دامنه تغییرات ضریب همبستگی بین ۱- تا ۱+ است.

- ضریب همبستگی ۱+ نشان می‌دهد که دو آزمون ۱۰۰ درصد با هم ارتباط مستقیم دارد. به عنوان مثال: فردی که در یکی از دو آزمون، رتبه اول را کسب کرده، در آزمون دیگر نیز همین رتبه را به دست آورده است.

- ضریب همبستگی صفر بیان می‌کند که بین دو آزمون هیچ ارتباطی وجود ندارد. به بیان دیگر نتایج یک آزمون نمی‌تواند نتایج آزمون دیگر را پیش‌بینی کند.

- ضریب همبستگی ۱- به معنی ارتباط معکوس بین دو آزمون است؛ به عنوان مثال: نفر رتبه اول در یک آزمون در آزمون دیگر رتبه آخر را کسب کرده است.

البته دو حد ۱+ و ۱- حالات مطلق بوده، هیچوقت اتفاق نمی‌افتد.

هر چه قدر مطلق ضریب همبستگی بین دو متغیر بیشتر باشد، مقدار رابطه بین آن دو بیشتر است؛ مثلاً: میزان ارتباط در ضریب همبستگی ۰/۷ از ضریب همبستگی ۰/۶ بیشتر و ارتباط در ضریب همبستگی ۰/۸۶- از هر دو بیشتر است. البته ضریب‌های همبستگی منفی نشان دهنده رابطه معکوس است.

ضریب همبستگی هیچگاه به صورت درصد تعبیر نمی‌شود ولی مجذور ضریب همبستگی را می‌توان به عنوان درصد وجه اشتراک بین دو متغیر به کار برد، به عنوان مثال: اگر ضریب همبستگی دو آزمون ۰/۸۴ باشد می‌توان گفت که در ۷۰/۵ درصد نتایج دو آزمون مشابه است (۰/۸۴ × ۰/۸۴ = ۷۰/۵٪).

روشهای تعیین پایایی

۱- روش بازآزمایی یا آزمون - پی آزمون^۱: در این روش باید آزمون را برای گروه معینی از دانشجویان در دو مقطع زمانی مختلف تکرار کرد. فاصله بین دو اجرای آزمون متغیر است (از چند دقیقه تا چند سال). در صورتی که از نتایج آزمون در تصمیم‌گیریهای اساسی استفاده می‌شود، باید فاصله دو نوبت آزمون را به طور معقول انتخاب کرد. اگر فاصله زمانی بین دو آزمون کم باشد، آزمون شوندگان پاسخهای خود را از آزمون قبلی به یاد دارند و پایایی آزمون به طور کاذب افزایش می‌یابد. اگر فاصله زمانی بیش از حد زیاد باشد نحوه نگرش و میزان معلومات امتحان‌شوندگان در این مدت تغییر کرده، پایایی به طور کاذب کاهش خواهد یافت. گفته شده مناسبترین فاصله زمانی بین دو اجرای آزمون، یکماه است. به هر حال هنگام گزارش میزان پایایی به دست آمده از این طریق، حتماً باید فاصله زمانی ذکر گردد. ضریب همبستگی بین دو بار اجرای یک آزمون واحد، ضریب پایایی آن را نشان می‌دهد و نشان‌دهنده پایایی پایداری^۲ آزمون است.

۲- روش موازی^۳: در این روش باید دو دسته سؤال بر مبنای یک جدول مشخصات آزمون طرح شده، سپس این دو دسته سؤال در یک جلسه آزمون به یک گروه واحد داده شود تا به آن پاسخ گویند. ضریب همبستگی بین این دو آزمون محاسبه خواهد شد و به عنوان شاخص پایایی برای هر یک از دو دسته سؤال مدنظر قرار خواهد گرفت.

۳- روش دو نیمه کردن آزمون^۴: در این شیوه، به علت وجود محدودیت‌های اجرایی روش قبلی، آزمون فقط یک بار اجرا می‌شود. پس از اجرای آزمون، سؤالات آن را به دو گروه فرد و زوج تقسیم کرده، نمره هر دانشجو در این دو دسته به طور مجزا تعیین می‌شود. هر دو دسته سؤالات باید از روی یک جدول مشخصات امتحان طرح شود تا درجه دشواری و سطوح یادگیری یکسانی داشته باشد. در مرحله بعد، ضریب همبستگی بین نمرات دو نیمه آزمون محاسبه خواهد شد. این ضریب همبستگی دقیقاً معادل ضریب پایایی برای هر یک از دو نیمه آزمون است. پس از محاسبه ضریب همبستگی دو نیمه آزمون می‌توان با استفاده از رابطه اسپیرمن - براون^۵ پایایی کل آزمون را محاسبه نمود:

$$\text{ضریب همبستگی بین دو نیمه آزمون} \times 2 = \text{پایایی کل آزمون} \\ \text{ضریب همبستگی دو نیمه آزمون} + 1$$

1- Test Re-test

2- Stability Reliability

3- Parallel

4- Split Half

5- Spearman - Brown

به عنوان مثال: نمرات پنج دانشجو را در یک آزمون پنجاه سوالی درس انگل شناسی در دو دسته سوالات فرد (X) و سوالات زوج (Y) محاسبه کنید:

دانشجو	X	Y	X ²	Y ²	XY
A	۱۵	۱۶	۲۲۵	۲۵۶	۲۴۰
B	۲۲	۲۲	۵۷۶	۴۸۴	۵۲۸
C	۲۰	۲۰	۴۰۰	۴۰۰	۴۰۰
D	۲۱	۱۹	۴۴۱	۳۶۱	۳۹۹
E	۱۷	۱۸	۲۸۹	۳۲۴	۳۰۶
Σ	۹۷	۹۵	۱۹۳۱	۱۸۲۵	۱۸۷۳

ضریب همبستگی نمرات دو نیمه آزمون:

$$r_{xy} = \frac{5 \times 1873 - (97 \times 95)}{\sqrt{[(5 \times 1931) - (97)^2][5 \times 1825 - (95)^2]}} = \frac{150}{\sqrt{246 \times 100}} = 0.95$$

$$\text{پایایی کل آزمون} = \frac{2 \times 0.95}{1 + 0.95} = 0.97$$

البته پس از دو نیمه کردن آزمون می‌توان از ضریب پایایی گاتمن^۱ نیز استفاده نمود:

$$r_G = \frac{2(St^2 - St_1^2 - St_2^2)}{St^2}$$

که در آن:

r_G = ضریب پایایی گاتمن؛ St^2 = واریانس نیمه اول؛

St_1^2 = واریانس هر دو نیمه آزمون؛ St_2^2 = واریانس نیمه دوم می‌باشد.

به هر حال ضریب پایایی که از این روش به دست می‌آید، نشان‌دهنده کمیته مقدار همسانی درونی^۲ سوالات آزمون است.

۲- روش کودر - ریچاردسون^۳: در این روش آزمون فقط یک بار اجرا می‌گردد و منظور از

1- Guttman's Coefficient

2- Internal Consistency

3- Kuder -Richardson Estimates of Reliability

تعیین پایایی در این روش، ارزیابی همسانی درونی کل آزمون است. در این روش می‌توان از رابطه کودر ریچاردسون ۲۰ و ۲۱ استفاده نمود:

$$KR_{r.} = \frac{n}{n-1} \left[1 - \frac{\sum pq}{s^2} \right]$$

تعداد سؤالات = n :

نسبت پاسخ درست در هر سؤال = p :

نسبت پاسخ غلط در هر سؤال = q :

انحراف معیار آزمون = s

مثال: نمرات پنج دانشجوی A, B, C, D, E در پنج سؤال منظور شده است. به پاسخ صحیح، نمره یک و به پاسخ غلط، نمره صفر تعلق گرفته است.

	A	B	C	D	E	p	q	pq
۱	۱	۰	۱	۱	۱	۰/۸	۰/۲	۰/۱۶
۲	۱	۰	۱	۱	۱	۰/۲	۰/۸	۰/۱۶
۳	۱	۰	۱	۱	۱	۰/۴	۰/۶	۰/۲۴
۴	۱	۰	۱	۱	۱	۰/۶	۰/۴	۰/۲۴
۵	۱	۱	۱	۱	۱	۱	۰	۰
x = نمره دانشجو	۲	۳	۴	۵	۳			

$$SD = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}$$

در این مثال، ابتدا انحراف معیار را با توجه به رابطه

محاسبه کنید.

$$SD = \frac{۶۳ - \frac{۲۸۹}{۵}}{۵} = ۱/۰۴$$

ضریب پایایی با توجه به فرمول اینچنین بدست می‌آید:

$$KR_{r.} = \frac{۵}{۴} \left[1 - \frac{۰/۸۸}{۱/۰۸} \right] = ۰/۲۳$$

به خاطر اینکه محاسبه نسبت پاسخهای صحیح در هر سؤال، بخصوص در آزمونهای بلند، خسته‌کننده است، رابطه کودر ریچاردسون بیشتر مورد استفاده قرار می‌گیرد.

این رابطه بدین قرار است:

$$KR_{\tau_1} = \frac{n}{n-1} \left[1 - \frac{m(n-m)}{ns_n^2} \right]$$

تعداد سؤالات آزمون = n:

میانگین نمرات دانشجویان = m:

انحراف معیار آزمون = S_n .

مثال: اگر در یک آزمون ۴۰ سؤالی، انحراف معیار برابر ۳ و میانگین نمرات دانشجویان در این درس ۳۰ باشد، ضریب پایایی برابر خواهد بود با:

$$KR_{\tau_1} = \frac{40}{40-1} \left[1 - \frac{30(40-30)}{40 \times 9} \right] = \frac{40}{39} \times \frac{60}{360} = 0/17$$

۵- پایایی بین نمره‌دهندگان^۱: برای مقایسه نحوه نمره‌دهی چند مشاهده‌کننده یا نمره‌دهنده متفاوت، از این روش استفاده می‌شود. در واقع نمره‌دهندگان متفاوت یک آزمون ذهنی یا عملی یا شفاهی باید تا حد امکان توافق داشته باشند وگرنه نتایج آزمون اعتبار چندانی نخواهد داشت. برای محاسبه این پایایی از محاسبه ضریب همبستگی بین نمره‌دهندگان استفاده می‌شود. این ضریب در یک آزمون مناسب و خوب نباید کمتر از ۰/۷ باشد.

پایایی در آزمونهای ملاک مرجع

از آنجا که در آزمونهای ملاک مرجع، میان نمرات، پراکندگی وجود نداشته یا بسیار کم است، روشهای ذکر شده در تعیین پایایی کاربرد کمتری دارند و اغلب برای تعیین پایایی این آزمونها می‌توان از روش زیر استفاده کرد:

از آنجایی که هدف از برگزاری این آزمونها تشخیص افراد دارای تسلط و بدون تسلط می‌باشد؛ برای این منظور لازم است دو امتحان موازی برگزار نمایید و پایایی را براساس درصد افرادی که در هر دو امتحان تسلط داشته یا نداشته‌اند معین کنید.

$$100 \times \frac{\text{تعداد افراد بدون تسلط در هر دو امتحان} + \text{تعداد افراد دارای تسلط در هر دو امتحان}}{\text{تعداد کل افراد}} = \text{پایایی}$$

البته این روش دو محدودیت دارد: اول آنکه دو نوع سؤال موازی مورد نیاز است و دوم، حد پایایی مطلوب در این روش مشخص نیست. به سبب همین محدودیتها معمولاً از برآورد پایایی در این آزمونها صرف نظر و بر اطمینان از روایی تکیه می‌گردد.

عوامل مؤثر بر پایایی

عوامل چندی بر پایایی مؤثرند، که می‌توان با توجه به آنها ضریب پایایی آزمون را افزایش داد. در این قسمت این عوامل ذکر می‌گردد.

- ۱- تجانس آزمون‌شوندگان: هر چه آزمون‌شوندگان از نظر توانایی مورد آزمون تفاوت بیشتری داشته باشند، ضریب پایایی بیشتر خواهد بود.
- ۲- تعداد آزمون‌شوندگان: ضریب پایایی با افزایش تعداد آزمون‌شوندگان بیشتر می‌گردد.
- ۳- تعداد گزینه‌ها: در صورت افزایش تعداد گزینه‌ها، ضریب پایایی افزایش می‌یابد.
- ۴- طول آزمون: با افزایش تعداد سؤالات، به علت افزایش حیطه‌های مورد اندازه‌گیری، ضریب پایایی بیشتر می‌شود. با استفاده از فرمول زیر که نتیجه از فرمول اسپیرمن-براون می‌توان اثر افزایش طول آزمون را بر ضریب پایایی محاسبه نمود:

$$r'_{11} = \frac{kr_{11}}{1 + (k-1)r_{11}}$$

r'_{11} = پایایی آزمون جدید؛
 r_{11} = پایایی آزمون اولیه؛
 k = نسبت تعداد سؤالات آزمون جدید به آزمون اولیه.

به عنوان مثال: اگر پایایی یک آزمون پنجاه سؤالی ۰/۷ بوده، تعداد سؤالات آزمون را به هفتاد افزایش دهیم، پایایی جدید آن بدینگونه محاسبه می‌شود:

$$r'_{11} = \frac{1/4 \times 0/7}{1 + 0/7 \times 0/4} = 0/76$$

تفسیر ضریب پایایی

ضریب پایایی به صورت نظری بین ۰ تا ۱+ تغییر می‌کند ولی در محاسبه، ضریب پایایی می‌تواند منفی هم به دست آید. اگر ضریب همبستگی منفی به دست آید، ضریب پایایی، صفر در نظر گرفته می‌شود. عدد صفر به معنی فقدان ارتباط بین دو آزمون و عدد

یک نشاندهنده ارتباط کامل آنهاست. در عمل، هیچگاه ضریب پایایی یک نخواهد بود. ضریب پایایی قابل قبول، عددی بیش از ۰/۷ است، در حالی که در بسیاری از آزمونهای پیشرفت تحصیلی به علت مطلوبیت طراحی و طولانی‌تر بودن آنها، ارقامی بیش از ۰/۹ مورد نظر هستند. از سوی دیگر اعداد بدست آمده از روشهای مختلف، یکسان نخواهد بود؛ برای ارائه گزارش باید روش محاسبه ضریب پایایی ذکر گردد.

مثال: معمولاً KR_{ρ_1} نسبت به ضریب پایایی اسپیرمن- براون حد پایین‌تری را از ضریب پایایی نشان می‌دهد چون روش KR_{ρ_1} براساس هماهنگی جوابهای دانشجویان به تک تک سؤالات است.

به هر حال پس از برآورد پایایی آزمون، می‌توان به راحتی با استفاده از نرم‌افزارهای آماری (مانند SPSS، STATISTICA) به تجزیه و تحلیل سؤالات آزمون براساس ضریب پایایی پرداخت و سؤالاتی را که حذف آنها موجب افزایش پایایی کل آزمون می‌گردد، شناسایی نمود. سپس با توجه به اولویت روایی آزمون به ویژه روایی محتوا، اقدام به اصلاح و تجدیدنظر در گزینه‌ها و احتمالاً حذف برخی از سؤالات نمود.

خطای معیار اندازه‌گیری^۲

خطای معیار اندازه‌گیری با توجه به ضریب پایایی محاسبه می‌شود. بنا بر تعریف، خطای معیار عبارت است از «انحراف معیار توزیع نمرات خطا». به عبارت دیگر این متغیر، میزان خطایی را که در تفسیر نمرات افراد باید در نظر گرفت، مشخص می‌کند. خطای معیار از رابطه زیر بدست می‌آید:

$$\text{خطای معیار} = SD \sqrt{1 - r}$$

SD = انحراف معیار؛ r = ضریب پایایی.

مثلاً: در یک امتحان سی سؤالی با ضریب پایایی ۵۰ درصد و انحراف معیار ۴، چنین خواهیم داشت:

$$\text{خطای معیار} = 4 \times \sqrt{1 - \frac{50}{100}} = 4 \times 0.7 = 2.8$$

خطای معیار مشخص می‌کند که به نمره هر دانشجو چند نمره باید اضافه یا کم کنیم تا حد معقولی برای برآورد نمره حقیقی او به دست آید. در مثال بالا، اگر دانشجویی در

امتحان نمره ۱۷ بدست آورد با توجه به خطای معیار، نمره امتحان او بین $۱۷ \pm ۲/۸$ قرار می‌گیرد. به این ترتیب به تفاوت‌های جزئی نمرات کمتر اهمیت داده می‌شود. علاوه بر رابطه بالا از جدول دیدریش نیز برای تعیین خطای معیار براساس تعداد پرسشها می‌تواند استفاده کرد که اعداد آن بسیار نزدیک به اعداد رابطه بالاست.

جدول دیدریش:

تعداد پرسشها	کمتر از ۲۴	۲۴-۲۷	۲۸-۸۹	۹۰-۱۰۹	۱۱۰-۱۲۹	۱۳۰-۱۵۰
خطای معیار	۲	۳	۴	۵	۶	۷

با توجه به جدول دیدریش هر اندازه بر تعداد سوالات آزمون افزوده شود، مقدار نسبی خطای معیار کاهش می‌یابد. مثلاً: در آزمون پنجاه سوالی، خطای معیار چهار و در آزمون صد سوالی پنج است. یعنی با وجود دو برابر شدن طول آزمون مقدار خطای معیار تنها $\frac{1}{۳}$ افزایش یافته است.

محاسبه خطای معیار با استفاده از ضریب پایایی بسیار سودمند است. ضریب پایایی امکان مقایسه انواع مختلف آزمونها را می‌دهد و خطای معیار برای تفسیر نمرات افراد به کار می‌رود. لازم به یادآوری است که هر چه میزان خطای معیار آزمون کوچکتر باشد ضریب پایایی آن بزرگتر خواهد بود.

در پایان ذکر این نکته ضروری است که با استفاده از نرم‌افزارهای آماری موجود بویژه SPSS محاسبه حداکثر ضرایب همبستگی و پایایی به راحتی امکان‌پذیر است.

برآورد نمره واقعی با استفاده از ضریب پایایی

نمره‌ای که هر داوطلب از یک آزمون به دست می‌آورد حاصل تأثیر خطا و نمره واقعی او (نمره‌ای که اگر هیچ خطایی وجود نداشت، وی کسب می‌نمود) می‌باشد. عوامل خطا به دو دسته تقسیم شده‌اند:

الف- خطای خوش‌شانسی که معمولاً برای افراد با نمرات بالا اتفاق می‌افتد؛

ب- خطای بدشانسی که بیشتر شامل حال افراد با نمرات پایین می‌گردد.

در حالاتی که خوش‌شانسی نقش زیادی داشته باشد و تمام موارد حدس یک فرد درست باشد، می‌توان فرض کرد که نیمی از نمره به دست آمده حاصل حدس دانشجویست.

در آزمونهایی که پایایی کمتری دارند، تفاوت نمره واقعی و نمره به دست آمده زیاد است. برای برآورد نمره واقعی می‌توانید از رابطه‌ای استفاده کرد که نانالی^۱ (۱۹۶۷) پیشنهاد کرده است:

$$X = \bar{X} + (r_{xx})(X' - \bar{X})$$

که در آن X برآوردی از نمره واقعی، X' نمره کسب شده، \bar{X} نمره میانگین، r_{xx} ضریب پایایی می‌باشد.

روشن است در حالاتی که ضریب پایایی ثابت باشد، با رد شدن از میانگین، تفاوت نمره کسب شده و نمره واقعی افزایش می‌یابد، اما وقتی ضریب پایایی تغییر کند، اندازه‌گیری انجام شده با ضریب پایایی کمتر، تفاوت بیشتری را بین برآورد نمره واقعی و نمره به دست آمده نشان خواهد داد. از سوی دیگر اگر ضریب پایایی معادل یک نباشد (یعنی آنچه در غالب موارد روی می‌دهد)، وقتی نمره کسب شده زیر میانگین است، برآورد نمره واقعی این نمره بیش از نمره به دست آمده خواهد بود و عکس این حالت در مواردی که این نمره بیش از میانگین باشد، برآورد نمره واقعی کمتر از نمره کسب شده می‌باشد. ذکر مجدد این نکته ضروری است که معادله فوق، فقط نحوه محاسبه برآوردی از نمره واقعی را ارائه می‌کند نه نمره واقعی را.

روایی

روایی از مهمترین شاخصهای یک امتحان مطلوب بوده، این گونه تعریف می‌شود: میزان توانایی آزمون در اندازه‌گیری آن چیزی که باید اندازه بگیرد. در توضیح این تعریف باید گفت هر ابزار یا آزمون باید به نحوی طراحی شود که تا حد امکان بتواند به صورت جامع و مانع تمام جنبه‌های موضوع مورد نظر را، به نحو مطلوب مورد اندازه‌گیری قرار دهد. به عنوان مثال: هر نظام آموزشی در ابتدا اهدافی در نظر دارد که برنامه‌های خود را برای رسیدن به آنها تنظیم می‌کند و در انتها با برگزاری آزمونی مشخص می‌نماید تا چه اندازه در رسیدن به اهداف مورد نظر موفق بوده است.

برای این منظور، آزمون باید کاملاً براساس اهداف آموزشی و متمرکز بر آنها طرح شده باشد تا بتواند میزان دستیابی نظام را به اهداف مذکور بسنجد. چنین آزمونی که منعکس کننده اهداف آموزشی است، یک آزمون روا محسوب می‌گردد و نتایج حاصل از آن معنی‌دار و قابل استناد بوده، می‌توان براساس آنها برنامه‌های نظام آموزشی را تکمیل و

اصلاح نمود؛ ولی در صورتی که امتحان روا نباشد، نتایج حاصل از آن به هیچوجه دقیق و قابل استفاده نخواهد بود.

فرض کنید یکی از اهداف شما این باشد که «دانشجو اعصاب مختلف بدن را شناخته بتواند آنها را در جسد مشخص کند» و شما در نظر دارید سؤالی درباره عصب حنجره‌ای راجعه^۱ مطرح کنید. ممکن است یکی از چندین سؤال زیر را پیشنهاد کنید:

سؤال (۱): عصب حنجره‌ای راجعه چه عملی بر عهده داشته، به کدام عضو عصب می‌دهد؟

سؤال (۲): منشأ عصب حنجره‌ای راجعه چیست و از کدام دسته اعصاب می‌باشد؟

سؤال (۳): در جسد باز شده رویرو، عصب حنجره‌ای راجعه را پیدا و با پنس مشخص کنید.

با توجه به هدفی که در نظر دارید فکر می‌کنید کدام سؤال پیشنهادی مناسب و روا می‌باشد؟

مشخص است که سؤالات ۱ و ۲ نمی‌تواند روا باشد چون عمل یا منشأ یک عصب جزو

هدفهای شما نبوده، فقط سؤال ۳ را می‌توانید به عنوان پرسشی مطلوب در آزمون به کار برید.

انواع روایی

روایی جنبه‌های مختلف دارد و هر آزمون، براساس اهداف آزمون‌گیرنده، باید در بردارنده یکی از انواع آن باشد. در این مبحث جنبه‌های مختلف روایی با تأکید بر کاربرد مبحث نظری روایی در آزمونهای پیشرفت تحصیلی، معرفی می‌گردد.

۱- روایی محتوا^۲

زمانی آزمون روایی محتوا دارد که نمونه دقیقی از محتوای برنامه درسی، زمان تدریس^۱ و اهداف آموزشی باشد. پس آنچه در این نوع روایی اهمیت دارد و بر آن تأکید می‌شود، نمونه‌گیری جامع و کامل از اهداف و برنامه‌های آموزشی است. البته در هر نظام آموزشی، اهداف بیشماری مد نظر است که ارزیابی همه آنها در یک آزمون به صورت تک‌تک امکان‌پذیر نیست. بنابراین باید اهدافی را به عنوان الگو انتخاب کرد که در برگیرنده کل اهداف بوده، به طوری که با سنجش آنها، همه توانایی‌های موردنظر آزموده شده باشد. برای اطمینان از روایی محتوای یک آزمون بهتر است از راهبردهای زیر استفاده نمایید:

(الف) اهداف نمونه و الگو که کلیه اهداف آموزشی را در برمی‌گیرد به دقت تعیین کنید.

ب) مواد درسی، محتوای آنها و سطوح یادگیری موردنظر و زمان تدریس برای هر ماده درسی را مشخص کرده، یادداشت نمایید.

ج) با توجه به موارد بالا جدول مشخصات تهیه کرده، برای طرح سؤالات آزمون از آن استفاده کنید (که روش ساخت آن در فصل یک ذکر شده است).

در آزمونهای فاقد روایی محتوا، بسیاری از مهارتهای موردنظر در دانشجو ارزیابی نمی‌شود و سؤالات غیر مربوط در امتحان وجود خواهد داشت که دانشجویان قادر به پاسخگویی به آنها نیستند و در کل، نتایج چنین امتحانی رضایت‌بخش و قابل استناد نخواهد بود. روایی، بیشتر بر مفاهیم ذهنی استوار است؛ اما برای اطمینان از روایی محتوای آزمون می‌توانید ابتدا اهداف تک‌تک سؤالات را با هدفهای اصلی بسنجید، آنگاه محتوای درسی و زمان صرف شده برای هر ماده درسی و تعداد سؤالات مربوط به آن را بررسی کنید. در صورت وجود تطابق کامل، آزمون شما روایی مناسبی دارد. البته در این راه از دیدگاههای کارشناسان نیز می‌توانید استفاده کنید. این کارشناسان می‌توانند همکاران شما در گروه آموزشی باشند که از اهداف آموزشی، محتوای درسی و منبع درسی مربوط اطلاع کافی دارند.

۲- روایی ظاهری (صوری)^۱

روایی ظاهری بیان‌کننده آن است که آیا آزمون «از نظر ظاهر آن، واقعاً رواست. تناسب ظاهری، منطقی و جذاب بودن آزمون، از ارکان روایی ظاهری است و بدان معناست که تا چه حد آزمون‌شونده، از سؤالات، همان چیزی را که مقصود آزمون‌گیرنده بوده، برداشت کرده است. بنابراین باید در طرح سؤالات آزمون، تا حد امکان، کلمات قاطع و عینی را به کار برد و از کلماتی که در شنوندگان یا خوانندگان برداشتهای متفاوتی ایجاد می‌کند، استفاده ننمود.

باید توجه داشت روایی ظاهری اساساً با سایر جنبه‌های روایی به ویژه روایی محتوا متفاوت است و جنبه‌هایی از آزمون را در بر می‌گیرد که بر عملکرد آزمون‌شونده مؤثر بوده، موجبات انگیزش، رغبت و تمایل او را فراهم آورده، ظاهر آزمون را از دیدگاه او مناسب می‌سازد.

بهترین راه برای اطمینان از روایی ظاهری، تأیید سؤالات آزمون، از جنبه یکسان بودن مفاهیم کلمات مورد استفاده، توسط چند نفر از جامعه آزمون‌شوندگان است که احتمالاً

می‌تواند با گفتگوی رو در رو نیز همراه باشد، اما با توجه به مشکلات اجرایی این امر، می‌توان همین کار را با استفاده از گروه‌های مشابه جامعه مزبور و یا گروه کارشناسان نیز انجام داد.

۳- روایی سازه^۱

روایی سازه یعنی میزان توانایی آزمون در اندازه‌گیری سازه یا صفت بخصوصی از فرد مثل: هوش، استعداد و ... که بیشتر در سنجش رفتارها و آزمونهای روانشناسی کاربرد دارد.

در امر آموزش، روایی سازه به این منظور تعیین می‌شود که عوامل مؤثر بر عملکرد افراد در آزمون و میزان این تأثیرات مشخص شود؛ زیرا اگر بخواهید نمرات یک آزمون را به طور دقیق تفسیر کنید باید بدانید که بجز تفاوت‌های موجود بین افراد از نظر یادگیری و مهارت، چه صفات و خصوصیات دیگری در نمرات آزمون تأثیر گذاشته است. باید صفات موردنظر و میزان تأثیر هر یک را در نتایج آزمون به صورت فرضیه‌های مشخصی بیان کنید مثلاً: «بی‌دقتی در پاسخگویی به سؤالات می‌تواند در نتایج آزمون تأثیر منفی زیادی داشته باشد». سپس آزمون را برگزار کرده، با توجه به نتایج آن میزان درستی فرضیه‌های خود را بررسی کنید. بدین وسیله می‌توانید عوامل فرعی را که در نمرات آزمون تأثیر می‌گذارد، حذف کرده، تفسیر درستی از نتایج آزمون به دست آورید.

معمولاً برای تأیید روایی سازه، از شواهد آماری توأم با استدلال منطقی استفاده می‌شود. شواهدی که مؤید آن باشند که واقعاً سازه مورد نظر اندازه‌گیری گردیده است (شواهد همگرا) و سازه‌های مشابه به جای سازه مورد نظر اندازه‌گیری نشده است (شواهد واگرا). به عنوان مثال در صورت اندازه‌گیری میزان اضطراب دانشجویان در آزمون، باید در پی جمع‌آوری شواهد و قراننی باشیم که واقعاً اضطراب مورد اندازه‌گیری قرار گرفته است (مثلاً همگرایی با علائم بالینی و یا کاهش نمرات در افراد مضطرب) و چیز دیگری به جای اضطراب (مثلاً ترس) اندازه‌گیری نشده است.

۴- روایی معیار یا ملاکی^۲

اگر بین نمرات دو آزمون همبستگی وجود داشته باشد، این آزمون دارای روایی معیار خواهد بود. پس این روایی بر مبنای همبستگی بین نمرات دو آزمون بوده، ضریب همبستگی بدست آمده، ضریب روایی معیار در نظر گرفته می‌شود. این روایی خود به دو

نوع تقسیم می‌شود:

الف) روایی پیش‌بین^۱ و ب) روایی همزمان^۲.

الف) روایی پیش‌بین. در صورتی آزمون، روایی پیش‌بین دارد که بتواند عملکرد خاصی را در آینده، در زمینه امور تحصیلی یا شغلی و ... پیش‌بینی کند.

به عنوان مثال: اگر گفته شود «شخصی که در امتحان ورودی دانشگاه موفق بوده حتماً در تحصیلات دانشگاهی نیز موفق خواهد بود»، بدان معنا است که امتحان ورودی دانشگاه می‌تواند موفقیت فرد را در امتحانات دانشگاهی پیش‌بینی کند. حال، برای پی‌بردن به درستی این گفته، می‌توان بین نمرات آزمون ورودی و نمرات دانشگاهی، ضریب همبستگی تعیین کرد. بالا بودن این ضریب، نشان‌دهنده صحت نسبی این گفته است و در غیر این صورت آزمون ورودی موردنظر، روایی پیش‌بین مطلوبی ندارد. بنابراین برای تعیین روایی پیش‌بین یک آزمون، باید دو آزمون برگزار کنید. یکی آزمون پیش‌بین که تعیین روایی آن مدنظر شماست و دیگری آزمون ملاک که ارزیابی براساس آن صورت می‌گیرد. هر چه ضریب همبستگی بدست آمده بین نتایج این دو آزمون بالاتر باشد آزمون رواتری خواهید داشت. از این نوع روایی می‌توان برای سنجش استعدادهای شغلی و تحصیلی استفاده کرد.

ب) روایی همزمان. این روایی توانایی آزمون را در تخمین عملکرد خاصی، در زمان حال نشان می‌دهد. بنابراین تنها تفاوت آن با روایی پیش‌بین در زمان می‌باشد، یعنی در روایی پیش‌بین فاصله زمانی بین آزمونهای پیش‌بین و ملاک، نسبتاً زیاد است ولی در روایی همزمان همانطور که از نام آن برمی‌آید دو امتحان یا همزمان یا بافاصله زمانی اندکی برگزار می‌شود.

این روایی موارد کاربرد زیادی دارد. مثلاً: فرض کنید دو امتحان با اهداف و محتوای یکسان طراحی کرده‌اید که یکی تعداد سؤال بیشتری داشته طولانی‌تر و دیگری سؤالات کمتری دارد. می‌خواهید بدانید این دو آزمون کوتاه و بلند از نظر نتایج مشابهند یا خیر. برای نیل به این منظور، آنها را همزمان یا با فاصله چند روز برای گروه مشخصی از افراد اجرا کرده، بین نمرات حاصل، ضریب همبستگی تعیین نمایید. اگر عدد بدست آمده، بزرگ باشد به این معنی است که این دو آزمون دارای روایی همزمان بوده، از یکی می‌توان به جای دیگری استفاده کرد.

همچنین می‌توانید نتایج یک امتحان را که تازه طراحی شده با نتایج امتحان دیگری که از روایی بالایی برخوردار است مقایسه کرده، به میزان کارآیی امتحان جدید پی ببرید یا ممکن است بخواهید آزمون ساده و مختصر را جایگزین امتحانی کنید که از نظر اجرا وقت‌گیر، پیچیده و پرهزینه است. در این صورت با تعیین روایی همزمان متوجه می‌شوید که این دو آزمون معادل است یا خیر. البته وقتی آزمون، روایی همزمان دارد نتایج آن تنها برای زمان حال مستند بوده، نمی‌توان از آن برای پیش‌بینی عملکرد آینده استفاده کرد.

روش محاسبه روایی

تعیین روایی محتوا به صورت اعداد ثابت و استفاده از روشهای محاسباتی چندان متداول و کارآمد نیست؛ در حالی که روایی ملاکی را می‌توان به راحتی با روشهای ریاضی محاسبه کرد.

بر طبق تعریف، روایی پیش‌بین یک سؤال، بیان‌کننده قدرت پیش‌بینی آن در مورد یک ملاک خاص می‌باشد. این روایی از راههای مختلف محاسبه می‌شود. یکی از روشهای متداول، روش همبستگی دو رشته‌ای نقطه‌ای است که رابطه آن به صورت زیر می‌باشد:

$$r_{pb} = \frac{Y_1 - Y}{S} \sqrt{\frac{n_1 n}{(n - n_1)(n - 1)}}$$

= n تعداد کل امتحان‌شوندگان؛

= n₁ تعداد کسانی که به سؤال مورد نظر پاسخ درست داده‌اند؛

= Y میانگین نمره خام همه امتحان‌شوندگان در آزمون ملاک؛

= Y₁ میانگین نمره خام امتحان‌شوندگانی که (در آزمون ملاک) به سؤال مورد نظر پاسخ درست داده‌اند؛

= S انحراف معیار نمره‌های خام همه امتحان‌شوندگان.

مثلاً: برای پیش‌بینی چگونگی موفقیت دانشجویان در دروس نیمسال اول پزشکی آزمونی طراحی کرده‌اید؛ اکنون می‌خواهید روایی را در مورد هر یک از سؤالات آزمون تعیین نموده، پرسشهای نامناسب را حذف کنید. برای بدست آوردن فاکتورهای لازم باید از رابطه قبل استفاده کنید. فرض کنید شما این آزمون را برای ۳۰۰ نفر برگزار کرده‌اید (n = ۳۰۰) که میانگین نمرات نیمسال اول آنها ۱۲ بوده است (Y = ۱۲). ۱۷۰ نفر از این تعداد به سؤال مورد نظر پاسخ صحیح داده‌اند (n₁ = ۱۷۰) که میانگین نمرات نیمسال اول آنها ۱۶

۲۱۰ آزمونهای پیشرفت تحصیلی در علوم پزشکی

می باشد ($Y_1 = 16$) و انحراف معیار نیز ۵ بوده است ($S = 5$). با توجه به اعداد فوق روایی سؤال مورد نظر ۰/۹۲ بدست می آید.

$$r_{pb} = \frac{16 - 12}{5} \sqrt{\frac{170 \times 300}{(300 - 170)(300 - 1)}} = 0.92$$

برای تعیین محدوده مناسب روایی، باید خطای معیار ضریب همبستگی دو رشته‌ای را با توجه به رابطه مذکور حساب کنید:

$$S_{rpb} = \text{خطای معیار ضریب همبستگی دو رشته‌ای؛}$$

$$n = \text{تعداد کل امتحان شوندگان}$$

$$S_{rpb} = \frac{1}{\sqrt{n}}$$

اگر $r_{pb} > S_{rpb} \times 1/96$ باشد آزمون از روایی خوبی برخوردار است.

S_{rpb} در امتحانی که شما اجرا کرده‌اید $\frac{1}{17} \approx 0.06$ بدست می آید و با بقیه محاسبات می بینید که $S_{rpb} \times 1/96 = 0.12$ و $0.92 > 0.12$ است. پس سؤال مورد نظر می تواند موفقیت دانشجویان نیمسال اول را پیش بینی کرده، از این جهت سؤال مناسبی می باشد.

روش دیگر، محاسبه روایی با استفاده از رابطه زیر است:

$$\text{روایی} = \frac{s^2_{CO}}{s^2_O} \quad \text{یا} \quad \text{روایی} = 1 - \frac{(s^2_O - s^2_e)}{s^2_O}$$

s^2_e = واریانس نمرات معیار؛

s^2_{CO} = کوواریانس (واریانس مشترک دو اندازه گیری)؛

s^2_O = واریانس نمرات مشاهده شده آزمونها.

ضریب روایی بین ۱ و -۱ تغییر می کند. اگر این ضریب روایی صفر باشد کوواریانس وجود ندارد یعنی آزمون به هیچوجه پیش بینی کننده نیست. اگر یک باشد همه واریانس مشاهده شده کوواریانس خواهد بود. یعنی آزمون می تواند صد درصد موارد، نتایج را پیش بینی کند. پس هرچه این ضریب به ۱ نزدیکتر باشد آزمون پیش بین مناسبتر خواهد بود و البته در عمل، بدست آوردن ضریب ۱ تقریباً غیر ممکن می باشد.

در صورتی که هدف از تعیین روایی، جایگزین کردن آزمون طولانی با آزمون کوتاه باشد ضریب روایی مطلوب حدود ۰/۷ یا بیشتر خواهد بود. اما در مورد روایی معیار (همزمان و پیش بین) ضریبهای پایین تر یعنی حدود ۰/۳ تا ۰/۴ هم می تواند مفید باشد.

عوامل مؤثر بر روایی آزمون

در هر امتحان سؤالات باید متناسب با نیازهای شغلی آینده دانشجو تعیین شود، یعنی مطالبی مورد سؤال قرار گیرد که دانستن آنها برای دانشجو مهم و ضروری باشد. در یک دانشکده پزشکی، پس از برگزاری چندین امتحان فیزیولوژی و آناتومی از چند پزشک با تجربه موفق درخواست شد که به سؤالات این آزمونها پاسخ گویند. یکی از پزشکان پس از بررسی سؤالات گفت:

«من و سایر پزشکان، نتوانستیم به سؤالات پاسخ دهیم. من نمی‌فهمم چه احتیاجی است که دانشجویان این مطالب را بدانند».

این بررسی ساده به خوبی نشان داد که در این امتحانات مطالبی مورد سؤال قرار گرفته که از نظر عملکرد شغلی در آینده اهمیتی ندارد و با توجه به اینکه دانشجو می‌خواهد در امتحان موفق باشد، صرفاً مطالبی را یاد می‌گیرد که ممکن است در امتحان طرح شود. بنابراین با در نظر گرفتن تأثیری که امتحان بر نحوه یادگیری دانشجویان دارد، آزمون‌گیرنده باید به روایی محتوای آزمون توجه فراوان داشته باشد و سعی کند تمام مطالب با اهمیت و مهارت‌های لازم را در امتحان بگنجانند. در این صورت یادگیری دانشجویان نیز بهتر خواهد شد.

گاهی در امتحانات فقط دانش و معلومات دانشجو ارزیابی شده، تواناییهای کاری وی نادیده گرفته می‌شود. این مسأله نیز به روایی محتوای آزمون لطمه می‌زند.

مثلاً: فرض کنید یکی از اهداف شما در زمینه آموزش بهورزان این است که «بهورز بتواند مادران را متقاعد کند که نوزادان خود را شیر دهند». در این صورت اگر از آزمون گشوده بپرسید «ارزش غذایی شیر مادر را به طور کامل توضیح دهید» سؤالی روا مطرح نکرده‌اید، زیرا این پرسش تنها قسمتی از معلومات نظری بهورز را مشخص کرده، مهارت‌های لازم برای رسیدن به هدف موردنظر مثل صحبت کردن، استدلال کردن و ... را نمی‌سنجد. بهتر است تواناییهای کاری دانشجو را به صورت عملی بسنجید، گرچه این کار دشوار و وقت‌گیر است ولی در عوض، امتحانی روا برگزار کرده‌اید که نتایج آن نیز قابل استناد می‌باشد.

علاوه بر موارد ذکر شده، عوامل متعدد دیگری نیز بر روایی آزمون تأثیر می‌گذارند. از سویی هر عاملی که بر عملکرد دانشجو در هنگام امتحان مؤثر باشد می‌تواند روایی آن را تغییر دهد. در صورتی که دستورالعمل پاسخگویی به سؤالات، مبهم و نامشخص بوده یا

سؤالات دو پهلو طرح شده باشد یا صورت سؤال مشکل و غیرقابل فهم باشد ممکن است دانشجو علی‌رغم داشتن اطلاعات لازم برای پاسخگویی به سؤال نتواند پاسخ صحیح را بنویسد و این موضوع روایی آزمون را دچار خدشه می‌کند.

نحوه قرار گرفتن سؤالات در امتحان نیز از نظر روایی اهمیت دارد. اگر دو امتحان «الف» و «ب» از لحاظ محتوا و سؤالات، کاملاً یکسان ولی در امتحان «الف» سؤالات مشکل در آخر برگه سؤالات و در امتحان «ب» در ابتدا باشد روایی آزمون «ب» نسبت به «الف» کاهش خواهد یافت. زیرا در امتحان «ب» دانشجو فرصت زیادی برای پاسخگویی به سؤالات مشکل که در ابتدای برگه آزمون قرار گرفته صرف کرده، دچار اضطراب و تشویش شده، نمی‌تواند به خوبی از عهده امتحان برآید. بنابراین باید در هر امتحانی ترتیب سؤالات از آسان به سخت باشد تا روایی افزایش یابد.

از عوامل مؤثر دیگر بر روایی آزمون، ارزش نسبی سؤالات است. گفته شد که با توجه به زمان تدریس، برای هر بخش و میزان اهمیت آن باید تعداد سؤالات مربوط به آن را طرح کرد. البته آنچه در این خصوص اهمیت دارد سهم نمره مربوط به هر بخش در پراکنندگی مجموع نمرات می‌باشد نه نسبت تعداد پرسشهای مربوط به آن. برای این منظور می‌توان از جدول مشخصات استفاده کرد. برای تعیین ارزش نسبی سؤالات به موارد زیر توجه داشته باشید:

ابتدا اهمیت مبحث مورد نظر را تعیین کرده، ارتباط منطقی آن مبحث را به کل موضوعات درس بسنجید و سپس نسبت و تعداد ساعات تدریس را برای این مبحث مشخص کنید. با توجه به این موارد و تجربه‌ای که به عنوان طراح سؤال دارید می‌توانید ارزش نسبی سؤالات را تعیین کنید.

در قسمتهای مختلف این فصل عوامل کاهش دهنده روایی آزمون مورد بحث قرار گرفت که در اینجا فهرست وار ذکر می‌گردد:

- ۱- وجود سؤالات و دستورالعملهای مبهم، نثر مشکل و عدم رعایت قواعد دستوری در سؤال؛
- ۲- گزینه‌های انحرافی نامناسب در سؤالات چند گزینه‌ای؛
- ۳- قرار دادن سؤالات مشکل در ابتدای آزمون؛
- ۴- اجرای ضعیف آزمون و اشتباهات نمره‌گذاری و تفسیر نتایج؛
- ۵- امکان حدس زدن پاسخ در سؤالات گزیده پاسخ؛
- ۶- اضطراب و نگرانی دانشجو.

ارتباط روایی و پایایی

باید گفت روایی تا حدودی به پایایی وابسته است ولی عکس آن ممکن است درست نباشد. به عبارت دیگر آزمون روا حتماً پایا است در صورتی که آزمون پایا ممکن است روا نباشد. آزمونی که به طور پایدار عامل اشتباهی را می‌سنجد پایاست ولی روا نیست. اگر آزمونی پایایی نداشته باشد قادر نخواهد بود چیزی را که باید، اندازه بگیرد، بنابراین فاقد روایی خواهد بود.

منابع:

- ۱- آذر، رسول. روشهای ارزیابی/آموزشی تهران: انتشارات نیما، ۱۳۶۹.
- ۲- بلوم؛ و همکاران. راهنمای ارزشیابی تکوینی و مجموعی، جلد اول، ترجمه ابراهیم کظیمی. تهران: انتشارات دانشگاه تربیت معلم، ۱۳۵۵.
- ۳- پاشا شریفی، حسن. اصول روان‌سنجی و روان‌آزمایی تهران: انتشارات رشد، ۱۳۷۳.
- ۴- سیف، علی‌اکبر. روانشناسی یادگیری و تدریس. تهران: انتشارات دانشگاه سپاهیان انقلاب ایران، ۱۳۵۵.
- ۵- گنجی، حمزه. اندازه‌گیری و ارزشیابی در آموزش و پرورش. تهران: انتشارات بعثت، ۱۳۷۲.
- ۶- مایزما، ویلیام؛ و همکاران. اندازه‌گیری و آزمون در تعلیم و تربیت. ترجمه غلامرضا خوئی‌نژاد. مشهد: انتشارات مؤسسه چاپ و انتشارات آستان قدس رضوی، ۱۳۷۲.
- ۷- هومن، حیدرعلی. اندازه‌گیری‌های روانی و تربیتی و فن تهیه تست. تهران: انتشارات فردوسی، ۱۳.
- 8- Abbatt FR. *Teaching For better learning*. 2nd ver.ed. Geneva: World Health Organization, 1992.
- 9- Bordage G. *Content validation of key features on a national examination of clinical decision making skills*. Acad Med. 1995; Apr 70 (4): 276-82.
- 10- Collidge F, Merwin M. *Reliability and validity of the college axis II inventory: a new inventory for the assessment*. J Pers Assess 1992; 10(2): 223-38.
- 11- Cox K. *The medical teacher, writing multiple choice question*. London: Churchill-Livingstone, 1982.
- 12- Hays RB, Iabb W. *Reliability of the fellowship examination of the royal Australian college of general practitioners*. Teach & Learn In Med 1995; 7: 43-5.
- 13- Hays RB, Vander V. *Longitudinal reliability of the royal Australian college of general practitioners certification examination*. Med Educ 1995; 29: 317-21.
- 14- Henerson ME, Morris LL, Fitz- Gibbon CT. *How to measure attitudes*. New York: Sage, 1987.
- 15- Neufeld V, Norman GR. *Assessing clinical competence*. New- York: Springer Publishing Company, 1985.
- 16- Salvia J; Yssldyke JE. *Assessment*. 7th rev. ed. Boston: Houghton Mifflin Company, 1998.
- 17- Taiwo A. *Fundamentals of classroom testing*. New Delhi: Vikas Publishing House, 1995.

فصل دهم

تجزیه و تحلیل سوالات آزمون

- خطای اندازه‌گیری
- تجزیه و تحلیل سؤال
- رواسازی محتوای سوالات
- رویکردهای قضاوتی برای تجزیه و تحلیل سؤال
- تجزیه و تحلیل آماری سؤال
- عوامل سؤال

تجزیه و تحلیل سؤالات آزمون

کار تهیه آزمون با نوشتن سؤالات آن پایان نمی‌یابد، بلکه پس از آن باید به دنبال جمع‌آوری شواهدی برای اطمینان از کیفیت مناسب سؤالات بود. برای چنین هدفی، طراح آزمون باید یک فرایند دقیق نظام‌دار را برای جمع‌آوری اطلاعات صحیح به انجام رسانده، بر پایه آن اطلاعات، قضاوت نماید. هدف چنین قضاوت‌هایی صرفاً شناسایی منابع احتمالی خطا و کاهش خطا در اندازه‌گیری است. به ویژه در مواردی که نتایج آزمون به عنوان مبنای انتخاب یا دسته‌بندی افراد و پذیرش آنها به کار می‌رود، باید نتایج آزمون از دقت کافی برخوردار باشد. معمولاً به علت فقدان وقوف کامل به اهمیت ارزشیابی، فقدان آگاهی لازم از روشهای تجزیه و تحلیل و احساس آن که تجزیه و تحلیل سؤالات به صرف وقت زیادی نیاز دارد، استادان تمایلی برای این کار نشان نمی‌دهند. در این فصل ضمن مرور مجدد مفهوم خطا در اندازه‌گیری راهکارهای مناسب جهت شناسایی منابع خطا، روشهای اندازه‌گیری خطا و راههای کاهش خطا ارائه خواهد گردید. با کاستن از عوامل خطای اندازه‌گیری می‌توان به افزایش کیفیت سؤالات آزمون امیدوار بود.

برای کشف خطای اندازه‌گیری دو رویکرد وجود دارد: ۱- قضاوت کارشناسان خبره‌ای که از ملاکهای اندازه‌گیری آگاهی دارند؛ ۲- استفاده از روشهای آماری مناسب. هر رویکرد شامل روشهای گوناگونی است که در این فصل عملی‌ترین روشهایی که به واسطه آن بتوان به قضاوتی مناسب در زمینه کیفیت سؤالات رسید، ارائه می‌گردد. یادآور می‌شود مبنای تجزیه و تحلیل آزمون می‌تواند هم براساس نظریه‌های سنتی (کلاسیک) و هم جدید صورت پذیرد که برای آگاهی بیشتر می‌توانید به منابعی که براساس هر یک از نظریه‌ها به رشته تحریر درآمده است، مراجعه نمایید. همچنین در این مبحث از ذکر رابطه‌های پیچیده آماری، به ویژه آنهایی که به راحتی با نرم‌افزارهای موجود قابل محاسبه‌اند خودداری شده است.

موضوعات این فصل شامل این موارد است:

- خطای اندازه‌گیری^۱؛
- تجزیه و تحلیل سؤال؛
- رواسازی محتوای سؤالات؛
- رویکردهای قضاوتی برای تجزیه و تحلیل سؤال؛
- تجزیه و تحلیل آماری سؤال؛
- عوامل سؤال^۲.

خطای اندازه‌گیری

هر سؤال در اندازه‌گیری آنچه در پی آن است توأم با مقداری خطاست و بهترین سؤال، سؤالی است که این میزان را به حداقل رساند. این خطاها بر دو نوع است:

۱- خطای تصادفی که عوامل مختلفی در آن دخیل بوده، بسیاری از آنها حتی قابل شناسایی هم نیست.

۲- تورش^۳ که به صورت نظام‌داری در اندازه‌گیریها رخ می‌نماید و سبب تخمین کمتر یا بیشتر یک متغیر نسبت به مقدار حقیقی آن می‌گردد.

در اینجا پیرامون خطای اندازه‌گیری و کاهش منابع بروز خطا بحث می‌شود. عبارت‌بندی سؤالات که باعث سردرگمی آزمون‌شونده گردد، اطلاعات غیر دقیق در سؤالات، فقدان هماهنگی یک سؤال با اهدافی که باید اندازه‌گیری شود و مانند آن، از منابع بالقوه خطای اندازه‌گیری است. چون این منابع به طور کل در سنجش‌های روانشناختی حذف شدنی نیست، بهترین شیوه به حداقل رساندن آنهاست.

تعریف خطای اندازه‌گیری

اگر در وضعیت مشخص، آزمون بی‌نهایت مرتبه تکرار گردد، مقدار انحراف نمره آزمون‌شونده نمایان‌کننده خطای اندازه‌گیری است و هر چه نمرات بیشتر پراکنده باشد، خطای اندازه‌گیری بیشتر خواهد بود. البته نمی‌توان یک آزمون را برای یک آزمون‌شونده در دفعات بی‌نهایت تکرار کرد و بنابراین خطای اندازه‌گیری براساس یک بار اجرای آزمون تخمین زده می‌شود که این تخمین نسبتاً دقیق است. در این حالت از خطای معیار

1- Measurement Error
3- Bias

2- Item Parameters

اندازه‌گیری به صورت نظری استفاده می‌شود:

خطای اندازه‌گیری - نمره مشاهده شده = نمره حقیقی

نمره واقعی، نمره‌ایست که آزمون‌شونده در صورتی که هیچ خطایی وجود نداشته باشد، کسب خواهد نمود. نمره مشاهده شده نمره‌ای است که یک آزمون‌شونده در شرایط موجود، در خلال اجرای آزمون در وضعیت واقعی کسب نموده است. اگر امتحان در دفعات بی‌نهایت تکرار گردد، نمره حقیقی برابر با میانگین نمرات آزمون‌شونده خواهد بود. با استفاده از نمودار، خطای معیار اندازه‌گیری به شکل توزیع نمرات در حوالی نمره واقعی برای هر فرد مشخص می‌گردد.

تجزیه و تحلیل سؤال

تجزیه و تحلیل سؤال، فرایند بررسی نقادانه سؤالات آزمون بوده، هدف آن شناسایی و کاهش منابع خطای اندازه‌گیری است. باید طراحان سؤال به طور مستمر تجزیه و تحلیل سؤالات را به عنوان معیار کیفیت و حذف سؤالات غیرقابل قبول، جایگزین ساختن آنها با سؤالات بهتر و نگه داشتن سؤالاتی که ملاکهای مطلوب دارد، به کار گیرند.

روشهای تجزیه و تحلیل سؤالات

تجزیه و تحلیل سؤال فوایدی را در بر دارد؛ که از آن جمله می‌توان به موارد زیر اشاره کرد:

- ۱- منجر به افزایش کیفیت سؤالات می‌شود؛
- ۲- در تهیه یک بانک سؤال کارآمد مورد استفاده قرار می‌گیرد؛
- ۳- منجر به بهبود و ارتقای مهارتهای ما در تهیه آزمون می‌گردد؛
- ۴- ارزش تشخیصی دارد به نحوی که با مشخص کردن نقاط قوت و ضعف دانشجویان در طراحی و تدوین فعالیتهای یادگیری در آینده کمک می‌کند؛
- ۵- مبنایی برای بحث در زمینه نتایج آزمون است و
- ۶- در صورتی که دانشجویان در این فرایند مشارکت داشته باشند و یا نتایج تجزیه و تحلیل سؤالات به آنها گفته شود، می‌تواند به عنوان یک تجربه یادگیری مورد استفاده قرار گیرد.

ارائه بازخورد به دانشجویان

همگان باید به این باور برسند که استاد، لازم است به عنوان بخشی از فرآیند آموزش، در مورد تک تک سؤالات آزمون با دانشجویان به بحث و گفتگو بپردازد. این امر موجب به بار آمدن نتایج خواهد شد که می‌توان به ذکر موارد زیر پرداخت:

- ۱- خطاها و اشتباهات دانشجویان به ویژه آن کسانی که نتایج ضعیفی به دست آورده‌اند یا مهارت‌های موردنظر را به درستی فرا نگرفته‌اند، اصلاح می‌گردد؛
- ۲- انگیزه دانشجویان برای کسب نتایج بهتر در آزمونهای بعدی افزایش می‌یابد؛
- ۳- به وسیله این کار می‌توان به دانشجویان شیوه سازماندهی رفتارها و عاداتهای مطالعه کردن را بر اساس اهداف آموزشی آموخت؛ و
- ۴- دانشجویان اجازه می‌یابند تا در مورد علت انتخاب پاسخ صحیح بحث کنند که این امر به رفع کج‌فهمی‌ها کمک می‌کند.

البته سؤالاتی که اینگونه مورد تجزیه و تحلیل قرار گیرند، دیگر سرّی نخواهند بود و بنابراین امکان کاربرد دوباره را از دست خواهند داد؛ اما با کمی توجه به مزایای بیشمار ارائه بازخورد به دانشجویان، برتری آن آشکار می‌گردد. علاوه بر آن ممکن است در مورد وقت زیادی که برای برگزاری چنین جلساتی باید صرف گردد، نگران شوید، اما توجه نمایید که اطمینان از یادگیری درست و کامل مطالب قبلی بسیار مهمتر از ارائه مطالب جدید است. تجزیه و تحلیل سؤالات به یکی از روشهای تجزیه و تحلیل آماری و یا با استفاده از رویکرد قضاوتی قابل انجام است. در تجزیه و تحلیل عددی بسته به توزیع پاسخهای هر سؤال از روشهای آماری مناسب استفاده می‌شود. بنابراین، روش مذکور عمدتاً بر پایه اجرای آزمون و تجزیه و تحلیل پس از اجرا مبتنی است. در رویکردهای قضاوتی از برخی افراد خواسته می‌شود در مورد ویژگیهای سؤال و میزان انطباق آن با برخی معیارها (ملاکها) اعلام نظر نمایند. ممکن است این افراد از جمله صاحب‌نظران در زمینه محتوا کارشناس یا از جمله ویراستاران و یا در زمره آزمون‌شوندگان باشند. در مرور بر مبنای داورى به دو مطلب مهم باید توجه داشت: اولاً هر مرورکننده باید در آن وظیفه^۱ صاحب صلاحیت و شایستگی بوده، و از طرفی خود وظیفه نیز یک فرایند نظامدار باشد. بدین معنا که روش انجام آن تعریف شده و ملاکهای مرور آن در دسترس باشد.

رواسازی محتوای سؤالات ۱

در رواسازی محتوای آزمون سعی در جمع‌آوری دلایل و شواهدی مبنی بر روایی محتوای آزمون است. این امر در تشخیص کیفیت آزمون کمک شایان توجهی می‌کند. طراح آزمون می‌تواند از اطلاعاتی که از مطالعات نظام‌دار به دست آمده، برای ارزشیابی و بهبود سؤالات استفاده نماید. در رواسازی باید از قدرت ارزیابی سؤالات لااقل در یکی از حیطه‌های محتوا اطمینان حاصل کرد. برای رسیدن به این اهداف از تشکیل نشست‌هایی با حضور کارشناسان برگزار می‌گردد؛ که ایشان در مورد میزان هماهنگی سؤال با محتوا قضاوت خواهند نمود. به طور کلی برای نظرخواهی از کارشناسان دو روش کلی وجود دارد:

الف) فهرستی از اهداف و سؤالات از قبل جور شده، برای نظرخواهی در اختیار داوران قرار داده می‌شود؛ به عبارت دیگر نویسنده هر سؤال، آن را با یک هدف یا حیطه ویژه مرتبط می‌سازد و نقش داور یا داوران آن است که وجود این ارتباط را تأیید یا رد نمایند.

البته پرسشنامه‌ای مناسب برای ثبت نظر داوران موردنیاز می‌باشد که می‌توان از مقیاس بلی - خیر یا مقیاس‌های درجه‌بندی (درجه‌ای از هماهنگی و ارتباط) استفاده نمود (فرم ۱-۱۰).
ب) در روش دیگر فقط اهداف و سؤالات در اختیار داوران قرار می‌گیرد. هر داور شخصاً به جور کردن سؤال با یکی از اهداف اقدام کرده، نظر خود را در یک برگه جمع‌آوری اطلاعات (مقیاس‌های درجه‌بندی) منعکس می‌کند (فرم ۲-۱۰).

این قضاوت‌ها به عنوان شواهد مربوط به روایی محتوا مورد توجه قرار می‌گیرد. برای بهتر شدن نتیجه کارشناسیها می‌توان از نشست‌های کور^۲ استفاده نمود. در این روش مسئول طرح با استفاده از تلفن، پست یا شبکه‌های رایانه‌ای اقدام به جمع‌آوری نظر گروه داوران می‌نماید. این امر کمک می‌کند تا میزان تأثیرگذاری داوران یا فشار ایشان برای متقاعد کردن یکدیگر تحت تأثیر مقام و موقعیت، مرتبه اجتماعی و یا جاذبه شخصی به حداقل ممکن برسد.

معمولاً جلب نظر چهار یا پنج داور برای هر سؤال کافی است، البته در صورت زیاد بودن تعداد سؤالات می‌توان سؤالات را به چند بخش تقسیم نمود و در مورد هر بخش نظر چهار یا پنج داور را جویا شد. در مواردی که نتیجه امتحان برای آزمون‌شوندگان حیاتی بوده امتحان دارای حدنصاب قبولی است، (مانند امتحانات جامع و پذیرش در مقاطع مختلف) بهتر است از نظر ده تا پانزده نفر داور استفاده کرد. داوران باید در زمینه مورد ارزیابی و نیز در جور کردن سؤالات با اهداف، خبره باشند.

فرم ۱-۱۰: نمونه برگه داوری برای بررسی میزان هماهنگی سؤالات و اهداف

نام.....
واهنمایی: هدف شماره یک را در زیر بخوانید. سپس اولین سؤال دفترچه آزمون را مطالعه کرده به دقت به میزان هماهنگی سؤال با مهارت مربوط توجه کنید. به منظور درجه بندی میزان هماهنگی از الگوی زیر پیروی نمایید.

با درجه هماهنگی زیاد = ۱

با درجه هماهنگی متوسط = ۲

با درجه هماهنگی کم یا نامشخص = ۳

اگر نظریه خاصی راجع به هماهنگی این سؤال دارید، در محل مخصوص بنویسید. پس از پایان قضاوت در مورد یک سؤال، به سراغ سؤال دوم رفته به همین نحو در تمام سؤالات قضاوت نمائید.

هدف شماره یک:

دانشجو بتواند کاربرد داروهای ضد فشارخون را در مورد درجات مختلف بیماری شرح دهد.

درجه بندی	نظریه
سؤال ۱
سؤال ۲
سؤال ۳
سؤال ۴
سؤال ۶
سؤال ۹

هدف شماره دو:

دانشجو بتواند برای انواع مختلف سرطان، دارو درمانی مناسب را پیشنهاد نماید.

درجه بندی	نظریه
سؤال ۳
سؤال ۵
سؤال ۱۰
سؤال ۱۱

فرم ۲-۱۰. نمونه برگه داوری در زمینه هماهنگی سؤالات با اهداف

جور کردن سؤالات/اهداف

نام مرور کننده.....

محتوای مورد نظر:

نخست، فهرست تمام حیطه‌های اختصاصی موردنظر و سؤالات آزمون را بخوانید. وظیفه شما آن است که مشخص کنید آیا هر سؤال یکی از حیطه‌های موردنظر را اندازه‌گیری می‌نماید یا خیر. این امر به شرطی است که شما احساس کنید پاسخ‌دهی آزمون‌شونده به یک سؤال بیانگر سطح خاصی از عملکرد در حیطه موردنظر است. در کنار هر هدف، شماره سؤالاتی را که فکر می‌کنید اندازه‌گیری آن هدف خاص است، یادداشت نمایید. احتمالاً در برخی از موارد، شما احساس می‌کنید که سؤال هیچ یک از حیطه‌های موردنظر را اندازه‌گیری نمی‌نماید، شماره چنین سؤالاتی را در پایان برگه بنویسید.

سؤالات جور شده	هدف
	۱
	۲
	۳
	۴

پاسخ به هیچ هدفی مربوط نیست.

کمی سازی نظر داوران

پس از جمع‌آوری مناسب داده‌های مربوط به ارزشیابی داوران اطلاعات مزبور باید تجزیه و تحلیل و سپس تفسیر گردد. طراح آزمون باید از توافق نظر داوران در زمینه انطباق سؤال با اهداف کاملاً مطمئن گردد، مثلاً، اگر پنج داور وجود دارد، چهار تن از آنها باید توافق داشته باشند یا در صورت استفاده از ده داور باید هشت تن آنها موافق باشند.

همچنین در تجزیه و تحلیل سؤال می‌توان از جلسات غیر رسمی و پرسش از افراد مناسب برای اطمینان از روایی ظاهری و محتوایی سؤالات بهره گرفت. مثلاً از سایر

همکاران شاغل در گروه آموزشی یا گروهی مشابه با افرادی که قرار است مورد آزمون قرار گیرند، می‌توان خواست در مورد سؤالات قضاوت کنند. مثلاً به سؤالات باز ذیل پاسخ گویند: به نظر شما، آیا این سؤال دانشجو را گیج می‌کند؟ آیا گزینه‌های غلط باور کردنی است؟ آیا کلمات استفاده شده با توجه به سن و مرتبه دانشجو قابل فهم می‌باشد؟ آیا نمودارها و تصاویر واضح است؟

البته این روش مشکلاتی را در بردارد که دشواری برپایی و ادارهٔ چنین جلساتی، از جمله آنهاست. همچنین پیدا کردن گروهی مشابه با افراد آزمون‌شونده به نحوی که سؤالات فاش نشود، بسیار مشکل است. گرچه چنین مشکلاتی وجود دارد، استفاده از این روش بسیار ارزشمند است.

روشهای آماری تجزیه و تحلیل سؤالات

برخی از روشهای آماری که در تجزیه و تحلیل سؤالات قابل استفاده است، در این مبحث مورد بررسی قرار خواهد گرفت. کروکر و آلیژینا^۱ (۱۹۸۶) شاخصهای مذکور را به سه دسته تقسیم نموده‌اند:

- ۱- شاخصهایی که بیان‌کنندهٔ چگونگی توزیع پاسخهای یک سؤال است (میانگین و واریانس پاسخهای هر سؤال)؛
- ۲- شاخصهایی که درجهٔ ارتباط بین پاسخ به هر سؤال و برخی ملاکهای موردنظر را توصیف می‌کند؛ و
- ۳- شاخصهایی که به صورت همزمان، بیان‌کنندهٔ واریانس سؤال و ارتباط سؤال با یک ملاک است.

توصیف نمرات آزمون

امتیازی که پس از تصحیح برگهٔ امتحانی به دانشجو داده می‌شود، نمرهٔ خام نام دارد که به خودی خود ارزشی ندارد، چون اولاً اطلاعات زیادی راجع به استاندارد بودن یا نبودن امتحان بدست نمی‌دهد، ثانیاً قادر به نشان دادن درجه قوت و ضعف دانشجویان نیست. برای به دست آوردن چنین آگاهی‌هایی احتیاج به توصیف نمرات آزمون است. برای توصیف نمرات آزمون می‌توان اندازه‌گیریهای زیر را انجام داد:

الف- محاسبه انحراف معیار آزمون

انحراف معیار، معدل فاصله نمرات از نمره میانگین است و میانگین نمرات، حاصل تقسیم مجموع نمرات بر تعداد آنها می‌باشد. برای به دست آوردن انحراف معیار، پس از مرتب کردن نمرات به ترتیب ارزش آنها $\frac{1}{6}$ بالایی نمرات و $\frac{1}{6}$ پایینی آن را مشخص کرده، با استفاده از فرمول زیر انحراف استاندارد بدست می‌آید:

$$\text{انحراف معیار} = \frac{\text{حاصل جمع } 1/6 \text{ بالایی} - \text{حاصل جمع } 1/6 \text{ پایینی}}{\text{نصف تعداد دانشجویان}}$$

مثال: اگر نمرات ۳۰ دانشجو را در آزمونی به ترتیب زیر در نظر بگیرید:

$$14 - 15 - 15 - 17 - 18 - 18 - 20 - 20 - 21 - 22 - 22 - 22 - 23 - 23 - 23 - 25 - 25 - 26 - 26 - 27 - 27 - 28 - 30 - 30 - 30 - 31 - 32 - 32 - 32 - 33 - 33 - 35 - 36 - 36 - 37$$

$$\text{حاصل جمع } \frac{1}{6} \text{ بالایی} = 177 = 37 + 36 + 36 + 35 + 33$$

$$\text{حاصل جمع } \frac{1}{6} \text{ پایینی} = 79 = 14 + 15 + 15 + 17 + 18$$

$\frac{1}{6}$ نمرات برابر ۵ عدد از نمرات است، بنابراین:

$$\text{انحراف معیار} = \frac{177 - 79}{15} = 6.53$$

اگر انحراف معیار بزرگ باشد نشان می‌دهد که آزمون توانسته است آزمون‌شوندگان را بهتر طبقه‌بندی کند ولی اگر انحراف معیار کوچک باشد بیان‌کننده عدم توانایی آزمون در نشان دادن تفاوت بین امتحان‌شوندگان است.

آزمونی مناسب است که منحنی نمرات آن به شکل زنگوله‌ای متقارن باشد. در چنین منحنی، میانگین نمرات در مرکز آن در نظر گرفته شده، واحدهای محور افقی براساس انحراف معیار محاسبه می‌شود. به این معنی که +۱ برابر انحراف معیار و +۲ دو برابر انحراف معیار و ...

در این منحنی اکثریت نمرات، نزدیک نمره میانگین متمرکز بوده، و هر چه از میانگین، دورتر شود تراکم نمرات آن منطقه کاهش می‌یابد. در یک منحنی توزیع طبیعی از کل نمرات ۳۴ درصد آنها از ۰ تا +۱ انحراف معیار، ۱۴ درصد از +۱ تا +۲ انحراف معیار و ۲ درصد از +۲ تا +۳ انحراف معیار و ۰/۱۳ درصد بیشتر از +۳ قرار می‌گیرد. همین ترتیب در مورد اعداد سمت چپ منحنی صدق می‌کند. پس می‌توان گفت امتحانی دارای توزیع طبیعی است که تقریباً $\frac{2}{3}$ (۶۸ درصد) نمرات آن در محدوده میانگین \pm یک انحراف استاندارد قرار

گیرد.

مثال: در امتحان مثال قبل نمره ۲۵/۶ که میانگین نمرات است را به عنوان نقطه صفر منحنی در نظر می‌گیریم و از دو طرف به اندازه ۶/۵۳ (یک انحراف معیار) و ۱۳/۰۶ (دو انحراف معیار) و ۱۹/۵۹ (سه انحراف معیار) پیش می‌رویم.

روش دیگر برای محاسبه انحراف معیار

با توجه به فرمولهای زیر نیز می‌توان انحراف استاندارد را محاسبه کرد:

$$\text{انحراف معیار} = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

$$\text{انحراف معیار} = \sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}$$

x = نمره هر دانشجو؛ x = میانگین نمرات دانشجویان

$\sum x^2$ = مجموع مجذور نمرات دانشجویان؛ $(\sum x)^2$ = مجذور جمع نمرات دانشجویان

n = تعداد کل دانشجویان.

ب- محاسبه حدنصاب قبولی (MPL)^۱

حد نصاب قبولی، برابر حداقل نمره‌ای است که دانشجو باید از یک درس بدست آورد تا شایستگی قابل قبول او در آن درس اثبات شود.

تعیین حد نصاب قبولی براساس قضاوت دو یا چند استاد صورت گرفته، مقدار آن متغیر است. برای تعیین MPL بعد از امتحان باید چند استاد در مورد سؤال نظر بدهند. با توجه به نظر اساتید برای کسب MPL دانشجوی ضعیف باید قادر باشد تعدادی از گزینه‌های انحرافی را که دور از ذهن هستند کنار گذاشته، بین بقیه جوابها به انتخاب بپردازد. در این صورت می‌توان گفت که دانشجو موفق به کسب حد نصاب قبولی شده است. برای تعیین حدنصاب قبولی می‌توانید، از دستورالعمل زیر استفاده نمایید:

- ۱- به گزینه‌های انحرافی که دانشجوی لب مرز بتواند به راحتی آنها را کنار بگذارد «۰» بدهید.
- ۲- به گزینه‌های انحرافی که دانشجوی متوسط نتواند غلط بودن آن را تشخیص دهد بسته به میزان نزدیک بودن آنها به جواب صحیح نمراتی نظیر ۰/۲۵ و ۰/۵ و ۰/۷۵ بدهید.
- ۳- برای پاسخ صحیح سؤال، نمره «۱» را در نظر بگیرید.

۴- اعداد بدست آمده را با هم جمع کنید.

۵- با توجه به فرمول زیر MPL را محاسبه نمایید:

$$MPL = \frac{\text{نمره منظور شده برای پاسخ صحیح}}{\text{جمع اعداد بدست آمده}}$$

مثال: تعداد کل اسید آمینه‌های شناخته شده چقدر است؟

- الف) ۲۰ عدد ۰/۵
- ب) ۵۰ عدد ۰/۲۵
- ج) ۳۰ عدد ۱
- د) ۱۰۰ عدد ۰
- ه) ۲۰۰ عدد ۰

$$MPL = \frac{1}{175} = 0.57 \text{ برای این سؤال خاص}$$

۶- حال برای محاسبه حدنصاب قبولی برای کل امتحان طبق فرمول زیر عمل کنید.

$$MPL = \frac{\text{مجموع حدنصاب تک تک سؤالات آزمون}}{\text{تعداد سؤالات}} \times 100$$

نمرات بالای حد نصاب قبولی شایستگی لازم را در درس موردنظر کسب کرده،

نمرات پایین‌تر از حدنصاب در امتحان موفق نشده‌اند.

ج- تعیین رتبه دانشجویان

یک راه آسان ولی غیردقیق برای مقایسه نمرات دانشجویان، رتبه‌بندی آنها بین خودشان است ولی ارزش رتبه داده شده تا حدود زیادی به تعداد اعضای امتحان‌شونده بستگی دارد.

مثلاً: کسب رتبه ۱۰ بین دویست نفر نشانه موفقیت است ولی همین رتبه بین ده نفر رتبه مناسبی نیست. بنابراین در تعیین رتبه باید تعداد افراد را در نظر گرفت. با توجه به این موضوع روش تعیین رتبه درصدی، روش مناسبی است. منظور از رتبه درصدی، درصد نمراتی است که در زیر یک نمره معین قرار دارد. به عنوان مثال: اگر رتبه درصدی فردی ۹۵ باشد، یعنی که نمرات ۹۵ درصد افراد پایین‌تر از نمره این فرد قرار گرفته است. اگر رتبه درصدی فردی پنجاه باشد یعنی ۵۰ درصد نمرات، پایین‌تر از نمره وی قرار گرفته است. رتبه درصدی پنجاه را میانه نیز می‌نامند.

برای تعیین رتبه درصدی به روش زیر عمل کنید:

- ۱- نمرات دانشجویان را حدوداً در ۱۰ طبقه جای دهید.
- ۲- فراوانی نمرات سه طبقه را محاسبه کنید (f).
- ۳- فراوانی تراکمی نمرات را تعیین کنید. برای این کار فراوانی هر نمره را با فراوانیهای نمرات پایین جمع کنید (fc).
- ۴- به نصف فراوانی هر نمره، فراوانی تراکمی نمره بعدی را اضافه کرده، حاصل را در $\frac{100}{N}$ (N = تعداد امتحان‌شوندگان) ضرب کنید تا رتبه درصدی بدست آید.

بعنوان مثال: نمرات تعدادی دانشجو به شرح زیر است:

۱۱- ۱۱- ۱۲- ۱۵- ۱۵- ۱۵- ۱۶- ۱۷- ۱۷- ۱۸- ۱۸- ۱۸- ۱۸- ۱۹- ۱۹- ۱۹- ۲۰-
 ۲۰- ۲۲- ۲۲- ۲۲- ۲۳- ۲۳- ۲۴- ۲۵- ۲۵- ۲۶- ۲۷- ۲۷- ۲۷- ۲۸- ۲۹- ۳۰- ۳۰-
 ۳۳- ۳۳- ۳۴- ۳۵- ۳۵- ۳۷

طبقات	f	fc	رتبه درصدی
۳۵-۳۷	۳	۴۰	$28/5 \times \frac{100}{40} = 96/25$
۳۲-۳۴	۳	۳۷	$25/5 \times \frac{100}{40} = 88/25$
۲۹-۳۱	۳	۳۴	$22/5 \times \frac{100}{40} = 81/25$
۲۶-۲۸	۵	۳۱	$28/5 \times \frac{100}{40} = 71/25$
۲۳-۲۵	۵	۲۶	$23/5 \times \frac{100}{40} = 58/25$
۲۰-۲۲	۵	۲۱	$18/5 \times \frac{100}{40} = 46/25$
۱۷-۱۹	۹	۱۶	$11/5 \times \frac{100}{40} = 28/25$
۱۴-۱۶	۳	۷	$5 \times \frac{100}{40} = 12/5$
۱۱-۱۳	۳	۳	$1/5 \times \frac{100}{40} = 3/25$

در این مثال: $71/25$ درصد افراد، دارای نمراتی زیر نمرهٔ افراد طبقهٔ ۲۸-۲۶ هستند. بهترین نتیجه‌ای که می‌توانید به دانشجو بدهید رتبهٔ درصدی اوست تا به این وسیله از موقعیت خود بین همکلاسه‌هایش آگاه شود؛ ولی هیچگاه رتبه‌های درصدی را به عنوان رتبهٔ کلی دانشجو به او ارائه نکنید، بلکه مجموع نمرات دانشجو را در چندین آزمون محاسبه کرده، آنگاه رتبهٔ درصدی او را حساب کنید.

شاخص نسبت صحیح^۱

احتمالاً معمولترین روش محاسبه شاخص دشواری سؤال برای سؤالات با مقیاس دوتایی (با نمرات صفر و یک)، شاخص مقدار P^1 می‌باشد. مقدار P نسبتی از آزمون‌شوندگان است که به یک سؤال ویژه، پاسخ صحیح داده‌اند. به عنوان مثال اگر یک سؤال را به صد نفر آزمون‌شونده بدهیم و هشتاد نفر ایشان به سؤال مزبور پاسخ صحیح دهند، مقدار P برابر با $0/8$ می‌باشد که نشان‌دهنده آن است که هشتاد درصد آزمون‌شوندگان به سؤال، پاسخ صحیح داده‌اند. یا اگر یک سؤال دیگر از ۳۱۱ آزمون‌شونده پرسیده شود و ۱۸۷ نفر ایشان پاسخ صحیح بدهند، مقدار P معادل $0/6$ خواهد بود (زیرا

$$\frac{187}{312} = 0/6$$

مقدار P به تنهایی بیان‌کننده خوب یا بد بودن یک سؤال نیست و فقط شاخصی است که نشان‌دهنده دشواری سؤال برای یک گروه خاص از آزمون‌شوندگان است. زیرا اگر همان سؤال از گروه متفاوتی از آزمون‌شوندگان پرسیده شود، احتمالاً مقدار P قابل پیش‌بینی نخواهد بود. جنبه‌های متفاوت مقدار P با توجه به گروه، در اصطلاح وابستگی به نمونه^۲ متجلی است و به مفهوم آن است که مقدار P به گروه ویژه‌ای که آزمون برای آنها اجرا می‌شود، بستگی دارد.

مثال: فرض کنید یک سؤال خاص را در مورد درمان فشار خون از دانشجویان سال اول و ششم پزشکی به صورت جداگانه پرسید، به احتمال بسیار قوی مقدار P در دو گروه متفاوت خواهد بود؛ چرا که پاسخ به چنین سؤالی برای گروه دوم راحت‌تر از گروه اول می‌باشد. بنابراین در انتخاب یک نمونه برای محاسبه ارزش P ، باید نمونه به نحوی انتخاب و نمونه‌گیری به گونه‌ای طراحی شود که واقعاً نماینده جمعیت مورد مطالعه باشد؛ که در این صورت نتایج به دست آمده قابلیت تعمیم‌پذیری مناسب دارد. این خصوصیت وابستگی به نمونه، یک جنبه مهم مقدار P به حساب می‌آید و همانگونه که خواهیم دید، سایر روشهای آماری در پی ارائه شاخصهای دشواری غیروابسته به نمونه آزمون‌شوندگان می‌باشد.

استفاده از مقدار P برای تجزیه و تحلیل سؤال

مقدار P بزرگترین کمک را به نویسنده سؤال در تجزیه و تحلیل آن می‌کند. با توجه به

1- The Proportion Correct Index

2- P-value

3- Sample Dependence

جدول ۱-۱۰. مقادیر P برای یک سؤال

	A	B	C*	D	بدون پاسخ	جمع کل
تعداد	۲۸	۱۷	۱۹۷	۴۱	۳	۲۸۶
مقدار P	۰/۱	۰/۰۶	۰/۶۹	۰/۱۴		

* پاسخ صحیح

مقدار P می‌توان به برخی اشتباهات طراح سؤال پی برد که از آن جمله است: عبارت‌بندی ظاهری سؤال که منجر به سردرگمی آزمون‌شوندگان می‌شود؛ شناسایی گزینه‌های معیوب و شناسایی سؤالاتی که سهواً کلیدشان اشتباه شده است. این شاخص همچنین می‌تواند در بررسی نحوه عملکرد در دو گروه یک جمعیت که تواناییهای کم و زیاد دارند، به نویسنده سؤال کمک نماید.

در ادامه، هر کدام از راهکارهای استفاده‌شده مقدار P توضیح داده خواهد شد. اما نخست باید این مقدار به نحوی بکار رود که تفسیر را آسان سازد.

ممکن است گاهی مقدار P به صورت منفرد به کار رود. اما برای تجزیه و تحلیل سؤالات مقدار P برای تمام گزینه‌ها، گزینه صحیح و گزینه‌های غلط محاسبه می‌شود. وقتی مقدار P به این صورت مورد استفاده واقع می‌گردد، تعدادی سؤال بی‌پاسخ نیز گزارش می‌گردد. جدول ۱-۱۰ مثالی از مقادیر P که برای هر گزینه گزارش شده، همراه با تعداد افرادی را که به سؤال پاسخ نداده‌اند، نشان می‌دهد.

همانگونه که در جدول ۱-۱۰ مشاهده می‌شود مقادیر P اطلاعات زیادی در جهت بهبود و اصلاح سؤالات به ما می‌دهد. در این مورد شاخص مزبور برای پاسخ صحیح ۰/۶۹ بوده است که نشان می‌دهد سؤال برای آزمون‌شوندگان نه چندان ساده و نه چندان دشوار بوده تقریباً تمام گزینه‌های غلط توسط ایشان انتخاب شده است. در اغلب موارد، مقدار P شاخص خوبی برای قضاوت در مورد کیفیت یک آزمون است. به هر حال، از آنجا که تعداد نسبتاً کمی (شش درصد) از آزمون‌شوندگان گزینه B را انتخاب کرده‌اند، بهتر است این گزینه غلط تا حد امکان اصلاح شود.

با در نظر گرفتن شانس در پاسخگویی به سؤالات چندگزینه‌ای انتظار داریم دانشجویان متوسط حدود نیمی از سؤالات را درست پاسخ گویند، در چنین آزمونهایی اثر

جدول ۱-۲. ارتباط میانگین مطلوب مقادیر P و نوع سؤالات آزمون.

نوع سؤال	میانگین مطلوب مقادیر P آزمون
کامل کردنی و کوتاه پاسخ	۰/۵۰
پنج‌گزینه‌ای	۰/۷
چهارگزینه‌ای	۰/۷۴
سه‌گزینه‌ای	۰/۷۷
صحیح - غلط (دو گزینه‌ای)	۰/۸۵

شانس با آسان شدن سؤالات کاهش می‌یابد و در نتیجه پایایی آزمون افزایش پیدا می‌کند. جدول ۱-۲ برای ضریب دشواری مطلوب در انواع آزمون‌ها ارائه شده است. البته باید توجه داشت که اعداد مزبور ارائه‌کننده مقادیری از P هستند که می‌توان به وسیله آنها حداکثر افتراق را در بین آزمون‌شوندگان قائل شد. ولی به طور کلی مقادیر مناسب P در آزمون‌های پیشرفت تحصیلی دانشگاهی در گستره‌ای از ۰/۴ تا ۰/۸ قرار دارد. حالا به مقادیر P که در جدول ۱-۲ برای یک سؤال دیگر ارائه شده است توجه نمایید. در این مورد مقدار P نمایان‌کننده چند عیب در این سؤال است. بسیاری از پاسخ‌دهندگان در ابهام مانده‌اند. این موضوع با پراکندگی پاسخ به گزینه‌های غلط جایگزین قابل مشاهده است. به عنوان مثال گزینه صحیح A کمتر از گزینه غلط C، انتخاب شده است (۲۷ درصد در مقابل ۲۵ درصد) و گزینه B، چنان نامعقول و غیرقابل قبول بوده که هیچ کس آن را انتخاب نکرده است. تعداد نسبتاً زیادی نیز هیچ گزینه‌ای را انتخاب نکرده‌اند که دلیلی دیگر بر مبهم بودن سؤال است. این یافته اشکال جدی در سؤال را مطرح می‌کند. باید توجه داشت، تهیه یک سؤال خوب با نوشتن یک سؤال پایان نمی‌یابد و برای این کار نیاز به تجزیه و تحلیل مناسب و بازنگری سؤالات در مراحل متعدد است.

سؤالات با کلید اشتباه

یکی دیگر از موارد استفاده از مقدار P، شناسایی سؤالاتی است که کلیدشان اشتباه شده است. متأسفانه وجود سؤالاتی با کلید اشتباه، زیاد اتفاق افتاده، در بسیاری از موارد به راحتی نمی‌توان چنین اشتباهی را شناسایی نمود. گاهی اوقات خستگی ناشی از مراحل تهیه سؤال سبب اعمال در توجه به جزئیات گردیده، کلید سؤال اشتباه می‌شود. در سایر موارد،

جدول ۳-۱۰. مقادیر P برای یک سؤال ضعیف

	A*	B	C	D	بدون پاسخ	جمع کل
تعداد	۷۷	۰	۱۳۰	۶۳	۱۶	۲۸۶
مقدار p	۰/۲۷	۰	۰/۴۵	۰/۲۲		

* پاسخ صحیح

جدول ۴-۱۰. مقادیر P برای یک سؤال با کلید اشتباه

	A	B*	C	D	بدون پاسخ	جمع کل
گزینه‌ها						
تعداد	۲۰۲	۳۱	۲۸	۲۵	۰	۲۸۶
مقادیر P	۰/۷۱	۰/۱۱	۰/۱۰	۰/۹		

* پاسخ صحیح

به ابهام و درک نادرست از سؤال یا عبارت‌بندی ضعیف آن توجه نمی‌شود و پاسخ صحیح آشکار نیست. گاهی هم ظرافتهای زبانی یا دشواری محتوا منجر به اشتباه در کلید سؤال می‌شود. وقتی سؤالاتی که کلیدشان اشتباه شده مشخص می‌شود که طراح آزمون جدولی از مقادیر P را برای تمامی سؤالات تهیه کند و آنگاه ببیند نتایج با آنچه او انتظار داشته، تفاوت بسیاری دارد.

به عنوان مثال جدول ۴-۱۰ نمایان‌کننده مقادیر عددی مربوط به یک سؤال است که در آن گزینه B به عنوان گزینه صحیح در نظر گرفته شده ولی اغلب آزمون‌شوندگان گزینه A را انتخاب نموده‌اند؛ این امر احتمالاً نشان‌دهنده اشتباه در کلید سؤال بوده است.

مقایسه مقادیر P برای گروههای بالا و پایین

گاهی مقایسه گروهها در جمعیت مورد آزمون مشخص‌کننده عملکرد یک سؤال است. در اغلب موارد به این منظور، جمعیت به دو گروه بالا و پایین تقسیم می‌شود که معمولاً شامل ۲۷ درصد از آزمون‌شوندگانی است که بیشترین نمره کل را آورده‌اند و ۲۷ درصد از کسانی که کمترین نمرات را کسب نموده‌اند. مبنای انتخاب ۲۷ درصد بر اساس محاسبات کلی^۱ (۱۹۳۹) می‌باشد. او نشان داد این عدد شاخصی پایدار از تفاوت بین گروههای پایین و بالا ارائه می‌کند. علاوه بر اینکه در تعیین شاخصهای پایایی درونی نیز از

1- Kelly

جدول ۵-۱۰. مقادیر P برای زیرگروه‌های آزمون‌شوندگان در مورد یک سؤال

بدون پاسخ	D	C	B*	A	گزینه‌ها
۱	۰/۰۲	۰/۰۸	۰/۶۱	۰/۲۹	مقادیر P برای ۲۷٪ بالا
۶	۰/۱۱	۰/۳۱	۰/۲۷	۰/۳۱	مقادیر P برای ۲۷٪ پایین
	-۰/۰۹	-۰/۲۳	۰/۳۲	-۰/۰۲	تفاوت

* پاسخ صحیح

همین عدد استفاده می‌شود. البته در کلاسهای کوچک با دانشجویان محدود رعایت عدد ۲۷ درصد الزامی نیست و به دلیل مشکلات ناشی از آن می‌توان کل آزمون‌شوندگان را به دو گروه بالا و پائین تقسیم نمود.

مقادیر جدول ۵-۱۰ نشان می‌دهد که تفاوت دو گروه در مورد گزینه صحیح ۳۲ درصد می‌باشد. این تفاوت احتمالاً بیان‌کننده آن است که سؤال برای آزمون‌شوندگان گروه بالا چندان دشوار نبوده ولی برای گروه پایین تا حدی مشکل بوده است. در برخی از سنجشها این نمای متفاوت برای یک سؤال مطلوب است. تفاوت بین گروهها برای سایر پاسخهای پیشنهادی مقدار منفی دارد. چنین مقادیر منفی نشان‌دهنده آن است که آزمون‌شوندگان با توان بیشتر کمتر از آزمون‌شوندگان با توانایی کمتر پاسخهای غلط را انتخاب کرده‌اند. به علاوه معمولاً این نما برای گزینه صحیح یک سؤال مثبت است. البته طراح آزمون باید توجه کند که ۰/۰۲ درصد اختلاف بین گروهها برای یک پاسخ غلط کم است و این مطلب مؤید آن است که گزینه A باید مجدداً مورد بازبینی قرار گیرد.

مثال: به مقادیر مربوط به سؤال دیگری که در جدول ۶-۱۰ آورده شده است، توجه کنید. توجه به نمای بالا نشان می‌دهد A و C گزینه‌های انحرافی خوبی هستند؛ اما D گزینه

جدول ۶-۱۰. مقادیر P برای زیر گروه‌های آزمون شوندگان در مورد یک سؤال.

بدون پاسخ	E	D	C	B*	A	گزینه‌ها
۰	۰/۱۷	۰	۰	۰/۸۲	۰	گروه بالا
۰	۰	۰	۰/۸	۰/۵۰	۰/۴۲	گروه پایین
	۰/۱۷	۰	-۰/۸	۰/۳۳	-۰/۴۲	تفاوت

* پاسخ صحیح

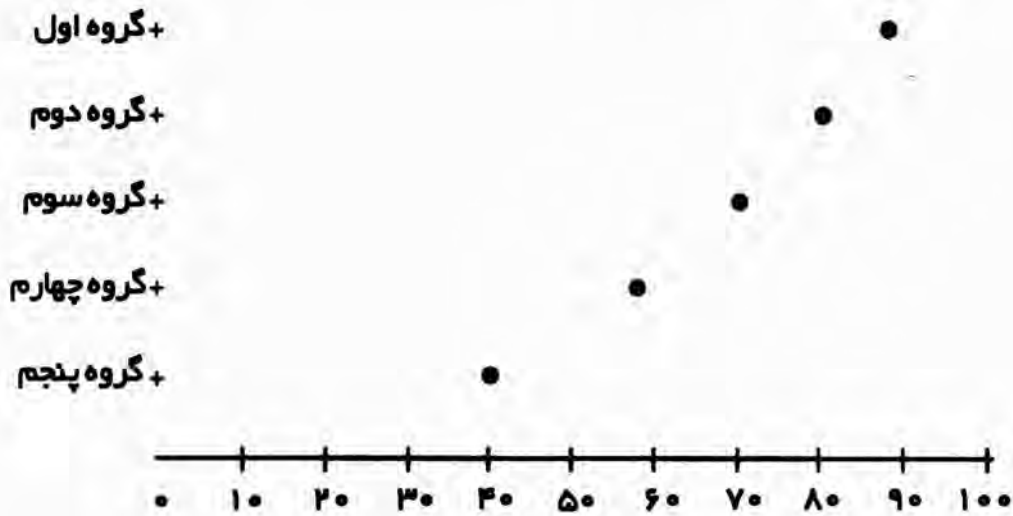
ضعیفی است زیرا هیچکس به ویژه افراد گروه پایین آن را انتخاب نکرده‌اند، E هم گزینه‌ی ضعیفی است که فقط افراد گروه بالا آن را انتخاب کرده‌اند. در این حال باید از خود بپرسیم چرا دانشجویان قویتر گزینه‌ی E را انتخاب کرده‌اند؟ آیا این امر به علت ابهام در سؤال یا گزینه‌ها بوده است؟ آیا دو پاسخ صحیح وجود داشته است؟ یا این امر مربوط به بی‌احتیاطی بخشی از آزمون‌شوندگان بوده است؟

انجام مقایسه چند طبقه‌ای

گاهی در تجزیه و تحلیل سؤالات، چند بخش جمعیت با یکدیگر مورد مقایسه قرار می‌گیرند. هر بخش یا طبقه جمعیت نماینده گروه و ویژه‌ای از آزمون‌شوندگان با مقدار مشخصی از توانایی است. اگر آزمون‌شوندگان براساس یک ملاک خوب با محتوای یکسان و ضریب پایایی مشابه مورد ارزیابی قرار گرفته دسته‌بندی شوند، ایده‌آل است. اما در عمل معمولاً چنین امکانی وجود ندارد و از نمره کل آزمون برای طبقه‌بندی افراد استفاده می‌شود. در مواردی که کل امتحان‌شوندگان حدود دو‌ست نفر یا بیشتر باشند می‌توان آنها را به پنج دسته که هر دسته نمایان‌کننده بیست درصد جمعیت است، طبقه‌بندی نمود. نمونه‌ای از این تقسیم‌بندی در نمودار ۱-۱۰ آورده شده است.

از آنجا که معمولاً توزیع نمرات دارای انحراف است، لازم نیست حتماً همه گروهها دارای ۲۰ درصد جمعیت باشند بلکه اگر اغلب آنها به ۲۰ درصد نزدیک باشند، کافی است. در مواردی که جمعیت به پنج طبقه تقسیم شود بهتر است از مقدار P به صورت نمودار استفاده گردد. (مانند نمودار ۱-۱۰ که در آن نمودار تصویری برای یک سؤال در ارتباط با زیر گروهها و مقدار P آورده شده است). در نمودار ۱-۱۰ مشاهده می‌شود که نسبت پاسخ صحیح بالاترین گروه که بیشترین نمره کل را آورده‌اند، بیش از گروههای دیگر است و به ترتیب، این نسبت در گروههای پایین‌تر کاهش می‌یابد. در موارد تفسیر آزمون براساس آزمونهای ملاک نسبی، سؤال مزبور می‌تواند در مورد نمره کل کسب شده توسط آزمون‌شونده پیش‌بینی صحیحی بدهد. میزان پاسخهای صحیح گروههای با توانایی کمتر به طور متناسب افزایش می‌یابد.

نمودار ۲-۱۰ برای سؤال دیگری ترسیم شده است. در اینجا سؤال عملکرد مناسبی نداشته، به تجدیدنظر نیاز دارد. همانطور که مشاهده می‌شود هیچ آزمون‌شونده‌ای با هر اندازه از توانایی گزینه انحرافی A را انتخاب نکرده است. همچنین اغلب آزمون‌شوندگان با

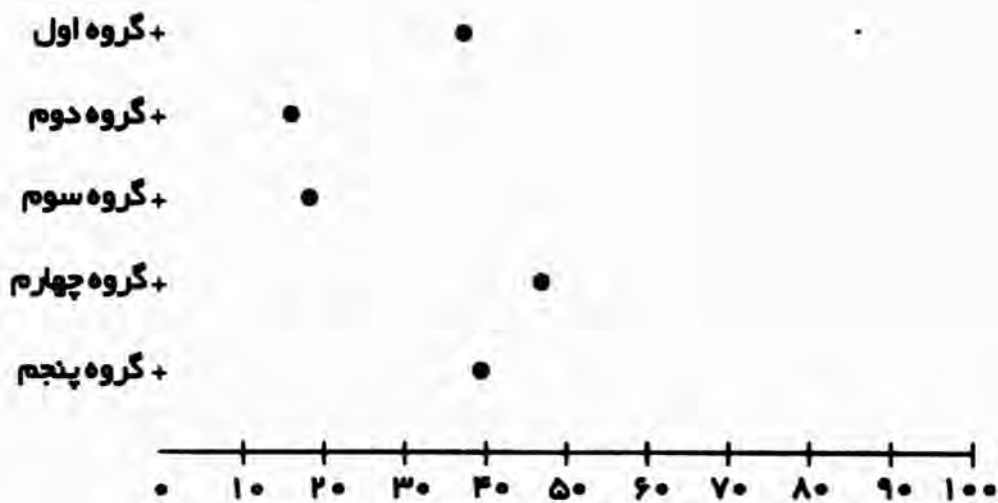


بدون پاسخ	A	B	C*	D	
۰	۸	۱۷	۲۰۷	۱۱	گروه پنجم
۰	۱۱	۳۹	۲۲۳	۲۵	گروه چهارم
۰	۱۶	۵۳	۱۸۳	۳۲	گروه سوم
۰	۱۷	۶۱	۱۱۳	۲۹	گروه دوم
۲	۲۳	۷۰	۱۰۶	۳۹	گروه اول
	۰/۰۷	۰/۱۸	۰/۶۴	۰/۱۰	مقدار P

* مقدار صحیح

نمودار ۱-۱۰. مقادیر P برای زیر گروه‌های مختلف آزمون‌شوندگان.

توانایی‌های مختلف، گزینه انحرافی C را به جای گزینه صحیح B انتخاب کرده‌اند، که احتمالاً به علت جمله‌بندی ضعیف گزینه‌هاست. این نمودار نمایان‌کننده آن است که دو گروه پایین نسبت به گروه بالا عملکرد بهتری داشته‌اند که به علت کلمات مبهم یا مشاهده پدیده بازداری پیش‌گستر^۱ می‌باشد. این پدیده یکی از انواع تورش‌های نظام‌دار است که گاهی به علت دقت کم در تهیه و انتخاب گزینه‌های انحرافی به وجود می‌آید. در این حالت به دلیل آنکه برخی از آزمون‌شوندگان با توانایی بالا، اطلاعاتی بیش از حد انتظار طراح آزمون دارند، با استفاده از آن اطلاعات، برخی گزینه‌های انحرافی را به نحو خاصی تعبیر و تفسیر می‌کنند و از انتخاب



بدون پاسخ	A	B*	C	D	
۰	۰	۲۷	۱۱۱	۲۱	گروه پنجم
۰	۰	۲۳	۱۱۶	۲۱	گروه چهارم
۰	۰	۱۶	۱۲۳	۱۳	گروه سوم
۰	۰	۱۰	۸۳	۳۲	گروه دوم
۲	۰	۸	۹۷	۲۴	گروه اول
	۰/۰۰	۰/۱۱	۰/۷۳	۰/۱۵	مقدار P

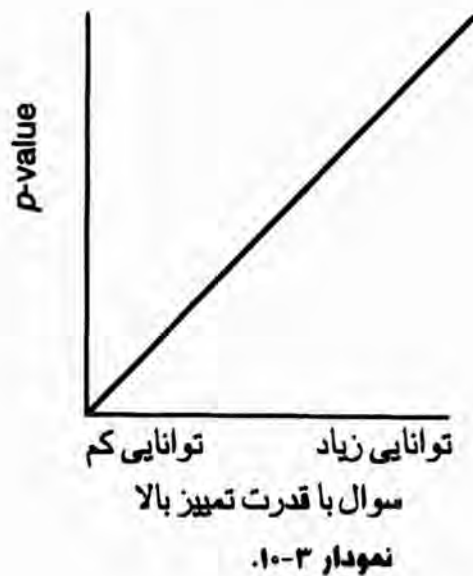
* مقدار صحیح

نمودار ۲-۱۰. مقادیر P برای یک سؤال فرضی ضعیف.

گزینه صحیح باز می‌مانند. در این مورد، سؤال به بازنگری و اصلاح کلی نیاز دارد به هرحال این نمونه یکی از موارد استفاده نمودارها و تجزیه و تحلیل آماری برای بهتر کردن سؤالات است.

شاخصهای تمیز سؤال

قدرت تمیز یکی دیگر از مفاهیم مهم برای قضاوت در مورد کیفیت آزمونهاست. شاخص تمیز یک سؤال می‌تواند به عنوان عامل ارتباط بین دشواری یک سؤال و توانایی آزمون‌شوندگان بیان گردد. قدرت تمیز سؤال، شاخصی است برای ایجاد افتراق بین افراد



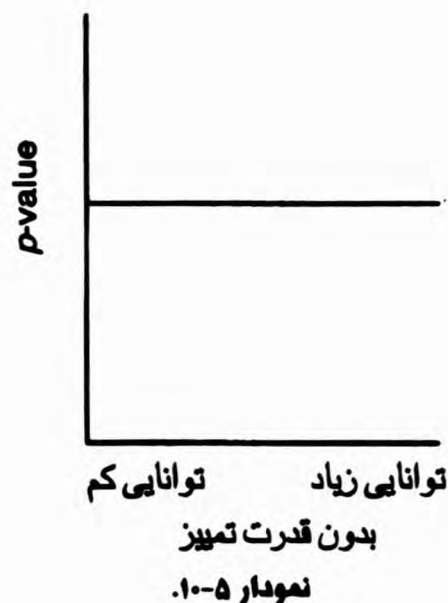
آزمون‌شونده در زمینه موضوع مورد بررسی یا سازه روانشناختی که باید اندازه‌گیری شود. اساس آن، این فرضیه است: کسانی که توان بیشتری دارند نسبت به کسانی که توانایی کمتری دارند باید به یک سوال ویژه، پاسخ صحیح بیشتری دهند. بنابراین سؤالاتی که تمام افراد به آن پاسخ صحیح دهند یا هیچ فردی نتواند به آن پاسخ گوید قدرت تمیز نخواهد داشت، زیرا اساساً نتوانسته است اطلاعاتی راجع به تفاوت بین افراد گوناگون بدهد.

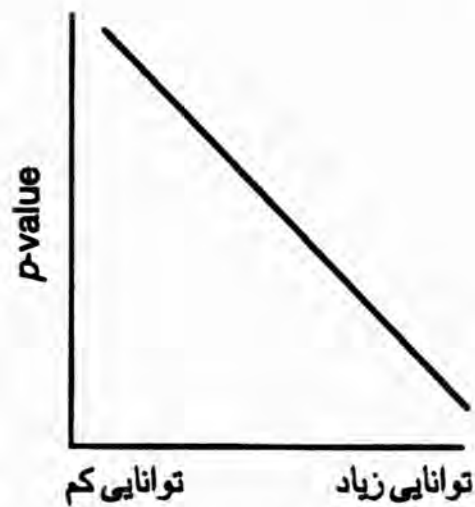
از نظر محاسباتی، شاخص تمیز سوال، دشواری سوال را به عنوان تابع توانایی جمعیت امتحان‌دهندگان در مجموعه مورد مطالعه تعریف می‌نماید. به عبارت دیگر، برای یک توانایی ویژه، تمیز در ارتباط با دشواری، ارتباط آنها در چهار نمودار آورده شده است: چهار نمودار ۱۰-۳ تا ۱۰-۶ بیان‌کننده چهار سطح تمیز است: قدرت تمیز عالی، قدرت تمیز متوسط، بدون قدرت تمیز (صفر) و با قدرت تمیز منفی.

سوال نشان داده شده در نمودار ۱۰-۳ سوال خوبی است، زیرا قدرت تمیز آن عالی می‌باشد. کسانی که توان بیشتری داشته‌اند، توانسته‌اند به این سوال پاسخ گویند و آنها که توان کمی داشته‌اند به این سوال پاسخ صحیح نداده‌اند. در نمودار ۱۰-۳ توجه کنید که هر چه توانایی آزمون‌شونده بیشتر می‌شود، ضریب دشواری نیز افزایش می‌یابد. داده‌های جدول ۱۰-۱ نیز نمایان‌کننده یک قدرت تمیز عالی است. نمودار ۱۰-۴ نمایان‌کننده ضریب تمیز متوسطی است. این نمودار نمایان‌کننده برقراری تمایز بین آزمون‌شوندگان است اما قدرت سوال نمودار ۱۰-۳ را ندارد.



نمودار ۵-۱۰ نمایان‌کننده یک سؤال بدون قدرت تفکیک بین دو گروه بالا و پایین است. احتمالاً سؤالی است که تمام پاسخ‌دهندگان به آن پاسخ صحیح و یا غلط داده‌اند. معمولاً چنین سؤالاتی اطلاعات کمی را به ما می‌دهند و اصلاح یا حذف می‌گردند. اما باید به خاطر داشت سؤالات با ضریب تمیز صفر نباید بدون درنگ و با عجله حذف شود زیرا انتظار می‌رود تمام پاسخ‌دهندگان به برخی سؤالات که مهارتهای مهم را برای یک هدف ویژه موردنظر آزمون اندازه‌گیری می‌کند، پاسخ صحیح دهند.





سوال با قدرت تمییز منفی

نمودار ۶-۱۰.

سرانجام نمودار ۶-۱۰ نمایش‌دهنده یک ضریب تمییز منفی است. در چنین سؤالاتی افراد با توانایی کمتر بیش از افرادی که توان بیشتری دارند به سؤال پاسخ صحیح داده‌اند که معمولاً به علت ضعف در واژه‌پردازی یا استفاده از کلماتی است که افراد با دانش بیشتر را سردرگم نموده است. در این مورد هم پدیده بازداری پیش‌گستر برای سؤال اتفاق افتاده است. سؤالات با قدرت تمییز منفی نیازمند بازنگری و اصلاح و یا احتمالاً حذف هستند.

ضریب تمییز مطلوب براساس دیدگاه‌های کارشناسان مشخص می‌گردد، البته برخی روش‌های آماری نیز در این زمینه راهگشا است. به هر حال، فقط در مواردی که در پی یافتن افراد بسیار معدودی در بین گروه زیادی هستیم، باید ضریب تمییز سؤالات خیلی زیاد باشد.

ضریب همبستگی دو رشته‌ای نقطه‌ای^۱

ضریب همبستگی دو رشته‌ای نقطه‌ای یکی از شاخص‌های تمییز است که با استفاده از آن می‌توان درجه ارتباط بین دو متغیر یعنی یک سؤال تنها از آزمون و نمره کل آزمون را تخمین زد. چنان که در قسمت قبل گردید نمره کل امتحان، به عنوان نمادی قابل قبول از توانایی آزمون‌شونده در حیطه موردنظر است.

باید توجه داشت محاسبه ضریب همبستگی دو رشته‌ای نقطه‌ای وقتی سودمند است

1- Point Biserial Correlation Coefficient

که سؤال آزمون دوارزشی^۱ باشد. (به عنوان مثال: پاسخ سؤالات یا فقط صحیح یا فقط غلط باشد؛ مانند سؤالات صحیح - غلط یا چندگزینه‌ای) و نمره کل نیز به طور ذاتی پیوسته^۲ باشد. ذکر این نکته لازم است که ضریب همبستگی دورشته‌ای نقطه‌ای از جمله ضرایب همبستگی گشتاوری^۳ است. گشتاورها عبارتند از انحراف استاندارد نمرات پیرامون میانگین. به این انحراف‌های پیرامون میانگین، گشتاورهای نوع اول اطلاق می‌شود و توان دوم این انحراف‌ها، گشتاور نوع دوم و توان سوم آن نوع سوم و ... نامگذاری می‌شود. بنابراین، با توجه به آنکه میانگین نمرات استاندارد شده صفر است، یک انحراف استاندارد از میانگین، معادل گشتاور نوع اول است. بدین ترتیب، ضریب همبستگی دو رشته‌ای نقطه‌ای عبارت است از همبستگی بین دو سری نمره استاندارد شده. با استفاده از این آماره، طراح آزمون می‌تواند ضریب تمیز یک سؤال را اندازه‌گیری کند. به عنوان مثال به داده‌های ارائه شده در جدول ۷-۱۰ توجه کنید. در اینجا چند سؤال که از نظر دشواری روند نزولی دارند ردیف و ضریب تمیز با استفاده از ضریب همبستگی دو رشته‌ای نقطه‌ای محاسبه و برآورد شده است. معمولاً

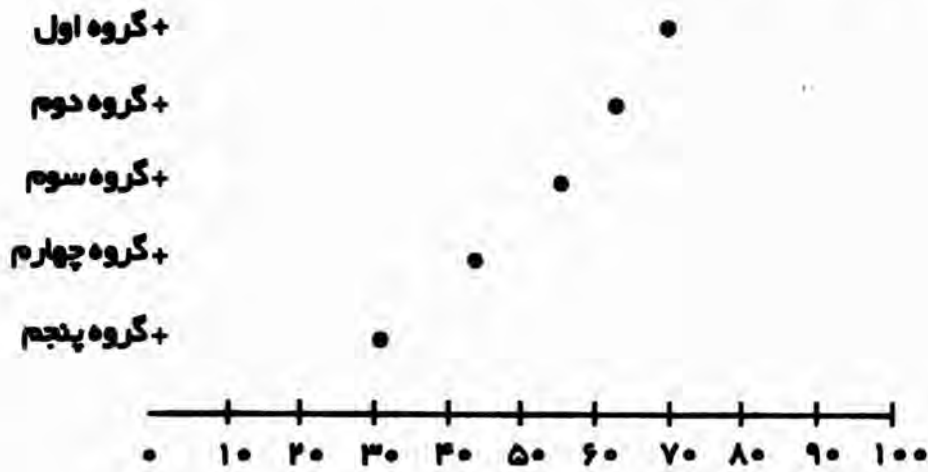
جدول ۷-۱۰. ضرایب دشواری و تمیز برای یک سؤال فرضی.

سؤال	مقدار P	ضریب همبستگی دو رشته‌ای نقطه‌ای
۳	۰/۹۲	-۰/۰۹
۱	۰/۸۶	۰/۰۶
۱۶	۰/۷۵	۰/۱۲
۹	۰/۷۳	۰/۲۵
۲	۰/۶۸	۰/۱۵
۱۱	۰/۶۲	-۰/۲۱
۱۵	۰/۶۰	۰/۳۱
۸	۰/۵۵	۰/۴۶
۸	۰/۵۱	۰/۲۵
۱۴	۰/۲۸	-۰/۲۱

1- Dichotomous

2- Continuous

3- Product-moment Correlation Coefficient



حذف	A	B	C*	D	
•	۲۱	۳	۱۹۳	۲۷	گروه پنجم
•	۱۵	۲	۱۲۵	۵۹	گروه چهارم
•	۳۰	۱۱	۱۶۱	۸۳	گروه سوم
•	۴۶	۲۳	۱۲۱	۸۷	گروه دوم
۱	۵۰	۲۸	۸۴	۸۱	گروه اول
	۰/۰۰	۰/۰۶	۰/۵۴	۰/۲۷	مقدار P
	-۰/۱۴	-۰/۲۰	۰/۲۹	-۰/۱۱	ضریب همبستگی دو رشته‌ای نقطه‌ای

* مقدار صحیح

نمودار ۷-۱۰. مقادیر P و ضریب همبستگی دو رشته‌ای نقطه‌ای برای یک سؤال.

سؤالاتی که ضریب همبستگی بالاتر دارند، دارای قدرت تمیز زیادتری هستند. برعکس، سؤالاتی که ضریب همبستگی دو رشته‌ای نقطه‌ای آنها کمتر است، قدرت تمیز کمتری دارند. به عنوان یک دستورالعمل کلی، سؤالاتی که ضریب همبستگی دو رشته‌ای نقطه‌ای آنها منفی است، باید حذف یا اصلاح گردند. می‌توانید با تلفیق نتایج حاصل از مقادیر P و ضرایب همبستگی دو رشته‌ای نقطه‌ای، کیفیت سؤالات را بررسی نمایید. وقتی یک سؤال ضریب دشواری مناسب داشته باشد (براساس هدف آزمون در گستره $0/۸۰ < P < 0/۴۰$) و قدرت تمیز آن نیز بالا باشد، آن را به عنوان سؤالی مناسب در نظر بگیرید. به عنوان مثال: اولین سؤال در جدول ۷-۱۰ (سؤال شماره سه آزمون) براساس ملاک بسیار ساده بوده، قدرت

تمیز آن ضعیف است. به هر حال این سؤال به عنوان یک سؤال معیوب باید بازبینی شود. سؤال شماره چهارده، بسیار مشکل است و باید مورد تجدیدنظر قرار گیرد. این سؤال برای اکثر آزمون‌شوندگان مبهم بوده است، زیرا آن تعداد کمی از آزمون‌شوندگان که به سؤال پاسخ صحیح گفته‌اند، به گروه توانمند تعلق ندارند. سؤال شماره نه خوب بوده، زیرا دشواری آن در گستره قابل قبول بوده، قدرت تمیز به نسبت بالایی دارد.

در اغلب موارد مقایسه ضریب همبستگی دورشته‌ای نقطه‌ای برای گزینه‌های صحیح و غلط به تجزیه و تحلیل سؤالات کمک می‌کند. انتظار آن است که قدرت تمیز برای یک سؤال واحد، در مورد گزینه صحیح، مثبت و برای سایر گزینه‌ها، منفی باشد. نمودار ۷-۱۰ نشان‌دهنده چنین آماره‌هایی در مورد یک سؤال است که این وضعیت را نشان می‌دهد. به این وسیله می‌توانید سؤالاتی را که دارای سطح بالایی از قدرت تمیز است، شناسایی نمایید.

معایب استفاده از ضریب دورشته‌ای نقطه‌ای

با وجود استفاده مکرر از ضریب همبستگی دورشته‌ای نقطه‌ای در تجزیه و تحلیل آزمونها، این آماره خالی از اشکال نیست. تأثیر نمره مربوط به هر سؤال در نمره کل یا اندازه توانایی این نگرانی را ایجاد می‌کند که اطلاعات به دست آمده در حقیقت مبنای درستی نداشته، تفاسیر اشتباه صورت پذیرفته است، اما با آنکه از نظر منطقی، این نظریه صحیح است، در عمل، اثر یک سؤال در نمره کل به حداقل می‌رسد؛ به ویژه هنگامی که تعداد سؤالات نسبتاً زیاد است (مثلاً ۲۵ سؤال یا بیش از آن) احتمالاً جای نگرانی نیست. اما وقتی ارزیابی بسیار دقیق مطلوب باشد یا تعداد سؤالات کمتر از ۲۵ سؤال باشد، این مشکل را باید مدنظر قرار دهید. از آنجا که نحوه محاسبه پیچیده بوده، در عمل به ندرت مورد استفاده است، از ذکر مباحث مربوط خودداری می‌گردد.

به هر حال حتی با توجه به مشکلاتی که گاهی در هنگام استفاده از این شاخص از نظر تکنیکی رخ می‌دهد، این شاخص بر شاخصهای نسبی قدرت تمیز برتری دارد.

همبستگی دورشته‌ای تخمینی^۱

یکی دیگر از آماره‌هایی که در تجزیه و تحلیل سؤالات به عنوان شاخص تمیز استفاده می‌شود، همبستگی دورشته‌ای تخمینی است. با تمام شباهتی که این آماره با

شاخص همبستگی دورشته‌ای نقطه‌ای دارد، یک تفاوت مهم این دو شاخص را از هم متمایز می‌سازد: در حالی که ضریب همبستگی دورشته‌ای نقطه‌ای وقتی محاسبه می‌شود که یکی از دو متغیری که با هم همبسته‌اند، حقیقتاً دو ارزشی^۱ باشند، در همبستگی تخمینی دورشته‌ای، هر دو متغیر به طور ذاتی پیوسته‌اند. همچنین فرض بر آن است که نمرات هر دو متغیر توزیع طبیعی (نرمال) دارند. به منظور تسهیل در محاسبه، یکی از دو متغیر (به‌طور دلخواه) به دو گروه تقسیم می‌شود. مثلاً: دو گروه آزمون‌شونده‌ای که درست یا نادرست به یک سؤال پاسخ داده‌اند یا دو گروهی که در یک آزمون (مثلاً بر مبنای نمره ده از بیست) قبول یا رد شده‌اند. در این جا یک متغیر پیوسته را به یک متغیر دو ارزشی تبدیل نموده‌ایم. این آماره محدودیت‌های تکنیکی ضریب همبستگی دو رشته‌ای نقطه‌ای را ندارد.

اگر بتوان فرض کرد که سازه مورد اندازه‌گیری در جمعیت آزمون‌شوندگان دارای توزیع طبیعی است، گستره این ضریب همبستگی بین -۱ و +۱ محدود می‌گردد. بنابراین، ضریب همبستگی تخمینی از آزمون‌شوندگانی که توانایی خیلی کم یا خیلی زیاد دارند، بیشتر تأثیر می‌پذیرد. پس در مواردی که بخواهید سؤالاتی را که در انتهای درجه دشواری است، این ضریب بر ضریب همبستگی دورشته‌ای نقطه‌ای ارجحیت دارد.

ضریب فی^۲

ضریب همبستگی فی یکی دیگر از ضرایب همبستگی است، که وقتی هر دو متغیر همبسته به صورت واقعی دو ارزشی هستند، در امر تجزیه و تحلیل سؤالات آزمونها قابل استفاده است. این ضریب مانند سایر ضرایب همبستگی در محدوده -۱ و +۱ تغییر می‌کند. ضریب فی برای تشخیص درجه ارتباط یک سؤال و برخی ملاکها مانند برخی قسمت‌های برنامه، جنس یا برخی ویژگیهای جمعیت‌شناختی^۳ مورد استفاده قرار می‌گیرد.

گاهی هر خلال تجزیه و تحلیل، بررسی اثر پاسخ‌دهی به یک سؤال (به عنوان یک ملاک) میزان پاسخ‌دهی به سؤال دیگر مطلوب نظر قرار می‌گیرد. برای چنین کاری می‌توانید جدولی مانند جدول ۸-۱۰ تهیه نمایید، که در آن توزیع نسبت پاسخهای صحیح به هر سؤال در یک گروه از آزمون‌شوندگان آورده می‌شود. مثلاً در این مثال سی درصد آزمون‌شوندگان به هر دو سؤال پاسخ صحیح داده‌اند و بیست درصد آنها به هر دو سؤال پاسخ غلط داده‌اند.

1- Dichotomy

2- Phi Coefficient

3- Demographic Characteristics

جدول ۸-۱۰. جدول چهارخانه توزیع پاسخهای صحیح دو آزمون شونده برای دو سؤال.

سؤال ۱

		صحیح	غلط	
سؤال ۲	صحیح	۰/۳۰	۰/۱۷	۰/۴۷
	غلط	۰/۳۳	۰/۲۰	۰/۵۳
		۰/۵۳	۰/۴۷	

در تجزیه و تحلیل سؤالات، مقدار ضریب فی به مقایسه درجه پایداری پاسخها برای همان سؤال به وسیله آزمون شوندگان در موقعیتهای متفاوت زمانی وابستگی تام دارد. به عنوان مثال: اگر بخواهید بدانید آیا بعضی متغیرها مانند جنس با نحوه عملکرد یک گروه آزمون شونده در اجزای متفاوت آزمون آناتومی ارتباط دارد یا نه، آماره ضریب فی، آماره مناسبی است.

استفاده از ضریب فی

گاهی در روشهای تجزیه و تحلیل سؤالات، مقایسه عملکرد دو گروه در یک سری سؤال و یا در یک عملکرد ویژه مورد توجه قرار میگیرد. مثلاً می خواهید عملکرد یک گروه آموزش دیده را که انتظار می رود بتوانند به سؤالات پاسخ گویند - و به آنها گروه ملاک^۱ گوئیم - را با گروهی که آموزشی ندیده اند مقایسه نمایید. برای هر سؤال، عملکرد این دو گروه مورد مقایسه قرار میگیرد. اگر آزمونی را به صورت همزمان برای هر دو گروه برگزار نمایید، می توانید با استفاده از ضریب فی، میزان اختلاف دو گروه را مورد بررسی قرار دهید. البته استفاده از آن، محدودیتهایی نیز دارد و آشکارترین آنها، این حقیقت است که به ندرت برنامه های آموزشی چنان خوب و در طول یک سری اهداف به دقت تعیین و مشخص شده، سازمان دهی شده اند که بتوان دو گروه مناسب را شناسایی نمود. عیب دیگر

1- Criterion Group

آن است که صرف آموزش دیدن گروه ملاک می‌تواند منجر به ارتقای سایر توانمندیها گردد. به هر حال این امر می‌تواند منجر به طولانی شدن، غیرعملی شدن یا ایجاد فاصله زمانی در برگزاری آزمونها گردد. با وجود این مشکلات، اگر یک ملاک با مدرک و دلایل کافی انتخاب شود و گروههای مناسب در دسترس باشند، این روش، روشی قدرتمند برای تجزیه و تحلیل سؤالات است.

معایب ضرایب همبستگی فی

یکی از محدودیتهای ویژه ضریب فی، آن است که یک ضریب همبستگی تخمینی است و مانند ضرایب همبستگی دورشتهای نقطه‌ای و دورشته‌ای، از ضریب همبستگی قدیمی پیرسون مشتق شده است. از آنجا که تمام همبستگی‌های گشتاوری به شکل یک نمره استاندارد بیان می‌شود، دو متغیر با نمرات یکسان الزاماً به طور کامل ارتباط خواهند داشت؛ بنابراین، در مواردی که مقدار P برای دو گروه مساوی است، ضریب فی +۱ خواهد بود.

عوامل سؤال

راهکارهای ارائه شده برای تجزیه و تحلیل سؤالات در بخشهای قبلی به طور عمده بر پایه دیدگاههای سنتی تجزیه و تحلیل سؤالات بنا نهاده شده بود. حسن این روشها کاربرد ساده و آسان آنها توسط استاد است. اما در سالهای اخیر، نظریاتی جدید در این حیطه ارائه گردیده و توسعه رایانه‌های قوی و هنر برنامه‌نویسی، به کاربردی نمودن این نظریات کمک شایانی نموده است. از مهمترین این نظریات جدید، رویکرد صفت مکنون^۱ می‌باشد، که براساس آن موفقیت هر فرد در وظیفه خود مبتنی بر یک صفت یا خصیصه یا ویژگی زیربنایی است و هر چه این صفت یا خصیصه قوی‌تر باشد، فرد دارای عملکرد بهتری خواهد بود.

با استفاده از این نظریه می‌توان سؤالات آزمون را به صورت تصویری و به نحوی که برخی ویژگیهای مهم آن به آسانی قابل مرور باشد، مورد بررسی قرار داد. این کار با کشیدن خطوط کشیده شده سؤال^۲ یا کارکردهای^۳ سؤالات آزمون انجام می‌شود. از نظر روش‌شناختی می‌توان شکل‌های ۳-۱۰ تا ۶-۱۰ این فصل را به عنوان خطوط کشیده شده سؤال موردنظر قرار داد ولی در عمل خطوط بیشتری به عنوان خطوط کشیده شده سؤال

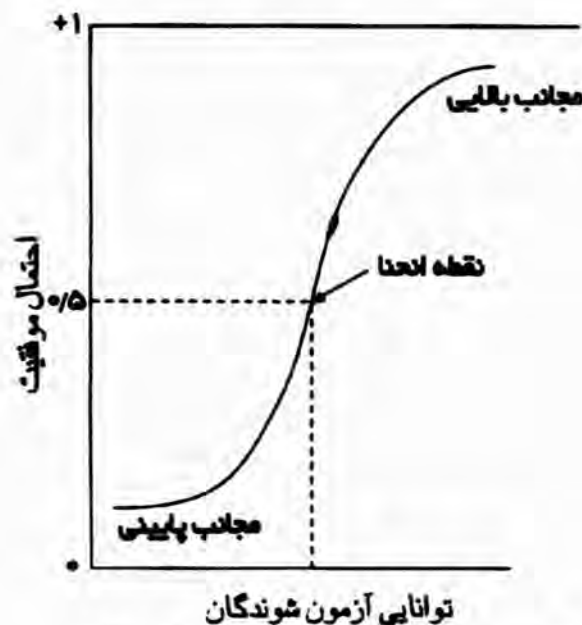
1- Latent Trait Approach
3- Functions

2- Item Trace Lines

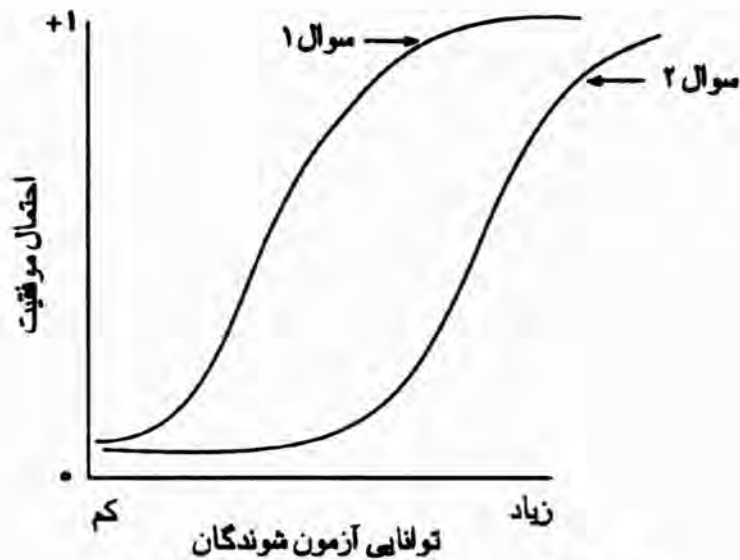
برای تجزیه و تحلیل سؤالات با رویکرد صفت مکنون مورد استفاده قرار می‌گیرد. خطوط کشیده شده سؤال در این مبحث، منحنی ویژگی سؤال^۱ نامیده می‌شود و در آنها اطلاعاتی درباره یک، دو یا سه عامل یا حد و مرز محاسباتی برای هر سؤال ارائه می‌گردد.

- ۱- عامل A: که نشانه شیب منحنی بوده، بیان‌کننده افزایش احتمال پاسخ صحیح به یک سؤال با افزایش توانایی پاسخگویان است. این عامل همان قدرت تمیز سؤال می‌باشد.
- ۲- عامل B: دشواری یک سؤال را نشان می‌دهد که با در نظر گرفتن نقطه‌ای بر منحنی که در آن متغیر مکنون (مانند یک سازه روانشناختی) شیب پیدا می‌کند تا به صورت نقطه تعبیر مکان از چپ به راست منحنی مشخص می‌گردد. به عبارت دیگر نمره معیاری است که در آن نیمی از آزمون‌شوندگان به سؤال موردنظر پاسخ صحیح داده‌اند.
- ۳- عامل C: شروع یا کف منحنی را نشان می‌دهد و نشانه احتمال پاسخ حدسی است، یعنی ارائه پاسخی صحیح از سوی آزمون‌شوندگانی که توانایی کمی دارند. این عامل شانسی یا شانسی کاذب^۲ نامیده می‌شود.

نمودار ۸-۱۰ یک منحنی ویژگی سؤال را نمایش می‌دهد. توجه کنید که درصد موفقیت در محور Y و خصیصه یا ویژگی (توانایی) در محور X قرار گرفته است. به سه نکته توجه کنید:



نمودار ۸-۱۰. خطوط کشیده شده برای یک سؤال.



نمودار ۹-۱۰. منحنی ویژگی سؤال برای دو سؤال.

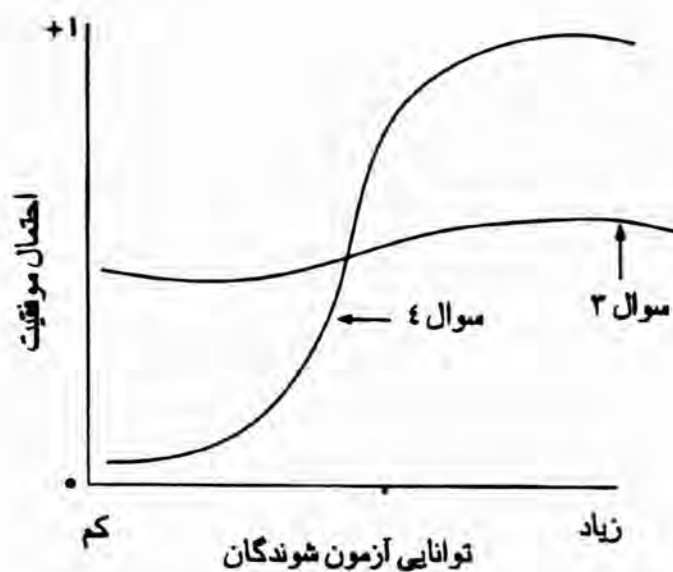
۱- روند شیب تمام منحنی‌ها یکسان است، یعنی همواره افزایش می‌یابد و هرگز کاملاً افقی نمی‌شود.

۲- نقطه انحنای با ترسیم یک خط افقی از یک نقطه (۵۰ درصد کسانی که پاسخ صحیح داده‌اند) بر محور y ها نشان داده می‌شود. با این کار میزان تغییر مکان از چپ به راست منحنی مشخص می‌گردد (عامل B).

۳- خطوط مجانب پایینی و بالایی که به آنها می‌توان نزدیک شد اما هرگز واقعاً به صفر و یا یک نمی‌توان رسید.

از آنجایی که می‌توان شاخصهای تمیز، دشواری و حدس (یا شانس کاذب) را در مورد هر سؤال، با استفاده از نمودارها بدست آورد، منحنی‌های ویژگی سؤال برای تجزیه و تحلیل سؤالات مفید هستند. نمودار ۹-۱۰ منحنی ویژگی را برای دو سؤال، با تشابه در برخی زمینه‌های مورد نظر و با تفاوت در میزان دشواری نشان می‌دهد.

توجه کنید که سؤال ۱ و ۲، شکل مشابهی دارند که نشان‌دهنده ضریب تمیز یکسان این سؤالات است، اما منحنی سؤال ۲ بیش از سؤال ۱ به سمت راست جابجا شده است. بنابراین، سؤال ۲ قادر به تمیز سطوح بالاتر توانایی می‌باشد. می‌توان گفت هر دو سؤال ضریب تمیز مشابهی دارند اما سؤال ۲، سؤال دشوارتری است. این اطلاعات به شما کمک می‌کند تا در مورد استفاده از سؤال ۱ یا ۲ برای گروه ویژه‌ای از آزمون‌شوندگان تصمیم بگیرید.



نمودار ۱۰-۱۰. منحنی ویژگی سوال برای دو سوال متفاوت.

به عنوان مثال: احتمالاً سوال ۱ برای استفاده در دوره فیزیوپاتولوژی با پیشرفت تحصیلی متوسط مناسب است؛ در حالیکه سوال ۲ فقط برای افراد توانمند دوره فیزیوپاتولوژی یا احتمالاً برای امتحان دوره پیش کارورزی قابل خواهد بود. با بررسی نمودار ۱۰-۱۰ مشاهده می‌شود که منحنی ویژگی سوال برای سوال ۳ کاملاً مسطح است که نشان می‌دهد ضریب تمیز، صرف‌نظر از سطح توانایی آزمون‌شوندگان خیلی کم است. معمولاً سؤالاتی که شمایی همانند سوال ۳ دارند، سؤالات ضعیفی هستند و باید مورد تجدیدنظر قرار گیرند. به شیب تند منحنی ویژگی سوال ۲ توجه کنید. ضریب تمیز این سوال فقط در یک نقطه، در ابتدای گستره توانایی بسیار خوب است. در حالت ثابت این ضریب تمیز مناسب است ولی اغلب آزمون‌گیرندگان چنین سؤالاتی را که دارای شیب بیش از حد تند می‌باشند، اصلاح یا حذف می‌کنند. برای بیشتر آزمونها، آزمون‌گیرندگان، سؤالاتی را که دارای منحنی ویژگی سوال ملایم و شکل کند داشته باشند، ترجیح می‌دهند.

تجزیه و تحلیل آزمونهای ملاک مرجع

مبنای تجزیه و تحلیل سؤالات آزمونهای ملاک نسبی، وجود تفاوتها در آزمون‌شوندگان است و ما در آزمونهای ملاک مرجع انتظار نداریم این تفاوتها وجود

داشته باشد و یا کمتر چنین انتظاری داریم و مطلوب آن است که تمام دانشجویان به تمامی اهداف رسیده باشند، بنابراین روشهای مرسوم تجزیه و تحلیل سؤالات از کارآیی چندانی برخوردار نیستند و تمایل بر آن است که با روشهای ویژه‌ای به تجزیه و تحلیل این‌گونه آزمونها پرداخته شود، اما برای رسیدن به روشهای کارا و قابل قبول تحقیقات بیشتری باید صورت پذیرد.

با توجه به آنچه ذکر شد، به دو روش اختصاصی محاسبه ضریب تمیز این آزمونها اشاره می‌شود.

۱- ضریب حساسیت^۱

ضریب حساسیت، شاخصی است برای تعیین تأثیر محتوای آموزشی در نمره آزمون. اگر نمره آزمون قبل و بعد از آموزش یکسان باشد بیان‌کننده آن است که آزمون، تأثیر آموزش را اندازه‌گیری نمی‌کند. یک آزمون در صورتی تأثیر آموزش را اندازه می‌گیرد که نمرات دانشجویان در امتحانات پایان دوره بیشتر از نمرات آنها در آزمون قبل از دوره باشد. برای تعیین ضریب حساسیت از فرمول زیر استفاده کنید:

$$S = \frac{RA - RB}{T}$$

S: ضریب حساسیت

RA: تعداد دانشجویانی که پس از طی دوره به سؤال، پاسخ صحیح داده‌اند.

RB: تعداد دانشجویانی که قبل از شروع دوره به سؤال، پاسخ صحیح داده‌اند.

T: تعداد کل دانشجویان

حداکثر ضریب حساسیت ۱ می‌باشد و ضریب حساسیت کم بیان‌کننده اشکال در آزمون یا نظام آموزشی است.

۲- ضریب تمیز تسلط - فقدان تسلط

این ضریب توسط برنان^۲ پیشنهاد شده، هدف آن تعیین قدرت سؤال در شناسایی افراد دارای حد تسلط مطلوب (مثلاً کسب هشتاد درصد نمره کل) می‌باشد. در این روش از یک مربع چهارخانه استفاده می‌شود. این ضریب با B نشان داده شده، بیان‌کننده تفاوت سطوح دشواری سؤال بین دو گروه با تسلط (مثلاً کسانی که از این نمره کمتر آورده‌اند) می‌باشد.

سؤال	صحیح	a	b
	غلط	c	d
		فقدان تسلط	تسلط

a نشانه تعداد افراد از گروهی است که نمره ملاک را کسب نکرده‌اند ولی به سؤال پاسخ صحیح داده‌اند، b افرادی هستند که با کسب نمره ملاک به سؤال پاسخ صحیح داده‌اند. c تعداد افرادی را نشان می‌دهد که نمره مطلوب کسب نکرده‌اند و به سؤال پاسخ غلط داده‌اند و d نیز نمایان‌کننده تعدادی افرادی است که نمره ملاک را کسب نموده‌اند اما نتوانسته‌اند به سؤال موردنظر پاسخ صحیح دهند. بنابراین $a + c$ مجموعه افرادی است که موفق به کسب نمره مطلوب (مثلاً هشتاد درصد) نشده و $b + d$ تعداد افرادی است که این نمره را کسب نموده‌اند. ضریب تمیز تسلط - فقدان تسلط با استفاده از فرمول زیر محاسبه می‌شود:

$$B = \frac{b}{b + d} - \frac{a}{a + c}$$

این ضریب بین -۱ و +۱ تغییر می‌کند و مطلوب آن است که سؤال دارای ضریبی مثبت باشد و در صورتی که سؤال دارای ضریبی نزدیک به صفر باشد، در ایجاد تمیز بین دو گروه مذکور نقش نداشته است.

ذکر این نکته ضروری است: اگر چه بیشتر بر مثبت بودن ضرایب تمیز تأکید می‌شود سؤالاتی را که دارای ضرایبی تمیز منفی هستند، نمی‌توان به راحتی حذف نمود، به ویژه اگر آن سؤال به هدف یا ملاک مهمی مربوط باشد؛ زیرا ضرایب تمیز منفی ممکن است مربوط به نقص در سؤال، فقدان کارآیی آموزش یا مشکلات یادگیری در بخشی از دانشجویان باشد.

تفسیر داده‌های حاصل از تجزیه و تحلیل سؤالات

به دلایل زیر تفسیر نتایج حاصل از تجزیه و تحلیل عددی سؤالات باید با احتیاط صورت پذیرد.

۱- این داده‌ها معادل برآورد روایی سؤال نیست. برای قضاوت در مورد میزان دقت در روایی آزمون باید از سایر ملاکهای خارجی بهره جست؛ در حالی که به علت ملاک قرار

دادن نمره کلی آزمون به عنوان عامل تقسیم دانشجویان به دو گروه قوی و ضعیف، از ملاک داخلی استفاده شده است. بنابراین، بیشتر هماهنگی درونی سؤالات مورد بررسی قرار می‌گیرد تا روایی.

۲- ضریب تمیز و یا تفاوت عملکرد دو گروه قوی و ضعیف همیشه بیان‌کننده کیفیت سؤال نیست و نباید سؤالاتی که ضریب تمیز کمی دارند به صورت خودکار حذف گردند بلکه سؤالاتی که دارای ضریب تمیز کم یا منفی هستند باید به دقت مورد توجه و بررسی قرار گیرند. در بسیاری موارد سؤالات با ضریب تمیز کم ولی مثبت به ویژه در آزمونهای ملاک مرجع باید حفظ شود زیرا گاهی چنین سؤالاتی در اندازه‌گیری میزان توانایی دانشجویان در بخش مهمی از اهداف آموزش سهم دارد، علاوه بر آن، همانطور که قبلاً ذکر شد گاهی برای ایجاد انگیزه در دانشجویان تعدادی سؤال آسان نیز مورد نیاز می‌باشد. دلایل متعددی برای کم شدن ضریب تمیز سؤال ذکر شده است که از آن جمله می‌توان به این موارد اشاره نمود:

الف- سؤالات بسیار آسان یا بسیار سخت، معمولاً ضریب تمیز کمی دارند. اما در بسیاری از موارد وجود چنین سؤالاتی ضروری است تا بتوان نمونه‌گیری کاملی از محتوا و اهداف آموزشی به عمل آورد.

ب- کارکرد سؤال در ارتباط با کل آزمون، قدرت تمیز آن را تحت تأثیر قرار می‌دهد. به خاطر بیاورید که نمره آزمون یک ملاک داخلی است.

۳- داده‌های حاصل از تجزیه و تحلیل آماری سؤال ناپایدار است. برخی تصور می‌کنند ضریب تمیز و یا مقدار P اعدادی ثابت هستند، در حالی که این تصور چندان واقعیت ندارد. داده‌های حاصل از تجزیه و تحلیل آماری سؤال تحت تأثیر طبیعت و ویژگیهای گروه مورد آزمون، تعداد افراد مورد آزمون، فرایند آموزشی به کار گرفته شده توسط استاد و خطاهای تصادفی قرار می‌گیرد. قضاوت ما در زمینه کیفیت یک سؤال باید بیشتر مبتنی بر پیش‌بینی قدرت سؤال در اندازه‌گیری یک هدف مهم آموزشی باشد تا آن که بر شاخصهای تمیز و یا مقادیر P تکیه کند؛ در حالی که ضریب تمیز مثبت نشان‌دهنده فقدان ابهام سؤال و یا نقایص و عیوب تکمیلی است و چنین سؤالاتی در کاربریهای احتمالی بعدی قابل استفاده‌اند.

۴- برای انتخاب سؤالات آزمون منحصراً از ویژگیهای آماری استفاده نکنید. یکی از بهترین راههای انتخاب سؤالات، انتخاب سؤالاتی است که بر اساس اهداف آزمون دارای قدرت

تفسیر و مقدار P مناسبی باشند؛ زیرا به ویژه در آزمونهای معلم‌ساخته که تعداد دانشجویان حدود سی تا چهل نفر است، نتایج از پایایی چندانی برخوردار نیست. علاوه بر آن که حتی در گروههای به حد کافی بزرگ هم، نتایج آماری به ویژه مقادیر P تحت تأثیر عامل حدس، محل‌گزینۀ صحیح و توالی سؤالات قرار می‌گیرد.

اما به هر حال تجزیه و تحلیل آماری سؤالات به مادر امر انتخاب سؤال کمک می‌کند و بنابراین به عنوان علامت شناسایی سؤالات برای آزمونهای دقیق مناسب است ولی بعد از چنین فرایندی هم سؤالات باید بازبینی شوند تا مشخص شود که واقعاً نماینده محتوا هستند یا خیر.

منابع

- ۱- بلوم؛ و همکاران. *راهنمای ارزشیابی تکوینی و مجموعی، جلد اول*. ترجمه ابراهیم کظیمی. تهران: انتشارات دانشگاه تربیت معلم، ۱۳۵۵.
- ۲- پاشا شریفی، حسن. *اصول روان‌سنجی و روان‌آزمایی*. تهران: انتشارات رشد، ۱۳۷۳.
- ۳- سیف، علی‌اکبر. *اندازه‌گیری و ارزشیابی پیشرفت تحصیلی*. تهران: انتشارات آگاه، ۱۳۷۱.
- ۴- گرون لوند، نورمن. *آزمونهای پیشرفت تحصیلی*. ترجمه خسرو مهندسی. تهران: انتشارات سازمان سنجش، ۱۳۵۶.
- ۵- مایزما، ویلیام؛ و همکاران. *اندازه‌گیری و آزمون در تعلیم و تربیت*. ترجمه غلامرضا خوئی‌نژاد. مشهد: انتشارات مؤسسه چاپ و انتشارات آستان قدس رضوی، ۱۳۷۲.
- ۶- هومن، حیدرعلی. *اندازه‌گیریهای روانی و تربیتی و فن تهیه تست*. تهران: انتشارات فردوسی، ۱۳.
- 7- Mehrens WA. *Measurement and evaluation in education and psychology*. 3rd rev. ed. New York: Holt Saunders, 1984.
- 8- Osterlind SJ. *Constructing test items*. 2nd rev.ed. London: Kluwer Academic Publishers, 1998.
- 9- Peplow PV. *Performance of medical students in case-based and essay components of written anatomy examination*. Med Educ 1991; 25: 293-9.
- 10- Peplow PV. *Students' written reports as an essential part of a case-based learning program in anatomy*. Med Educ. 1992 Nov 11; 26(6): 462-5.
- 11- Stocking ML. *Item response theory*. In: Keevees jp. *Educational research methology and measurmant*. 2nd rev.ed. Cambridge: Cambridge University Press, 1997.

فصل یازدهم

اجرا و نمره‌گذاری

- برگزاری آزمون
- نمره‌گذاری آزمون

برگزاری آزمون

بدون شک هر یک از ما در امتحانات متعددی شرکت کرده، کم و بیش با نحوه اجرای آزمون آشنایی داریم. اما عوامل اجرایی آزمون تا چه حد در نتیجه آزمون مؤثر است و به افزایش نمره دانشجوی کمک می‌کند. آنچه در پی می‌آید پاسخ به این سؤال و مطالبی در مورد تقلب، اضطراب و نحوه اجرای آزمون و نمره‌دهی است.

ابتدا به نکات کلی در مورد چگونگی اجرای آزمون اشاره می‌نماییم.

توزیع سؤالات امتحانی

قبل از توزیع سؤالات باید مطمئن شد که تمام کتابها و جزوه‌های مربوط به امتحان از دانشجویان دور یا از دسترس آنها خارج است. سپس باید پاکت مهر شده حاوی سؤالات را به همه نشان داد. البته باید برگه سؤالات را از پشت، روی میز قرار دهید به طوری که امکان خواندن سؤالات وجود نداشته نباشد و همه در یک زمان مشخص، پاسخ به سؤالات را شروع کنند و کسی که اول سؤالات را می‌گیرد با کسی که آخر، آن را دریافت می‌کند وقت یکسانی داشته باشد. همچنین موظفید که وقت امتحان را اعلام کنید. (می‌توانید روی تخته سیاه یا بالای برگه سؤالات بنویسید.)

مدیریت

مدیریت و صداقت در آن، اهمیت بسزایی در اجرای آزمون دارد. به عنوان مدرس، باید در محل امتحان از شروع تا پایان حاضر باشید؛ به نحوی که دائماً دانشجویان حضور شما را احساس کنند. به علاوه به هیچوجه نباید خود را پشت میز مشغول نگه داشته یا به مطالعه کتابهای داستانی بپردازید؛ بلکه باید کاملاً توجه خود را به آزمون‌دهندگان معطوف دارید. نباید از دانشجویانی که اوقات فراغت دارند برای مسؤولیت جلسه استفاده کرد. قبل از اجرای آزمون نظارتی بر برگه‌های حاوی سؤالات داشته، از باز نشدن مهر آن مطمئن شوید. بهتر است برای امتحانات از سؤالات ماشین‌نویسی شده استفاده کنید.

عوامل اجرایی آزمون تأثیر مستقیمی در نمرات آزمون خواهند داشت که این تأثیر تا حدود زیادی در کنترل استاد می باشد.

در اجرای آزمون به این نکات زیر توجه کنید:

- ۱- به دانشجو وقت کافی بدهید، تا خود را برای امتحان آماده کند.
- ۲- هرگز آزمون را بدون اطلاع قبلی دانشجو برگزار نکنید؛ زیرا تواناییها را محدود کرده، باعث ایجاد اضطراب در دانشجو می شوید.
- ۳- به دانشجو اطلاعات لازم را در مورد زمان برگزاری، تعداد سؤالات و نوع آنها، میزان تأثیر آزمون در نتیجه کلی دوره درسی، حدود موضوعاتی که در امتحان مطرح می شود و ... بدهید؛ زیرا اولاً مطالعه او را هدایت می کند، ثانیاً نگرانی دانشجو را کاهش داده، آمادگی وی را از نظر روانی افزایش می دهد.
- ۴- فضای مناسبی را برای امتحان انتخاب کنید. هر دانشجو باید فضای کافی برای قرار دادن وسایل خود داشته و فاصله دانشجویان از هم مناسب باشد. برای جلوگیری از تقلب یا حرکات غیرقانونی امتحانی، برای نشستن دانشجویان روی صندلی باید مواردی را در نظر بگیرید، یعنی صندلیها نزدیک هم نباشد؛ دوستان کنار هم ننشینند؛ برای اطمینان بیشتر، جای نشستن هر فرد در هر امتحان باید فرق کند و همواره ثابت نباشد و دانشجویان برای هر امتحان باید جابجا شوند.
- ۵- زمان مناسبی را برای امتحان انتخاب کنید. هیجان و خستگی مانع از اجرای خوب آزمون می شود. زمان آزمون را طوری تنظیم کنید که قبل از ناهار یا قبل از پایان ساعت اداری به اتمام برسد.
- ۶- اطلاعات کافی را در مورد نحوه اجرا و سازماندهی آن (از جمله توزیع سؤالات و پاسخنامه) به اجرا کنندگان آزمون بدهید تا از لحاظ زمان شروع، اتمام و ... همسانی وجود داشته، امکان مقایسه دانشجویان را با یکدیگر فراهم آورد.
- ۷- به هر یک از دانشجویان، یک برگه به عنوان پاسخنامه بدهید.
- ۸- در هنگام اجرای آزمون اگر دانشجو سؤالی دارد باید به آن پاسخ دهید، به گونه ای که دیگر دانشجویان تشویق به پرسیدن نشوند؛ زیرا با زیاد شدن تعداد سؤال کنندگان نظم جلسه به هم می خورد و این کار باعث حواس پرتی می شود.
- ۹- زمان باقیمانده را با فواصل معین روی تخته سیاه بنویسید.

- ۱۰- بعد از امتحان به دانشجویان اجازه دهید تا سؤالات و پاسخهای پیشنهادی^۱ را ببینند.
- ۱۱- قبل از تصحیح برگه‌ها یک جلسه برگزار کنید تا دانشجویان پیشنهادهای سازنده خود را برای تغییر در تصحیح ارائه دهند. این پیشنهادها را در زمان کافی بررسی کرده، انواع قابل قبول را در فرآیند نمره‌گذاری از پیش تعیین شده خود دخالت دهید و طبق دستورالعمل اصلاح شده برای تصحیح اقدام کنید.
- ۱۲- بعد از تصحیح و اعلام نتایج اجازه دهید که دانشجویان برگه‌های خود را ببینند. (کاری که معمولاً در دانشکده‌ها به علت تعداد زیاد دانشجو انجام نمی‌شود؛ اما این امر برای ارائه بازخورد به دانشجو کاملاً ضروری است.)

تعیین مدت زمان آزمون

از جمله تصمیمات مهم در طراحی آزمون، تعیین تعداد سؤالات و طول زمان آزمون است.

تعداد سؤالات با توجه به محتوای درسی تعیین شده، طول زمان آزمون نیز براساس نوع سؤالات، میزان نیاز سؤالات به دقت و توجه دانشجو و میزان خستگی دانشجویان مشخص می‌شود.

با توجه به اینکه استادان بیشتر مایلند میزان توانایی و ادراک کلی دانشجو را بسنجند و نه سرعت پاسخ‌دهی او را، اغلب آزمونهای پزشکی از نوع آزمون قدرت است. بنابراین، در این آزمونها زمان را باید طوری تعیین کنید که بیش از ۹۰ درصد دانشجویان وقت کافی برای پاسخ‌دهی به سؤالات داشته باشند. این زمان با توجه به تجربه آزمون‌گیرنده تعیین می‌شود. پیشنهاد می‌گردد که در آزمونهای قدرت برای هر سؤال چندگزینه‌ای یک دقیقه در نظر بگیرید، سپس طبق نظر خود مدت زمان خاصی را به آن اضافه کنید. مثلاً: اگر آزمون ۶۰ سؤال دارد احتمالاً زمان کافی یک ساعت و ده دقیقه خواهد بود.

جمع‌آوری برگه‌ها

همان‌طور که تأکید شد همانند زمان شروع امتحان در موقع جمع‌آوری برگه‌ها نیز باید نظم را رعایت کنید تا از هرج و مرج جلوگیری شود. دانشجویی که زودتر از زمان معین،

آزمون را تمام می‌نماید باید دستش را بلند کند و یکی از مسؤولین برگه او را گرفته وی را به آرامی از جلسه خارج کند تا مزاحمتی برای دیگران ایجاد ننماید. کسانی که تا پایان وقت در جلسه حضور دارند باید سر جای خود بمانند تا مسؤولان برگه‌ها را به صورت منظم جمع‌آوری کنند. معمولاً این شیوه، از برخوردهایی که ممکن است در هر جلسه آزمون پیش آید جلوگیری می‌کند. همچنین شما می‌توانید شخصاً برگه را از جلوی امتحان دهنده برداشته از گم شدن آن پیشگیری نمایید. علاوه بر آن از شلوغی و ایجاد شرایط برای تقلب کردن افراد فرصت طلب جلوگیری می‌شود.

تقلب

تقلب پدیده‌ای است که هیچگاه نمی‌توان آن را به طور کامل از بین برد اما با رعایت این نکات می‌توان آن را به حداقل رساند:

- ۱- سؤالات آزمون را به طور پنهانی تهیه و تکثیر کنید؛
- ۲- آزمون را در محیط کاملاً آرام و همراه با نظم اجرا کنید؛
- ۳- فاصله دانشجویان را به گونه‌ای تعیین کنید که امکان رونویسی به حداقل برسد؛
- ۴- دو نمونه از سؤالات یک آزمون را با ترتیب مختلف به کار ببرید؛
- ۵- آزمون را به گونه‌ای ترتیب دهید که از نظر کیفی بی‌محتوا و خلاف عدالت تصور نشود؛
- ۶- اگر روایی محتوایی آزمون به درستی مورد توجه قرار نگیرد احتمال تقلب زیاد می‌شود؛
- ۷- امکان تقلب در سؤالات تشریحی کمتر از سؤالات چند گزینه‌ای است اما نوع سؤال بیشتر چگونگی و نحوه تقلب را تحت تأثیر قرار می‌دهد تا میزان آن را.

برای نظارت بر جنبه‌های مختلف تقلب، در دانشکده‌های پزشکی، بخش دانشجویی انجمن پزشکی آمریکا^۱ در سال ۱۹۸۹، کمیته ارتباطات تحصیلات پزشکی^۲ را تشویق به ایجاد یک نظام ابداعی کرد که بر مبنای آن به دانشجویان گواهی اخلاق اعطا می‌گردید. این نظام بر سه اصل استوار بود:

- ۱- گواهی اخلاق، باید بتواند رفتارهای غیر اخلاقی، از جمله تقلب را کاهش دهد؛
- ۲- اعطای گواهی اخلاق به فرد بر پایه صداقت و درستی باشد؛
- ۳- این گواهی باید بتواند در درازمدت رفتارهای درست و اخلاقی را در پزشکی تقویت نماید.

در واقع گواهی اخلاق مجموعه امتیازاتی بود که دانشجو در طول دوران تحصیل خود بر مبنای رفتارهای اخلاقی و صادقانه خود بدست آورده بود و توسط مسئولان دانشکده به او اعطا می‌شد. گرچه هر کس دوست دارد مورد اطلاق صداقت و درستی قرار گیرد، این جملات معمولاً در حد فرضیه بوده، وقایع جاری آنها را تأیید نمی‌کند.

بروک^۱ و همکارانش در مطالعه‌ای که بر روی گروهی از دانشجویان پزشکی دارای گواهی اخلاق انجام دادند به این نتیجه رسیدند که فقط چهار نفر از هفتاد نفری که متعهد شده بودند در صورت مشاهده تقلب همکلاسهایشان به دانشکده گزارش دهند حاضر به گزارش شدند و فقط دو مورد گزارش شده توسط یک کمیته مسئول مورد بررسی قرار گرفت. پس مسئولان دانشکده هم برای اعمال اخلاقی که خود، تعیین کرده بودند چندان ارزشی قائل نبودند. (در این دانشکده یکی از اعمال اخلاقی که برای دانشجویان امتیاز مثبت به دنبال داشت گزارش موارد تقلب بود.)

میسیریال^۲ نشان داد میزان تقلب در آزمونهایی که توسط گواهی اخلاق کنترل می‌شود در مقایسه با آزمونهای اجرا شده تحت نظارت، خیلی بالاست. اگر چه اکثریت دانشجویانی که میسیریال گزارش داد اعطای گواهی اخلاق را آخرین راه برای جلوگیری از رفتارهای غیر اخلاقی می‌دانستند، هیچکدام وقوع تقلب را به مراجع دانشجویی و مسئولان دانشکده گزارش نکردند.

تصور اینکه اعطای گواهی اخلاق به تنهایی باعث راستی و صداقت دانشجویان می‌شود بیش از حد آرمان‌گرایانه است. استفاده از گواهی اخلاق به معنی انتظار شکل‌گیری رفتار اخلاقی در دانشجو به محض ورود به دانشگاه است؛ در حالی که رفتار اخلاقی هر شخص از بدو دستیابی به قوه تمیز خوب و بد، شکل می‌گیرد نه با سپردن تعهد صادقانه یا ادای یک جمله در دانشگاه. صداقت و درستی در پزشکی در طول تحصیل کسب می‌شود و با عملکرد اخلاقی فرد در خلال حرفه خود تثبیت خواهد شد. پس اگر اعطای این گواهی به دانشجویان فایده‌ای ندارد چه باید کرد؟

شاید بهتر باشد پدیده تقلب را به عنوان یک معضل جدا در نظر نگرفت و آن را در زمره مسائل اخلاقی پزشکی قرار داد. رفتار اخلاقی در پزشکی به معنی انجام ندادن کلیه فعالیت‌هایی است که مراقبت از بیمار را خدشه دار می‌کند. یادگیری رفتار اخلاقی، یادگیری پذیرش مسئولیت توسط یک پزشک است.

همه دانشکده‌های پزشکی علاقه‌مند به آموزش علم پزشکی هستند ولی بخشی از این زمان باید صرف ایجاد حس مسئولیت در دانشجویان شود تا پزشکانی با اخلاق تربیت گردند. آموزش مسائل اخلاق پزشکی باید از روز اول ورود به دانشکده آغاز گردیده، در تمام طول تحصیل ادامه یابد. مطالبی که باید در این تعلیم اخلاقی در نظر گرفته شود عبارت است از:

- ۱- چگونگی رفتار با برخوردهای مبارزه‌طلبانه بیمار یعنی وقتی بیمار قصد همکاری با پزشک را ندارد و برخورد پرخاشگرانه‌ای نشان می‌دهد.
- ۲- چگونگی تلفیق جنبه‌های مالی و تجاری در پزشکی با جنبه‌های اخلاقی به طوری که در عین برطرف‌سازی معقول نیازهای مالی، مسائل اخلاقی نیز پایمال نشود.
- ۳- شناسایی تقلب در هر چهره‌ای که باشد خواه تقلب در امتحانات دانشجویی باشد یا تقلب در حرفه پزشکی به شکل ارائه نکردن خدمات بهینه به بیمار. چون ابتدا فرد باید بپذیرد که تقلب سر جلسه امتحان بخصوص در دانشکده پزشکی بسیار ناپسند است تا پس از آن به تقلب نکردن در حرفه خود عادت کند. چنین تغییری در برنامه آموزشی احتیاج به یک انقلاب اساسی در آموزش اخلاقی دانشجویان دارد تا پزشکانی تربیت شوند که در عین اینکه در امتحانات دانشگاهی تقلب نمی‌کنند در شغل آینده خود نیز مسؤول و صادق باشند.

البته رفتارهای غیر اخلاقی را نمی‌توان به طور کامل از بین برد اما تغییر شیوه و روش تفکر دانشجویان راجع به رفتارهای خود می‌تواند یک قدم به سوی بهبود صداقت در آنان باشد.

اضطراب آزمون^۱

آزمون باعث عکس‌العملهای مختلف در میان افراد می‌شود. در عده‌ای حس رقابت را تقویت کرده، اشتیاق و کارآیی فرد را افزایش می‌دهد. در گروه دیگر اضطراب مختصری ایجاد کرده باعث افزایش کارآیی آنها می‌شود. مثلاً: ورزشکاران در هنگام عصبانیت و هیجان در مسابقات بزرگ، قدرت بدنی بیشتری از خود بروز می‌دهند. این اضطراب را تسهیل‌کننده^۲ (مثبت) می‌گویند. در عده‌ای دیگر اضطراب به قدری شدید است که قدرت عمل

آنها را کم می‌کند. این اضطراب را مهارکننده^۱ (منفی) می‌گویند. گاهی اضطراب علاوه بر مشکلات روانی باعث ایجاد مشکلات جسمی مثل تهوع و موارد مشابه می‌شود که این حالت زیاد شایع نیست. از جمله در تحقیقی در سال ۱۹۹۳، چهل نفر دانشجوی پسر و ده نفر دانشجوی دختر انتخاب شدند که برای اولین بار در آزمون NBME شرکت می‌کردند. از شش ماه تا ده دقیقه قبل از امتحان؛ سرعت نبض، فشار خون و الکتروکاردیوگرام (ECG) آنها را بررسی کردند. سرعت نبض و فشار خون، افزایش قابل ملاحظه‌ای نسبت به گروه کنترل داشت اما در ECG تغییر مهمی دیده نشد.

اغلب دانشجویان تمایل دارند موفق نشدن خود را در امتحان به وجود اضطراب نسبت دهند و نه نداشتن آمادگی لازم برای امتحان، اما یک همبستگی منفی متوسط، بین میزان اضطراب و عملکرد موردنظر برای فرد در آزمون (نمره پیش‌بینی شده) وجود دارد (از ۰/۳۵- تا ۰/۲۱-). این ضریب نشان‌دهنده تضعیف کارایی دانشجو در اثر اضطراب است.

اینکه فقدان آمادگی برای آزمون باعث اضطراب در امتحان می‌شود، یک امر کاملاً شناخته شده و قابل قبول است. اضطراب امتحان به میزان اطلاعات و آشنایی با موقعیت آزمون بستگی دارد. چون واکنش افراد در مقابل آزمون متفاوت است نتیجه می‌گیریم که اضطراب ایجاد شده در افراد مختلف به میزان متفاوت خواهد بود. برخی دانشجویان حتی با داشتن آمادگی کافی به صرف شرکت در امتحان دچار اضطراب می‌شوند. در اشخاصی که آمادگی شرکت در آزمون ندارند، میزان اضطراب زیاد می‌شود. اضطراب زیاد شده کارایی را کاهش می‌دهد و این چرخه معیوب ادامه یافته، آثار نامطلوبی ایجاد می‌کند. به عنوان مثال در سال ۱۹۷۳، در ژاپن ۳۴۱ نفر به علت ترس از شکست در امتحانات ورودی دانشگاه دست به خودکشی زدند. هرگاه آزمون تعیین‌کننده آینده دانشجو باشد، اضطراب، تنها پاسخ منطقی در برابر آزمون بوده، در این موارد بهتر است آزمون را در چند نوبت اجرا کرد.

برای کاهش اضطراب دانشجو از امتحان باید در حین تدریس هم اقداماتی انجام دهید. می‌توانید ارتباط صمیمی و دوستانه با دانشجویان خصوصاً آنها که ترس از شکست در امتحان دارند برقرار کرده، سعی نمایید امتحانات را به تعداد بیشتر و در زمان کوتاه‌تر برگزار کنید. اعتماد نداشتن به مهارتهای استاد در عمل ارزیابی، باعث ایجاد اضطراب منفی شده، شخص پایین‌تر از حد دانش و تواناییهای خود شناخته می‌شود.

همچنین باید از میزان یأس و ناامیدی دانشجو که ناشی از شکستهای متوالی او بوده

بکاهید. به عنوان مثال برای ایجاد آمادگی در کسانی که از نظر درسی ضعیف هستند، آزمونهایی را در طول دوره طراحی کنید. احساس موفقیت باعث ایجاد انگیزه در دانشجو شده، اضطراب او را کاهش می‌دهد. اگر دانشجو احساس کند که می‌تواند از عهده آزمون برآید نگرانی او کم می‌شود و اگر حس کند امتحان حربه‌ای است که استاد، به کمک آن، دانشجویان را به صف می‌کند، اطمینان او به استاد از دست می‌رود. اگر هدف از امتحان گرفتار کردن دانشجو باشد مسلماً ترس و اضطراب او افزایش می‌یابد. سعی کنید حین تدریس، این‌گونه به دانشجویان قوت قلب ببخشید: «من می‌خواهم شما خوب یاد بگیرید چون درک کرده‌ام که شما خوب فهمیده‌اید. هدف من از گرفتن امتحان، ایجاد فرصتی برای درک بهتر است و ...».

با توجه به اینکه اطمینان نداشتن باعث اضطراب می‌شود، امتحان گرفتن بدون اطلاع قبلی، امتحانات غیر قابل پیش‌بینی که دانشجو نمی‌داند چه مهارتی را از او خواهند سنجید و یا وجود استنادانی که بی‌قاعده و یا برعکس با وسواس نمره می‌دهند ترس و نگرانی را افزایش می‌دهد. برعکس، اگر استاد، دانشجو را برای امتحان کاملاً آماده کرده، دانشجو اهداف امتحان را به خوبی بداند و از نمره‌گذاری براساس اصولی منطقی مطمئن باشد اضطراب آزمون کم می‌شود.

گاهی دانشجو به خوبی برای شرکت در آزمون آماده است ولی فقدان مهارت استاد در برگزاری آزمون و نمره‌گذاری، باعث اضطراب وی می‌شود؛ به همین دلیل باید مهارت استاد را در زمینه‌های مختلف اندازه‌گیری کرد. در این راستا می‌توان با نظرخواهی از دانشجویان در مورد استاد و ارائه بازخورد به وی، موجب تقویت این مهارتها گردید.

نمره‌گذاری

نمره نشانه‌ای از پیشرفتهای یک دانشجو است. نمره را یا به صورت حروف (الف، ب، ج، د) یا اعداد (۰-۱۰-۲۰) و یا کلمات (خوب، متوسط، ضعیف) نشان می‌دهند. بهتر است نمره به صورت عدد بیان شود تا عبارت؛ زیرا کلمات، معمولاً مبهم هستند. همچنین در مورد کلمات نمی‌توان جمع و تفریق و ... انجام داد. نمره‌گذاری به صورت صفر تا صد مناسب نیست، زیرا با افزایش تعداد طبقات نمره‌گذاری، قدرت تمیز و تفاوت قائل شدن آن کم می‌شود. علاوه بر آن اگر تعداد طبقات کم باشد اعتبار نمره کاهش می‌یابد. بنابراین بهترین روش نمره‌گذاری، به شیوه نه یا ده طبقه‌ای است.

برای اینکه یک نمره‌گذاری، بهترین باشد لزوماً نباید از نمودار زنگوله‌ای (توزیع نرمال) تبعیت کند. در اکثر موارد نمرات هم نتیجه اندازه‌گیری و هم، نتیجه قضاوت استاد است. در صورتی که قضاوت یک امر ذهنی بوده، معمولاً صراحت ندارد. قضاوت شخصی استاد اغلب پایایی و روایی آزمون را کاهش می‌دهد. بنابراین، تا حد ممکن نمره‌گذاری را براساس نتایج اندازه‌گیری تبیین کنید نه قضاوت شخصی.

روش نمره دادن وابسته به نوع پاسخهاست. برای امتحانهایی که پاسخ آن به صورت واضح و روشن است، امتحان دهنده، دور پاسخ درست را خط کشیده یا سیاه می‌کند یا پاسخها را در ورقه‌ای به نام «ورقه جوابها» می‌نویسد، تصحیح و نمره دادن ساده است. برای این سؤالاها یک نفر می‌تواند جواب را بخواند و دیگری ورقه جواب را کنترل کند. در مورد سؤالات چند گزینه‌ای می‌توان محل جواب درست را روی ورقه سوراخ کرده مانند کاغذ کپی روی ورقه امتحانی قرار دهید و سؤالات درست را کنترل کنید. قبل از جمع کردن پاسخهای درست باید دقت کنید که برای هر سؤال فقط یک پاسخ داده شده باشد. وقتی که پاسخ سؤالاها جمله‌های کوتاه یا مثال باشد باید جمله‌ها هدف سؤال کننده را شامل شود. در این نوع نمره دادن باید تصمیم بگیرید آیا برای غلطهای املایی^۱ و نگارش نیز نمره کم می‌کنید یا فقط دنبال مفهوم پاسخ می‌گردید.

این تصمیم‌گیری به هدف امتحان بستگی دارد. اگر منظور، سنجش قدرت دانشجو در نگارش است باید نمره کسر شود و گرنه این کار ضرورتی ندارد. دقت کنید صحت نمره دادن درست و مورد اطمینان باشد. برای این کار باید نمونه دقیق جوابهای امتحانی را تهیه کرده، تمام نکات مهم مورد نظر را در آن قید کنید و برای هر نکته یک نمره معین در نظر بگیرید. مثلاً: نمره هر نکته مهم سؤال را در جلو آن بنویسید.

مثال: پنج نوع آزمون را نام برده، مختصراً درباره هر کدام توضیح دهید. (۲۵ نمره)

نمره تعیین شده در پراگماتیک به معنی نمره کامل برای سؤال است. نمره‌دهنده بعداً بخشهای جواب را نسبت به اهمیت موضوعی که باید توضیح داده شود نمره‌دهی می‌کند:

الف) پنج نمره برای فهرست کردن هر پنج نوع امتحان، یک نمره برای هر نوع؛

ب) بیست نمره برای توضیح مختصر هر نوع آزمون؛ بنابراین چهار نمره برای هر توضیح کامل درست. نمره دهنده ممکن است این چهار نمره را بین هر قسمت از جواب، توزیع کند. مثلاً علاقه‌مند باشد نمره معینی برای تعریف، موارد استفاده، مزایا و معایب در نظر بگیرد.

نمره‌دهنده باید در ورقه جوابها این نکات را به خوبی پیش‌بینی کرده، نمره‌های معینی برای هر قسمت در نظر بگیرد. این مثال تا حدودی نحوه نمره‌گذاری سؤالات توضیحی را بیان می‌کند. فوایدی که نمره دادن به دانشجو در پی خواهد داشت عبارت است از:

- ۱- اطلاعاتی را در مورد پیشرفتهای یک دانشجو در اختیار دست‌اندرکاران آموزش گذاشته، بررسی وضع تحصیلی و میزان پیشرفت و ترقی او را امکان‌پذیر می‌سازد؛
- ۲- دانشجو را در تعیین مسیر تحصیلی یا شغلی خود یاری می‌دهد؛
- ۳- عامل محرک مهمی برای افزایش تلاش و کوشش دانشجو به حساب می‌آید.

در برخی از دانشگاهها به جای استفاده از نظام نمره‌دهی عددی از نظام قبولی-ردی^۱ استفاده می‌کنند. در این نوع به جای تعیین نمره برای دانشجو تنها قبولی-ردی او مشخص می‌شود؛ مثلاً در دانشگاه میشیگان دانشجویان را به چهار دسته (عالی، بسیارخوب، خوب، مردود) طبقه‌بندی می‌نمایند. این روش در بسیاری از مراکز آموزشی برای کاهش اضطراب دانشجو به کار می‌رود. همچنین محیطی ایجاد می‌کند که در آن باید دانشجو صرفاً برای یادگرفتن درس بخواند. چون وقتی توجه دانشجویان از نمرات و رقابت با یکدیگر کم شود، توجه به مقوله یادگیری بیشتر می‌گردد.

چگونگی نمره‌گذاری و تصحیح آزمون

برای آسانتر شدن تصحیح آزمون، بهتر است در مورد نحوه پاسخگویی دستورالعمل خاصی در اختیار دانشجویان قرار دهید. مثلاً: همه پاسخها را در سمت راست صفحه بنویسند.

گاهی استادان به بعضی سؤالات به علت اهمیت آنها نمرات بیشتری می‌دهند. این کار در افزایش ثبات و اعتبار آزمون نقشی ندارد؛ فقط کار تجزیه و تحلیل و تصحیح آزمون را مشکل می‌کند.

در سؤالات کوتاه پاسخ و سؤالات تشریحی تغییر یافته که طرح پرسشی در آن مبتنی بر بررسی مشکل بیمار و اداره وی است، تصحیح، برای موضوعاتی که در حد یادآوری باشد آسانتر است؛ زیرا پاسخهای صحیح به صورت کلمات کلیدی، عبارات یا جملات واضح خواهد بود. اما تصحیح موضوعاتی که در حد درک و کاربرد باشد نیاز به

الگوی از پیش تعیین شده برای پاسخها^۱ داشته، تصحیح آنها بیشتر جنبه ذهنی دارد تا عینی. نقش الگوی نمره‌گذاری بر پایه بخشهای ادراکی پنج دقیقه‌ای به ازای هر یک نمره است. به این معنی که زمان را به بخشهای پنج دقیقه‌ای تقسیم نمود، به تعداد آن بخشها مفاهیم پاسخ را دسته‌بندی کرده، برای هر مفهوم یک نمره در نظر گرفته می‌شود. مثلاً: پاسخ یک سؤال تشریحی طولانی با وقت ۲۵ دقیقه باید دارای پنج قسمت پنج دقیقه‌ای باشد و هر کدام یک نمره بگیرد.

نمره منفی

سه عامل سبب می‌شود دانشجو پاسخ را حدس بزند:

۱- کل مسأله ارائه شده را نداند؛

۲- برای پاسخگویی وقت، کافی نباشد؛

۳- هیچ‌انگیزه‌ای برای موفقیت در آزمون نداشته باشد.

هر چه اثر این عوامل کمتر شود اعتبار آزمون بالاتر می‌رود. عامل حدس در پاسخ به سؤالهای چند گزینه‌ای موقعی جدی است که آزمون موردنظر از نوع آزمون سرعت یا دارای سؤالهای دو گزینه‌ای بوده یا تعداد سؤالها کم و زمان آزمون کوتاه و یا سؤالات مشکل باشد. در این موارد می‌توانید نمره منفی منظور کرده، در جلسه امتحان حتماً این مسأله را به دانشجو متذکر شوید؛ اگر چه بهتر است به جای این کار به رفع اشکالات مذکور اقدام کنید.

در آزمونهای گزیده پاسخ که دارای چهار یا پنج گزینه باشد یا آزمونهای قدرت یا آنهایی که تعداد سؤالات آن زیاد و در سطح معلومات دانشجویان است، عامل حدس و گمان چندان مؤثر نیست و لزومی ندارد نمره منفی منظور کنید. در این صورت در جلسه امتحان متذکر شوید که دانشجویان همه سؤالها را، حتی در صورتی که احساس می‌کنند کوچکترین اطلاعی در مورد آن ندارند، جواب دهند.

نمره منفی در سؤالات عینی

دادن نمره منفی به سؤالات عینی مثل کم کردن $\frac{۱}{۳}$ پاسخهای غلط از پاسخهای صحیح یک روش کم‌فایده است؛ زیرا این روش معمولاً به ضرر دانشجویی است که درصد اطمینان

او به درستی پاسخ، بیشتر از دیگران است. هتلاً: کسی که بین چهار گزینه نمی‌تواند هیچکدام را به عنوان پاسخ غلط شناسایی کند و پاسخ درست را بیابد و کسی که دو گزینه غلط را تشخیص داده، بین دو گزینه باقیمانده، پاسخ صحیح سؤال را پیدا نمی‌کند نمرهٔ یکسان می‌گیرند؛ در صورتی که از نظر سطح معلومات با هم تفاوت دارند.

معمولاً برای محاسبهٔ نمرهٔ منفی در امتحانات عینی از رابطهٔ زیر استفاده می‌نمایند:

$$S = R - \frac{W}{K-1}$$

S = نمره ورقه؛

R = تعداد پاسخهای درست؛

W = تعداد پاسخهای غلط؛

K = تعداد گزینه‌ها.

هتلاً در یک آزمون چهار گزینه‌ای نمرهٔ فردی که ۷۰ پاسخ درست و ۳۰ پاسخ غلط دارد، ۶۰ از ۱۰۰ خواهد بود.

$$70 - \frac{30}{4-1} = 60$$

طبق این رابطه در آزمون چهار گزینه‌ای به ازای هر سه سؤال غلط یک نمرهٔ منفی تعلق می‌گیرد؛ زیرا امکان انتخاب گزینه‌های انحرافی سه برابر امکان انتخاب گزینهٔ صحیح است. دانشجویی که هیچ اطلاعی از درس ندارد باید تعداد پاسخهای درست او برابر با $\frac{1}{4}$ پاسخهای غلط باشد.

نمرهٔ منفی در امتحانات چند گزینه‌ای نسبت به تعداد سؤالات، $\frac{1}{4}$ یا $\frac{1}{8}$ در نظر گرفته می‌شود. البته اگر امتحان، گزینهٔ «نمی‌دانم» هم دارد برای انتخاب این گزینه نباید نمرهٔ منفی قائل شد.

در یک سؤال گزیده پاسخ اگر تعداد گزینه‌ها را K فرض کنیم، احتمال حدس پاسخ صحیح $\frac{1}{K}$ خواهد بود. هتلاً: در یک آزمون چهار گزینه‌ای با ۱۰۰ سؤال اگر فردی همه را حدسی جواب دهد ($\frac{1}{4} \times 100 = 25$) انتظار نمرهٔ ۲۵ یا حدود آن را خواهیم داشت. این رابطه در صورتی صحیح است که:

۱- تمامی آنهایی که پاسخ صحیح را نمی‌دانند حدس زده، درست و غلط بودن حدسشان نیز تابع قانون احتمالات باشد. از طرفی بعضی از دانشجویان بی‌اطلاع ولی محافظه‌کار، از ترس نمرهٔ منفی از پاسخ دادن خودداری می‌کنند؛ و برخی از دانشجویان بااطلاع‌تر

بعضی از گزینه‌ها را به راحتی حذف کرده، بین دو یا سه گزینه به انتخاب می‌پردازند؛ در نتیجه در بیشتر اوقات قانون احتمالات وجود ندارد.

۲- در تمامی پرسشها، همه گزینه‌ها توجه دانشجویان بی‌اطلاع را به یک اندازه جلب کند اما گاهی اطلاعات دانشجو به اندازه‌ای است که از انتخاب بعضی گزینه‌های انحرافی خودداری می‌کند؛ در عین حال فاقد آگاهی کافی برای انتخاب گزینه درست است.

بنابراین با توجه به آن که با برقراری نمره منفی اضطراب دانشجویان افزایش می‌یابد و از طرفی دانشجویانی که از نظر خصوصیات فردی محافظه‌کارترند، تمایل کمتری برای پاسخدهی به سؤالات نشان می‌دهند و نمی‌توانند تمام دانسته‌های خود را بروز دهند؛ همچنین اطمینان از وجود چنین وضعی مشکل است؛ نظر کلی بر آن است که اگر دانشجو پاسخ پرسش را با اطمینان نمی‌داند باید بتواند بهترین پاسخ را حدس بزند.

ذکر این نکته لازم است که دانشجویان بیشتر بر این باورند که هنگام پاسخ دادن به سؤالات چند گزینه‌ای باید بر نخستین گمان تکیه کرد. ولی پژوهشها نشان داده است که در خلال اجرای آزمون، بیشتر پاسخها از غلط به صحیح تغییر می‌کند نه از صحیح به غلط. یعنی بیشتر دانشجویانی که پاسخهایشان را تغییر داده‌اند به نمراتشان افزوده شده است.

اثر شانس بر میزان موفقیت

شانس در گذراندن هر امتحانی نقش دارد، این موضوع را تقریباً همه قبول دارند. بنابراین برای اطمینان از قبولی در امتحان، مهارت و شانس هر دو لازم است. فرض کنید فردی که کاملاً برای امتحان آمادگی دارد به علت یک بیماری ساده یا سایر حوادث غیر منتظره از شرکت در امتحان بازمانده یا نمی‌تواند به خوبی به سؤالات پاسخ دهد و با وجود مهارت و آگاهی به علت بدشانسی مردود می‌شود. در حالی که از فرد دیگری با اطلاعات کم و محدود ممکن است تنها سؤالاتی پرسیده شود که وی در مورد آنها آگاهی داشته است؛ بنابراین با وجود اطلاعات کم در امتحان به علت خوش‌شانسی قبول می‌شود.

اما نکته اینجاست که در موفقیت شخص در امتحان، شانس تا چه اندازه اهمیت دارد؟ آیا شانس و معلومات در قبولی فرد نقش یکسان دارند یا بخت و اقبال فقط تا حدودی بر آن مؤثر است؟

برای پاسخگویی به این سؤال، در سال ۱۹۹۰ تحقیقی براساس گزارش سالیانه دانشکده زبان یکی از دانشکده‌های دانشگاههای علوم پزشکی انجام شد. این گزارش در

مورد نتایج امتحانات پس از فراغت از تحصیل دانشجویان و چگونگی موفقیت آنان در این آزمونها بود. بسیاری از افراد در اولین بار شرکت در امتحان پس از فراغت از تحصیل مردود شدند ولی بافاصله زمانی کمی دوباره در آزمون شرکت کرده، در دفعات دوم یا سوم به موفقیت دست یافتند. سؤالی که به ذهن می‌رسید و اساس کار این تحقیق می‌باشد این بود که آیا مهارت افرادی که در دفعات دوم یا سوم قبول شدند نسبت به دفعه اول افزایش یافته یا آنها در این امتحان خوش شانس‌تر بوده‌اند؟

البته در صورتی که از خود افراد پرسیده شود، این میل باطنی وجود دارد (هم در مورد آزمون گیرندگان و هم آزمون شونده‌گان) که موفقیت خود را به افزایش توانایی نسبت دهند ولی این موضوع همیشه درست نیست.

در این تحقیق ابتدا فرضیه‌هایی بیان گردید و طرحهایی برای بررسی درستی یا نادرستی آنها اجرا شد. سپس تطابق نتایج حاصل از این طرحها با گزارشهای واقعی و اطلاعات حقیقی بررسی گردید.

فرضیه اول: معلومات افرادی که در امتحان اول مردود شده و در امتحانات بعدی موفق بوده‌اند افزایش نیافته؛ یعنی شانس، عامل اصلی موفقیت بوده است. برای بررسی این نظریه به افراد مردود شده در اولین امتحان اجازه دادند بدون هیچگونه پیشرفتی در توانایی بالقوه خود در امتحان بعدی شرکت کنند و امتحان مجدد بدون هر گونه آمادگی اضافی و بلافاصله گرفته شد. نتایج حاصل از این امتحان با گزارشهای واقعی مطابقت نداشت. پس این فرضیه می‌توانست درست باشد. البته به ندرت در مواردی ممکن است به دلیل ضریب تمیز پایین امتحان و همراهی سایر عوامل، قبولی در امتحان بدون آمادگی اضافی حاصل شود، ولی این موارد نادر دلیل درستی فرضیه نیست.

فرضیه دوم: افراد در بین امتحانات، توانایی خود را افزایش می‌دهند. برای بررسی این فرضیه از دانشجویان درخواست شد پس از مردودی در امتحان اول تلاش کرده، توانایی واقعی خود را افزایش دهند و سپس در امتحان مجدد شرکت کنند. نتایج این امتحان با اطلاعات واقعی مطابقت بیشتری داشته، مشخص کرد کسانی که معلومات خود را افزایش داده بودند درصد قبولی بیشتری داشتند.

با بررسی نتایج و اطلاعات واقعی دانشکده زبان مشخص شد که پیشرفت نسبتاً زیادی در توانایی بالقوه افراد در فاصله شرکت در امتحان اول و دوم ایجاد می‌شود. همچنین پیشرفت اضافی مختصری نیز قبل از سومین امتحان ایجاد شده، قبل از چهارمین

امتحان توانایی واقعی کاهش می‌یابد و این روال ادامه یافته که احتمالاً به علت خستگی فرد از امتحان و ناراحتی روحی حاصل از چندین بار شکست پیاپی است. از نتایج این تحقیق در کل مشخص شد که اگر امتحان ضریب تمیز بالایی داشته باشد بین اطلاعات حقیقی فرد و امکان قبولی او در این امتحان همبستگی وسیعی وجود دارد. یعنی عامل شانس در صورت مناسب بودن امتحان نسبتاً بی‌اهمیت و بی‌تأثیر است. در ضمن مشخص شد که توانایی افراد از چهارمین شرکت آنان در امتحان به بعد کاهش می‌یابد. این نتیجه درستی عمل دانشگاهها را در جهت ایجاد محدودیت برای تعداد دفعاتی که یک فرد می‌تواند در یک آزمون شرکت کند، نشان می‌دهد. بدون وجود محدودیت ممکن است افرادی برای بیست بار یا بیشتر در یک آزمون شرکت کرده، هیچ موفقیتی کسب ننمایند.

تبدیل نمرات آزمون چندگزینه‌ای

گاهی نمرات آزمون چندگزینه‌ای را بعد از نمره دادن کامل، به پنج طبقه^۱ تقسیم‌بندی می‌نمایند. در این روش، مبنای قضاوت نهایی، نمرات طبقه‌بندی شده به دست آمده است. اما این روش ارزش نمرات را نسبت به هم از بین می‌برد؛ زیرا یک محدوده نمره‌ای را به یک طبقه واحد تبدیل کرده، گاهی باعث می‌شود یک دانشجوی خوب و مستعد که بقیه موضوع را خوب فهمیده است نتواند آن درس را بگذراند.

مثال: در یک امتحان فرضی، سه موضوع مجزا مورد سؤال قرار گرفت. هر کدام از آنها ۱۰۰ نمره داشت. نمرات هر موضوع به صورت زیر به یک نظام طبقه‌بندی تبدیل شد.

$$۷۴ = ۵ \text{ بالای}$$

$$۷۴ \text{ تا } ۵۶ = ۴$$

$$۵۵ \text{ تا } ۴۶ = ۳$$

$$۴۵ \text{ تا } ۳۶ = ۲$$

$$۱ = \text{نمره زیر } ۳۶$$

نمره قبولی برای کل امتحان ۹ واحد از کل طبقات بود.

الف) یک دانشجوی معمولی نمرات زیر را از سه موضوع امتحان گرفت:

موضوع یک: ۴۶ موضوع دو: ۴۶ موضوع سه: ۴۶

نمره حقیقی: $۴۶ + ۴۶ + ۴۶ = ۱۳۸$

نمره ۴۶ در طبقه‌بندی معادل ۳ بود، پس نمره طبقه‌بندی: $۳ + ۳ + ۳ = ۹$

نمره قبولی ۹ در نظر گرفته شده بود، پس:

«دانشجو قبول شد».

ب) دانشجوی دیگر نمرات زیر را بدست آورد و اگر چه نمره حقیقی بالاتری کسب کرد، نسبت به مورد «الف» مردود شد.

موضوع یک: ۴۵ موضوع دو: ۴۵ موضوع سه: ۵۵

نمره حقیقی: $۴۵ + ۴۵ + ۵۵ = ۱۴۵$

نمره طبقه‌بندی: $۲ + ۲ + ۳ = ۷$

پس: «دانشجو مردود شد».

ج) یک دانشجوی نسبتاً باهوش نمرات زیر را گرفت:

موضوع یک: ۵۶ موضوع دو: ۵۶ موضوع سه: ۷۳

نمره حقیقی: $۵۶ + ۵۶ + ۷۳ = ۱۸۵$

نمره طبقه‌بندی: $۴ + ۴ + ۴ = ۱۲$

پس: «دانشجو قبول شد».

د) یک دانشجوی بسیار قوی و باهوش نمرات زیر را گرفت:

موضوع یک: ۷۲ موضوع دو: ۷۲ موضوع سه: ۷۲

نمره حقیقی: $۷۲ + ۷۲ + ۷۲ = ۲۱۶$

نمره طبقه‌بندی: $۴ + ۴ + ۴ = ۱۲$

پس: «دانشجو قبول شد».

مشاهده می‌شود که دانشجوی «ج» و «د» علی‌رغم نمره طبقه‌بندی مشابه از نظر نمره حقیقی با هم تفاوت دارند.

عوامل مؤثر در نمرات آزمون

عواملی که نمرات آزمون را تحت تأثیر قرار می‌دهد عبارت است از:

الف) عوامل مربوط به آزمون‌گیرنده؛

ب) عوامل مربوط به آزمون‌شونده؛

ج) عوامل اجرایی آزمون.

حال به شرح آنها می‌پردازیم.

الف) عوامل مربوط به آزمون‌گیرنده

۱- خطاهای مربوط به ارفاق: برای جلوگیری از این خطا و در صورت استفاده از نظام طبقه‌ای از شاخصی استفاده کنید که تنها یک حالت نامطلوب در بین پنج حالت ارزیابی داشته باشد مثلاً:

ضعیف - متوسط - خوب - خیلی خوب - عالی.

در این صورت ارزیابی‌ها به صورت متقارن در اطراف عبارت «خوب» پخش می‌شوند.

۲- تمایل به میانگین^۱: بیشتر آزمون‌گیرندگان تمایل ندارند که ارزیابی‌های خیلی دقیق برای دانشجویان انجام دهند؛ در نتیجه، نمرات همه داوطلبان در اطراف میانگین، متمرکز می‌شود. برای بررسی احتمال وقوع این مشکل از شاخصهای مرکزی استفاده کرده تا حد امکان سعی کنید دانشجویان متوسط را از هم جدا نمایید.

۳- اثر هاله‌ای^۲: گاهی یک نوع رفتار دانشجو به نظر آزمون‌گیرنده به قدری مهم جلوه می‌کند که بر ارزشیابی کلی تأثیر می‌گذارد. در این حالت این رفتار ویژه چنان مطلوب یا ناپسند جلوه می‌کند که قضاوت مدرس را در سایر زمینه‌ها با مشکل مواجه می‌سازد. تورندیک^۳ این پدیده را اثر هاله‌ای می‌نامد. اگر در ارزشیابی، زمینه‌های متفاوت و جنبه‌های گوناگون یک مسأله را بررسی کنید، این اثر کم می‌شود.

۴- خطای قیاسی^۴: این اشتباه شبیه اثر هاله‌ای بوده، هنگامی ایجاد می‌شود که آزمون‌گیرنده فرض می‌کند ارتباطی بین دو متغیر مورد ارزشیابی وجود دارد و هرگاه یک متغیر دارای یک نظم خاص باشد متغیر دوم نیز از آن پیروی خواهد کرد. اگر ارزشیابی مبنی بر یک عمل قابل مشاهده باشد از میزان این خطا کاسته می‌شود.

۵- خطا در مقابله^۵: هرگاه آزمون‌گیرنده، خیلی منظم و مرتب باشد ممکن است تصور کند دیگران برخلاف او نامنظم هستند. البته عکس این حالت نیز صادق است. به عنوان مثال: مردم گمان می‌کنند دیگران شبیه آنها هستند و با مشاهده خلاف آن متعجب می‌شوند.

۶- خطای نزدیکی^۶: نزدیکی دو ارزشیاب یا آزمون‌گیرنده با یکدیگر و امکان تماس و تبادل نظر بین آن دو باعث تأثیرپذیری آنها از همدیگر و در نتیجه تأثیر بر نمره‌گذاری آزمون می‌شود.

ب) عوامل مربوط به آزمون‌شونده

۱- میزان آشنایی با نوع آزمون: فردی که آشنایی قبلی با نوع سؤالات آزمون دارد نمره

1- Central Tendency

2- Halo Effect

3- Thorondik RL.

4- Logical Error

5- Contrast Error

6- Proximity Error

بیشتری می‌گیرد.

۲- حدس زدن: گاهی حدس پاسخ صحیح، باعث می‌شود فرد به طور تصادفی نمرهٔ
بیشتری بگیرد.

۳- تغییر پاسخها: معمولاً دوباره خوانی و تغییر پاسخها باعث می‌شود فرد به اشتباهات خود
پی برده، نمرهٔ بیشتری بگیرد.

۴- خوشنویسی و املا: بخصوص در آزمونهای تشریحی مرتب نوشتن و صحیح نوشتن
باعث می‌شود فرد نمرهٔ بیشتری بگیرد.

۵- کزافه و اضافه نویسی: اگر مصحح آزمون فرد با حوصله‌ای باشد اضافه نویسی باعث
می‌شود فرد نمرهٔ بیشتری بگیرد.

۶- دشواری متن پاسخها: معمولاً پاسخهایی که با متن ساده نوشته شده باشد، نمرهٔ
بیشتری می‌گیرد.

۷- اضطراب تسهیل کننده: با توجه به توضیحات قبل سبب می‌شود فرد نمرهٔ بیشتری
بگیرد.

ج) عوامل اجرایی آزمون

۱- استفاده از پاسخنامه جداگانه: پاسخنامه جداگانه، تصحیح و نمره‌دهی را دقیقتر کرده،
مقایسهٔ پاسخنامه‌های افراد مختلف و نمره‌گذاری واقعی‌تر را امکان‌پذیر می‌سازد.

۲- نحوهٔ اجرای آزمون: مدیریت داشتن در اجرای آزمون، دقت در تعیین زمان آزمون،
جلوگیری از تقلب، توزیع و جمع‌آوری برگه‌ها هر کدام به نوعی نمرات آزمون را تحت
تأثیر قرار می‌دهد.

۳- شرایط مکانی آزمون: سر و صدا، گرما و سرمای طاقت فرسا، راحت نبودن صندلی،
نبودن جای کافی برای بازکردن پرسشنامه، نبودن صندلی مناسب برای افراد چپ دست
و غیره در چگونگی پاسخ‌دهی دانشجو و نمرهٔ آزمون مؤثر است.

عینی بودن نمرهٔ آزمون^۱

عینی بودن نمرهٔ آزمون به رعایت عدالت، انصاف و نیز روایی و پایایی آن مربوط
است.

عواملی که عینی بودن نمره را کاهش می‌دهد عبارت است از:

۱- بد خطی دانشجو:

۲- خستگی مصحح:

۳- ذهنیت قبلی مصحح نسبت به دانشجو.

موارد مذکور باعث می‌شود دو نمره متفاوت به پاسخ یکسان در پاسخنامه‌های دو فرد داده شود. این مشکلات اکثراً در مورد امتحانات تشریحی وجود دارد.

با کاربرد روشهای زیر می‌توانید عینی بودن نمرات را افزایش دهید:

۱- اوراق را صرف‌نظر از غلط املائی، نگارش و دستور زبان تصحیح کنید:

۲- پاسخنامه را بدون اسم تصحیح کنید:

۳- اوراق را قبل از تصحیح مخلوط کنید، نامعلوم نشود هر ورقه مربوط به کیست:

۴- ابتدا یک سؤال را از پاسخنامه همه دانشجویان تصحیح کرده بعد به سراغ سایر سؤالات بروید:

۵- نمره هر سؤال را روی یک برگه جداگانه بنویسید تا هنگام تصحیح هر سؤال ندانید نمره سؤال قبل چه میزان بوده است.

تاثیر نمره دادن امتحان به صورت بدون اسم در میزان قبولی و ارتباط آن با جنس و نژاد

در گذشته گمان می‌شد ارزشیابی بالینی دانشجویان پزشکی و رزیدنتها در ارتباط با جنس و نژاد تحت تأثیر تورش قرار می‌گیرد و علت آن را زیاد بودن دلهره در خانمها و اقلیتهای نژادی می‌دانستند.

در یک پژوهش طی سالهای ۱۹۸۸-۱۹۹۱ در دانشگاه ایلنویز جنوبی^۱ نمره‌گذاری همراه با مشخص بودن هویت دانشجویان انجام شد. طی سالهای ۱۹۹۲-۱۹۹۵ برگه‌های امتحانی فقط با شماره‌هایی مشخص می‌شد که این عدد را فقط شخص دانشجو و یکی از کارمندان آموزش می‌دانستند ولی مصحح از آن اطلاعی نداشت. طی این سالها انواع آزمونهای چند گزینه‌ای، تشریحی، کوتاه پاسخ، عملی و آزمایشگاهی را اجرا کردند.

نمره قبولی در کل برای هر دانشجو از تقسیم مجموع نمرات قبولی بر تعداد کل امتحانات بدست آمد. میزان قبولی برای همه دانشجویان قبل از اجرای سیاست ۸۷/۰ درصد و بعد از آن ۸۷/۱ درصد بود. میزان قبولی برای مردان قبل از اجرای سیاست ۸۷/۵ درصد و بعد از آن ۸۷/۷ درصد بود. برای زنان قبل از اجرای سیاست ۸۶/۰ درصد و بعد از آن ۸۶/۱ درصد بود. در مورد امتحانات تشریحی میزان قبولی ۶۹/۳ درصد قبل و ۷۳/۶ درصد بعد از

اجرای سیاست بود. در مورد سفیدپوستان میزان قبولی قبل از اجرای سیاست ۷۱/۸ درصد و بعد از آن ۷۶/۹ درصد بود و در دانشجویان مرد سفیدپوست ۷۱/۶ درصد و بعد از آن ۷۷/۴ درصد بود.

پس با توجه به آمار ارائه شده، هیچ تفاوت قابل ملاحظه‌ای بین نتایج امتحانات برای مردان و زنان یا سیاهپوستان و سفیدپوستان قبل و بعد از اجرای سیاست وجود نداشت. در یک نظرخواهی از صاحب‌نظران، ۵۵ درصد با ادامه سیاست نمره‌دادن بدون اسم موافق بوده، معتقد بودند امکان تورش را از بین می‌برد. اما بقیه با این روش مخالف بودند؛ زیرا اعتقاد داشتند این روش موجب کاهش اطلاع استاد از وضعیت تحصیلی دانشجویان می‌گردد. همچنین میزان بازخورد را برای دانشجویان کاهش می‌دهد. البته ممکن است در این تحقیق تورش برای جنس و نژاد نشان داده نشده باشد ولی نمی‌توان گفت این تورش به طور کلی وجود ندارد. بنابراین بهتر است تصحیح اوراق بدون اسم باشد اما بعد از تصحیح، برای بررسی وضعیت تک‌تک دانشجویان با توجه به شماره اوراق، صاحب هر برگه را شناخته، در بهبود وضع تحصیلی، او را یاری دهید.

منابع:

- ۱- پاشا شریفی، حسن. اصول روان‌سنجی و روان‌آزمایی. تهران: انتشارات رشد، ۱۳۷۳.
- ۲- گرون لوند، نورمن. آزمونهای پیشرفت تحصیلی. تهران: انتشارات سازمان سنجش، ۱۳۵۶.
- ۳- گنجی، حمزه. اندازه‌گیری و ارزشیابی در آموزش و پرورش. تهران: انتشارات بعثت، ۱۳۷۳.
- ۴- گیلبرت. راهنمای آموزش برای تعلیم بهداشتکاران. ترجمه کیومرث ناصری. تهران: انتشارات تهران، ۱۳۶۴.
- ۵- مایزما، ویلیام؛ و همکاران. اندازه‌گیری و آزمون در تعلیم و تربیت. ترجمه غلامرضا نی‌نژاد. مشهد: انتشارات مؤسسه چاپ و انتشارات آستان قدس رضوی، ۱۳۷۲.
- 6- Dorsey Jk. *Effect of anonymous test grading on passing rates as related and race.* Acad Med 1995; 70(4): 321-29.
- 7- Mehrens WA. *Measurement and evaluation in education and psychology.* 3rd rev. ed. New York: Holt Saunders, 1984.
- 8- Pamphlett R. *Effect of anxiety on performance in multiple choice examination.* Med Educ 1995; 29: 297-302.
- 9- Robins I. *The effect of pass/fail grading and weekly quizzes on first-year student's performance and satisfaction.* Acad Med 1995; 70 (4): 327-29.
- 10- Tooth P. *Anxiety and study methods in preclinical students: causal relation to examination performance.* Med Educ 1989; 23(5): 416- 21.
- 11- Vargo D. *How can we deter cheating in medical school?* JAMA 1991 Nov 11; 266(17): 2456.

فصل دوازدهم

آزمونهای استاندارد

- مراحل استاندارد سازی آزمون
- امتحان استاندارد شده USMLE

تهیه یک آزمون استاندارد نیاز به صرف وقت زیاد و حوصله کافی دارد. استانداردسازی آزمون، طی چند مرحله صورت می‌گیرد؛ بدین ترتیب که از نتیجه هر مرحله برای انجام مرحله بعد استفاده می‌شود. فعالیت‌هایی که در این زمینه صورت گرفته، وقت‌گیر و تا حدی پیچیده است ولی برای داشتن یک آزمون خوب الزامی است. در اکثر آزمون‌ها، دانشجویان ماهر فقط به دلیل استاندارد نبودن آزمون، موقعیت‌های زیادی را برای نشان دادن توانایی خود از دست می‌دهند. در نظر گرفتن این مطلب، به کارگیری نیرو و صرف وقت کافی را برای طراحی یک آزمون مناسب توجیه می‌کند. البته می‌توانید سؤالات به دست آمده از یک دوره فعالیت استانداردسازی را برای آزمونهای بعدی نیز استفاده نمایید ولی توجه کنید که انجام این کار دو شرط دارد: نخست آنکه محتوای تدریس شما برای چند دوره متوالی آزمون ثابت مانده باشد و دوم آنکه دانشجویان از سؤالات شما در آزمونهای قبلی اطلاعی نداشته باشند.

مراحل استانداردسازی آزمون

- الزینا و کروکر^۱ یک برنامه ده مرحله‌ای را به عنوان حداقل تلاش لازم، برای اطمینان از داشتن یک وسیله اندازه‌گیری مناسب پیشنهاد کرده‌اند. این مراحل به ترتیب عبارتند از:
- ۱- تعیین اهداف آزمون؛
 - ۲- تعیین رفتارهای مهم در حوزه‌ای که باید اندازه‌گیری شود؛
 - ۳- فراهم کردن وسایل و امکانات لازم برای تهیه آزمون؛
 - ۴- طراحی سؤالات آزمون؛
 - ۵- بازنگری و مرور سؤالات آزمون؛
 - ۶- انجام پیش‌آزمون و مرور سؤالات؛

- ۷- انجام آزمون مقدماتی^۱ برای گروهی از داوطلبان آزمون آینده؛
- ۸- تجزیه و تحلیل سؤالات آزمون مقدماتی و جمع‌آوری اطلاعات برای برآورد پایایی و روایی سؤالات، به منظور شناسایی سؤالات ضعیف و بازسازی پرسشهای نارسا؛
- ۹- تعیین خط مشی برای انجام آزمون؛
- ۱۰- نمره‌دهی و تفسیر نمرات.

تعیین اهداف آزمون

هدف آزمون چیزی است که آزمون برای رسیدن به آن طراحی می‌شود. فرض کنید قرار است آزمونی برای تأیید شایستگی تعدادی دکتر داروساز^۲ برگزار شود. هدف از چنین آزمونی باید سنجش توانایی و شایستگی ویژه داروسازان برای انجام وظایف شغلی باشد. این شایستگی‌ها توسط گروه نمونه‌ای از داروسازان جامعه، داروسازان بیمارستانی و پژوهشگر و همچنین متخصصان سایر حرفه‌های وابسته تعیین می‌شود. شایستگی‌هایی که آنها مشخص می‌کنند می‌تواند شامل موارد زیر باشد:

- ۱- داشتن اطلاعات درمانی و دارویی در مورد داروهای نسخه‌ای و بی‌نسخه؛
- ۲- توانایی ارزیابی نحوه نسخه‌نویسی پزشکان؛
- ۳- نظارت و ارزیابی دارو درمانی؛
- ۴- ارائه دارو بر طبق نسخه پزشک؛
- ۵- مشورت و راهنمایی بیمار در زمینه نحوه مصرف داروهای تجویز شده؛
- ۶- همکاری با بیماران یا افراد سالم در زمینه ارائه اطلاعات دارویی؛
- ۷- مشورت با بیمار در مورد انتخاب، استفاده و آثار داروهای تجویز نشده؛
- ۸- آگاهی از موارد قانونی در مورد نحوه توزیع دارو و عملکرد داروخانه؛
- ۹- داشتن قدرت قضاوت و تصمیم‌گیری مناسب؛
- ۱۰- دارا بودن استانداردهای اخلاقی و عملی.

انجام پیش‌آزمون

با انجام چند پیش‌آزمون می‌توانید سؤالات نامناسب را از آزمون اصلی حذف کرده، آن را برای تعدادی از افرادی اجرا کنید که صلاحیت دارند در امتحان اصلی شرکت کنند.

1- Pilot

2- Pharmacy Doctor: Pharm. D

آزمونها به صورت کتاب باز^۱ ولی با نظارت مراقبان انجام شود. بعد از آزمون از امتحان شوندها بخواهید که در مورد وضوح سؤالات و ارتباط آنها با دامنه اطلاعاتی که آزمون جهت سنجش آن طراحی شده، اظهار نظر کنند. همچنین می‌توانید شاخص دشواری و قدرت تمیز را مشخص کرده، سؤالاتی را که خارج از محدوده مناسب هستند، حذف نمایید. تعیین میانگین نمره، نما، بالاترین و پایین‌ترین نمره و ضریب همبستگی نیز به شما کمک خواهد کرد.

انجام آزمونهای مقدماتی

برای این آزمونها دو نمونه مختلف از سؤالات تهیه کرده، در اختیار دو گروه متفاوت از آزمون شوندها داوطلب قرار دهید. با بررسی نمرات می‌توانید عوامل مختلفی که آزمونها را تحت تأثیر قرار می‌دهد مشخص کنید. به عنوان مثال با توجه به اینکه نمرات دانشجویان در گروه «الف» به طور قابل توجهی از نمرات گروه «ب» پایین‌تر بوده می‌توانید حدس بزنید سؤالات گروه «ب» پیچیده‌تر از سؤالات گروه «الف» است. در این مقایسه، شاخص دشواری هر کدام از انواع «الف» و «ب» که به حد مطلوب نزدیکتر است به عنوان شاخص دشواری استاندارد در نظر گرفته می‌شود و نمرات آزمون دیگر با نمرات آزمون با شاخص دشواری استاندارد، معادل می‌گردند. با انجام آزمونهای مقدماتی می‌توانید یک آزمون را برای دو گروه مختلف در زمانهای گوناگون برگزار کنید و نقش زمان دقیق برگزاری امتحان را بر عملکرد دانشجویان در امتحان بررسی نمایید.

برای بررسی روایی محتوای آزمون، گروهی از متخصصان در رشته مورد آزمون می‌توانند سؤال را از نظر داشتن توانایی برای سنجیدن اهداف اصلی آزمون، بررسی کنند. اگر روایی محتوا حتی توسط یک نفر از آنها تأیید نشود باید سؤال را به طراحان اصلی و نخست ارجاع دهید تا تصحیح یا حذف گردد.

با استفاده از همه اطلاعاتی که بدست آوردید، به حذف و اصلاح پرسشهای نامناسب پرداخته، تا رسیدن به حد مطلوب از شاخصهای مورد نظر، از پیش آزمون و آزمونهای مقدماتی و تجزیه و تحلیل آماری آزمونها استفاده نمایید.

برای آشنایی بیشتر با نمونه اجرایی این آزمونها به معرفی و شرح یکی از آنها می‌پردازیم.

USMLE

با توجه به برگزاری مداوم آزمونهای جامع علوم پایه، کارورزی و دستیاری توسط وزارت بهداشت، درمان و آموزش پزشکی و تأثیر آن بر روند جاری در دانشگاههای علوم پزشکی کشور در این فصل طراحی، برگزاری و نتایج اینگونه آزمونها مورد بحث قرار خواهد گرفت.

آزمون تأییدی مدرک پزشک در ایالات متحده^۱: عنوان مجموعه امتحاناتی است که توسط دانشگاههای علوم پزشکی کشورهای آمریکا، کانادا و پورتوریکو به منظور پذیرش دانشجویان خارجی در رشته‌های تخصصی پزشکی و همچنین تأیید شایستگی دانش‌آموختگان داخلی و خارجی رشته پزشکی برگزار می‌شود. این آزمون سه مرحله^۲ دارد. که بسیار شبیه امتحانات جامع علوم پایه، پیش‌کارورزی و دستیاری کشور ماست.

مرحله اول و دوم در مراکز وابسته به NBME و ECFMG^۳ انجام می‌شود و مرحله سوم به وسیله مؤسسات و حوزه‌هایی که گواهینامه پزشکی می‌دهند برگزار می‌گردد.

در مرحله اول مفاهیم کلیدی علوم پایه با تأکید بر مکانیسمهای تندرستی، بیماری‌راههای درمان به صورت محتوای تلفیق شده^۴ از تمامی دروس پایه، مورد سؤال قرار می‌گیرد. رشته‌های مورد آزمون عبارتند از: بیوشیمی، علوم تشریحی، فیزیولوژی، میکروبی‌شناسی، آسیب‌شناسی، داروشناسی و علوم رفتاری. تعداد سؤالات هشتصد عدد است و در چهار دفترچه تهیه می‌شود.

تقریباً نیمی از سؤالات با توصیف یک موقعیت بالینی آغاز شده، از فرد مورد آزمون می‌خواهد که علت داده‌های ارائه شده در مورد بیمار را توضیح دهد، یافته‌های بعدی را پیش‌بینی کرده، اقدام بعد در مراقبت از بیمار را به طور خاص بیان کند. البته آنچه از دانشجو خواسته می‌شود براساس اصول و مفاهیم علوم پایه است.

در مرحله دوم توانایی دانشجویان در فهم و به کار بستن علوم بالینی مورد سنجش قرار می‌گیرد تا مشخص شود دانشجو تا چه حد می‌تواند از این علوم در انجام مراقبت پزشکی تحت نظارت فرد دیگری استفاده کند. تعداد سؤالات این آزمون نیز هشتصد عدد بوده، در دو روز انجام می‌گیرد. رشته‌های مورد آزمون طب داخلی، اطفال، جراحی،

1- United State Medical Licencing Examination

2- Step

3- Educational Commission for Foreign Medical Graduates

4- Integrated

روانشناسی، زنان و زایمان و پزشکی پیشگیری است.

در مرحله سوم اساس برخورد بالینی و وظیفه پزشکی در برابر بیمار، بدون نظارت فرد دیگر مورد آزمون قرار می‌گیرد. در این آزمون تأکید خاصی بر اداره بیمار، شامل کارهای مقدماتی، مراقبتهای اورژانسی، معاینه فیزیکی، اطلاعات تشخیصی، تشخیص و درمان بیماری می‌شود. امتحان دارای ۷۲۰ سؤال بوده، مدت آن دو روز است.

در آزمونهای USMLE درصد سؤالاتی که هدف آنها یادآوری است، کاهش یافته، بیشتر سؤالات در زمینه تفسیر جداول و تصاویر و موقعیتهای بالینی و سایر مهارتهای حل مسأله است.

سؤالات آزمون از نوع چند گزینه‌ای، جورکردنی و نوع جدیدی به نام جورکردنی وسعت یافته^۱ است. سؤالات جورکردنی وسعت یافته همان سؤالات جورکردنی قبلی است؛ با این تفاوت که گزینه‌های آن طولانی‌تر شده، از کلمات بیشتری استفاده کرده است. سؤالات این آزمون نمره منفی ندارد. قبل از آزمون پرسشنامه‌ای محتوی ۱۲۰ سؤال نمونه در اختیار دانشجو قرار می‌گیرد تا با شیوه کلی طرح سؤال آشنا شوند.

پیش از این، استانداردهای قبولی-ردی در این آزمون براساس رتبه فرد در مقایسه با گروه مورد امتحان یعنی طبق شیوه ملاک نسبی بود و توانایی فرد را در مقایسه با افراد مورد امتحان می‌سنجید. اما اکنون استانداردهای قبولی-ردی بر طبق شیوه ملاک مطلق بوده، نمره‌ای که برای گذراندن درس لازم است از طریق مرور و بررسی سؤالات آزمون و تعیین سطح مهارت لازم در مورد هر کدام تعیین می‌شود. برای قبول شدن در این آزمونها دانشجویان حداقل باید به ۵۵ درصد تا ۶۵ درصد سؤالات پاسخ صحیح دهند. البته حداقل نمره قبولی در مرحله اول از مرحله دوم بیشتر است.

نمرات ارائه شده برای هر فرد در مرحله اول دربردارنده اطلاعات زیر است:

۱- طرح قبولی-ردی USMLE؛

۲- نمره کلی عددی؛

۳- میزان درصد آن نمره؛

۴- طرحی مختصر راجع به تواناییهای فرد در مورد هر یک از سیستمهای بدن.

به علاوه، دانشکده‌ها نیز جداولی دریافت می‌کنند که در آنها نمرات کلی فهرست شده و نتایج قبول یا رد دانشجویان به دانشکده گزارش می‌شود. همچنین گزارش کلی راجع به

تواناییهای کل گروه دانشجویان در مورد هر عضو بدن و رشته درسی نیز ارائه می‌گردد. سؤالات امتحان برای تمام آزمون شوندگان مشابه است ولی ترتیب آن فرق دارد. همه آزمون‌شوندگان حداقل یکبار در طول آزمون باید محل نشستن خود را عوض کنند. برای شرکت در مراحل اول و دوم، دانشجو باید دارای یکی از خصوصیات زیر باشد:

- ۱- دانشجوی پزشکی رسمی در یکی از کشورهای آمریکا، کانادا و یا پورتریکو بوده، توسط کمیته مسؤل آموزش پزشکی (LCME) تأیید شده باشد؛ یا
- ۲- دانشجوی پزشکی در کشوری غیر از سه کشور بالا بوده، توسط کمیته آموزشی دانشجویان خارجی (ECFMG) تأیید شده باشد.

اگر دانشجو یکی از خصوصیات بالا را داشته باشد، می‌تواند در آزمونهای مرحله اول و دوم در هر زمانی از دوره تحصیلی یا بعد از آن شرکت کند. برای شرکت در مرحله دوم لازم نیست امتحان مرحله اول، را پشت سر گذاشته باشد اما برای شرکت در مرحله سوم باید مراحل اول و دوم را طی کرده باشد. به دانشجویان توصیه می‌شود سه مرحله را در طول هفت سال تحصیل پزشکی خود طی کنند؛ چون در این صورت برای شرکت در آنها، هفت سال فرصت دارند که با پشت سر گذاشت اولین مرحله شروع می‌شود. هر فرد حداکثر شش بار می‌تواند در هر آزمون شرکت کند و هر مرحله، دو بار در سال برگزار می‌شود. برای شرکت در مرحله سوم فرد باید حائز تمام ویژگیهای زیر باشد:

- ۱- مدرک پزشکی یا معادل آن و یا مدرک DO داشته باشد؛
- ۲- مرحله اول و دوم را طی کرده باشد؛
- ۳- اگر دانشجوی خارج از سه کشور آمریکا، کانادا و پورتریکو است، توسط ECFMG تأیید شده باشد؛
- ۴- ویژگیهای اختصاصی دانشگاهی که قصد ادامه تحصیل را در آن دارد، داشته باشد؛ مثل: گذراندن دوره‌های آموزشی خاص بعد از تحصیل.

تأثیر آزمونهای USMLE بر روند آموزشی دانشکده‌های پزشکی

۱- تأثیر بر آزمونهای پیشرفت تحصیلی محلی هر دانشگاه

از آزمونهای دانشکده‌ها استنباط می‌شود که سؤالات از نوع صحیح-غلط و الف/ب/هر دو/هیچکدام^۱ از بین رفته‌اند؛ زیرا مسؤلان دانشکده‌ها علاقه‌مندند دانشجویان

1- A/B/Both/Neither

خود را با قالب به کار رفته در امتحانات USMLE آشنا کنند. از آنجا که کاربرد این نوع سؤالات برای آزمون در سطح یادآوری بوده، این حذف و تغییر، بسیار مثبت است. به علاوه دانشکده‌ها سعی دارند آزمونهایی اجرا کنند که حاصل تلفیق علوم پایه و بالینی باشد. انجام این کار - به خصوص اگر طراحان، پیش زمینه گسترده‌ای در زمینه بیولوژی نداشته باشند مشکل است. - برای رفع این مشکل، متخصصان رشته‌های مختلف می‌توانند در طراحی یک آزمون، در پایان سال، با هم همکاری داشته باشند.

۲- تاثیر بر برنامه آموزشی دانشکده

مسئولان دانشکده‌ها اغلب این سؤال را طرح می‌کنند که آیا آزمونهای USMLE آنچه را در دانشکده‌ها تدریس می‌شود مشخص می‌کند یا می‌خواهد آنچه را باید تدریس شود، معین نماید. پاسخ به این سؤال برای اعضای NBME بسیار مشکل است؛ زیرا اگر بگویند امتحان نشان‌دهنده چیزی است که تدریس می‌شود آنگاه متهم به «جلوگیری از نوآوری در آموزش» می‌شوند و اگر بگویند این آزمون نشان‌دهنده چیزهایی است که باید تدریس شود متهم به «رهبری برنامه آموزشی» می‌شوند.

به طور کلی جنبه‌های گسترده علمی که در امتحانات مرحله‌ای مورد آزمون قرار می‌گیرد کاملاً با اهداف اعطای مدرک پزشکی و دانسته‌های موردنیاز برای فارغ‌التحصیل شدن از بیشتر دانشکده‌ها، هماهنگی دارد.

برای اینکه بدانیم USMLE چه چیزی را می‌سنجد و چه چیزی را در بر می‌گیرد لازم است این مقوله‌ها از هم جدا شود:

۱- اهداف نظام آموزشی^۱؛

۲- عملکرد یا فرآیند نظام آموزشی^۲.

هدف، چیزی است که باید فراگرفته شود و عملکرد، دوره‌های درسی و روشها^۳ به کار رفته در آن است. می‌توان در عین تشابه در اهداف آموزشی، عملکردهای مختلفی اتخاذ کرد. هثل: یادگیری بر مبنای حل مسأله^۴، روش سنتی و ... هدف تمامی این عملکردها آن است که دانشجو درک کاملی از علوم پایه بالینی و تلفیق آنها داشته، به راحتی عهده حل معضلات بیماران برآید (یعنی به اصلاح یک حلال مشکل^۴ خوب باشد).

هر دانشکده در برنامه آموزشی خود سه چیز را مشخص می‌کند:

1- Curriculum Goal
3- Problem Based Learning

2- Curriculum Process
4- Problem Solver

۱- چگونه تدریس کند^۱؛

۲- چه چیزی تدریس کند^۲ و

۳- چه وقت تدریس کند^۳.

نکته اساسی آن است که هر دانشکده باید فرآیند تدریس در راستای امتحانات جامع تخصصی باشد؛ یعنی دانشکده، تدریس خود را با هدف شرکت در امتحانات مذکور پایه‌ریزی کند؛ حتی اگر چنین برنامه‌ریزی با برنامه‌های یک یا چند گروه دانشگاهی در تضاد قرار گیرد. البته سؤالات آزمونهای USMLE نسبت به عملکرد برنامه آموزشی بی‌طرف بوده، هیچ گرایشی به ساختارهای آموزشی مختلف ندارد.

۳- تاثیر بر کیفیت ارتقای دانشجویان

بیشتر دانشکده‌های پزشکی برای تصمیم‌گیری در مورد نحوه ارتقای دانشجویان به سال بالاتر خواستار گزارش نمرات دانشجویان در مرحله اول و دوم از NBME هستند. البته به دلیل فقدان انطباق کامل محتوای این آزمونها با اهداف و زمان‌بندی برنامه آموزشی دانشگاهها این کار نادرست است. حتی اگر این مشکل هم وجود نداشته باشد، NBME توصیه می‌کند از نمرات دانشجویان در امتحان USMLE برای سیستم قبول-رد استفاده نکنید؛ چون دانشجوی خوبی که نتواند در امتحانهای مرحله‌ای موفق شود سرانجام آن را پشت سر گذاشته، دانشجوی ضعیفی که به سختی از آن، نمره قبولی بگیرد در عنوان مراحل بعدی تحصیل حتماً دچار مشکل خواهد شد. اگر آزمون USMLE به عنوان تنها معیار ارتقا به کار رود قضاوت دانشکده‌ها راجع به دانشجویان خود بی‌اعتبار خواهد شد. در مجموع نتایج این آزمونها برای بهتر روشن نمودن قضاوت مسؤولان دانشکده‌هاست نه جایگزین آن.

۴- تایید فراغت از تحصیل

معیار قرار دادن موفقیت در امتحان مراحل اول و دوم در فراغت از تحصیل افراد نامطلوب است؛ چون هیچ آزمون کتبی وجود ندارد که بتواند همه ویژگیهای مهم را در آموزش اندازه بگیرد. پس آزمونهای USMLE تنها معیار فراغت از تحصیل نیست، ولی به عنوان یک مرحله و مانع مستقل و جداگانه بسیار سودمند است. به عنوان مثال: اگر دانشجوی سال چهارم هنوز موفق به گذراندن امتحان مرحله اول نشده، می‌توان نتیجه گرفت او آماده تقبل مسؤلیت یک کارورز نیست.

۵- تاثیر بر سیاست اخراج از دانشگاه

دانشجویان حداقل سه و حداکثر شش بار می‌توانند در هر آزمون مرحله‌ای شرکت کنند و موفق نشدن پس از شش بار شرکت در آزمون می‌تواند بر سیاست اخراج دانشجویان اثر بگذارد.

۶- تاثیر بر یادگیری

آزمونهای USMLE بشدت، نوع مطلب و نحوه خواندن آن را توسط دانشجو تحت تأثیر قرار می‌دهد. یادگیری دانشجو به طرز مشخص فقط متکی بر فعالیت برنامه آموزشی یا تدریس نیست. پس آزمون USMLE می‌تواند بدون تأثیر در برنامه آموزشی، یادگیری دانشجو را متأثر کند.

مثال: فرض کنید در یک دانشکده یک برنامه آموزشی یادگیری بر مبنای حل مسأله در بردارنده محتوای ژنتیک مولکولی نباشد، در نتیجه دانشجویان برای آماده شدن در امتحانات USMLE به احتمال زیاد باید خودشان این سرفصل‌ها را مطالعه کنند. حتی ممکن است برنامه‌ای ترتیب دهند که گروه بیوشیمی دانشکده تعدادی مقاله در این زمینه ارائه نماید.

نتیجه کلی

آزمونهای USMLE چند اثر کلی دارد که از آن جمله می‌توان موارد زیر را نام برد:

- ۱- ترغیب امتحان شوندگان برای تلفیق اطلاعات خود از مقوله‌های مختلف؛
- ۲- صرف وقت بیشتر آزمون شوندگان بر جنبه‌های کلیدی و اصولی علوم پایه که در امتحان مرحله اول مورد آزمون قرار می‌گیرد، (به جای حفظ مفاهیم خشک علوم پایه)؛
- ۳- وادار کردن دانشجو به مطالعه بیشتر آن دسته از سرفصلهای درسی که احتمالاً در آینده کاربرد خواهد داشت؛
- ۴- بیگانه نبودن دانشجو نسبت به پیشرفتهای علمی آینده.

منابع

- 1- Battles JB. *Analyzing and adjusting for variables in a large - scale standardized- patient examination.* Acad Med 1994; 69(5): 370-6.
- 2- Becker D. *Results of the initial administration of NBME comprehensive part I, II examinations.* Acad Med 1991; 66: 516-20.
- 3- Carline J. *Predicting performance during clinical years from the new medical college admission test.* J Med Educ 1983; 58 (1): 18-25.
- 4- Carter JH. *Curricular changes and improved performance by highrisk students on the national boards part II.* Assess Acad Minor phys 1983; 4(3): 82-8.
- 5- Case S. *Characteristics of new licensing examination.* Acad Med 1992; 67(10): 513-15.
- 6- Case S. *Validity of NBME Part I, II scores for selection of residents in orthopaedic surgery - dermatology and preventive medicine.* Acad Med 1993; 68(2Supp): S51- 6.
- 7- Fielding D. *Developing an Assessment of pharmacy practice knowledge.* Am Pharm Educ 1994 Win; 58: 361-8.
- 8- Ronai A. *Relationship between past academic performance and results of specialty in-training examinations.* J Med Educ 1984; 59 (4): 341-4.
- 9- Swanson D. *Impact of the USMLE step 1 on teaching and learning of the basic biomedical science.* Acad Med 1992; 67 (9): 553-6.

واژه‌نامه

ژ

آزمون گزیده پاسخ، ۳۲	آزمون، ۹، ۸، ۷، ۶، ۵، ۴، ۳، ۲، ۱، ۰، ۱۱، ۱۲، ۱۳، ۱۴، ۱۵، ۱۶، ۱۷، ۱۸، ۱۹، ۲۰
آزمون معلم ساخته، ۳۰، ۳۸، ۳۷	۲۱، ۲۲، ۲۳، ۲۴، ۲۵، ۲۶، ۲۷، ۳۰، ۳۱، ۳۲، ۳۳
آزمون ملاک مطلق، ۳۹، ۴۸، ۴۵	۳۴، ۳۵، ۳۶، ۳۷، ۳۸، ۳۹، ۴۰، ۴۱، ۴۲، ۴۳، ۴۴، ۴۵، ۴۶، ۴۷، ۴۸، ۴۹
آزمون ملاک نسبی، ۴۸، ۴۵	۵۰، ۵۱، ۵۲، ۵۳، ۵۴، ۵۵، ۵۶، ۵۷، ۵۸، ۵۹، ۶۰، ۶۱، ۶۲، ۶۳، ۶۴، ۶۵، ۶۶، ۶۷، ۶۸، ۶۹، ۷۰، ۷۱، ۷۲
آزمون مهارت، ۵۱، ۴۹	۷۳، ۷۴، ۷۵، ۷۶، ۷۷، ۷۸، ۷۹، ۸۰، ۸۱، ۸۲، ۸۳، ۸۴، ۸۵، ۸۶، ۸۷، ۸۸، ۸۹، ۹۰، ۹۱، ۹۲، ۹۳، ۹۴، ۹۵، ۹۶، ۹۷، ۹۸، ۹۹، ۱۰۰، ۱۰۱، ۱۰۲
آزمونهای پیشرفت تحصیلی، ۲۰۲، ۳۹	۱۰۳، ۱۰۴، ۱۰۵، ۱۰۶، ۱۰۷، ۱۰۸، ۱۰۹، ۱۱۰، ۱۱۱، ۱۱۲، ۱۱۳، ۱۱۴، ۱۱۵، ۱۱۶، ۱۱۷، ۱۱۸، ۱۱۹، ۱۲۰، ۱۲۱، ۱۲۲، ۱۲۳، ۱۲۴، ۱۲۵، ۱۲۶، ۱۲۷
۲۸۶، ۲۳۱، ۲۰۵	۱۲۸، ۱۲۹، ۱۳۰، ۱۳۱، ۱۳۲، ۱۳۳، ۱۳۴، ۱۳۵، ۱۳۶، ۱۳۷، ۱۳۸، ۱۳۹، ۱۴۰
آزمونهای تشریحی، ۱۱، ۷۸، ۱۰۱، ۱۰۲	۱۴۱، ۱۴۲، ۱۴۳، ۱۴۴، ۱۴۵، ۱۴۶، ۱۴۷، ۱۴۸، ۱۴۹، ۱۵۰، ۱۵۱، ۱۵۲، ۱۵۳
۲۷۲، ۱۰۳، ۱۰۳	۱۵۴، ۱۵۵، ۱۵۶، ۱۵۷، ۱۵۸، ۱۵۹، ۱۶۰، ۱۶۱، ۱۶۲، ۱۶۳، ۱۶۴، ۱۶۵، ۱۶۶، ۱۶۷، ۱۶۸، ۱۶۹، ۱۷۰، ۱۷۱، ۱۷۲
آزمونهای رغبت، ۵۰، ۴۹	۱۷۳، ۱۷۴، ۱۷۵، ۱۷۶، ۱۷۷، ۱۷۸، ۱۷۹
آزمونهای شخصیت، ۱۸۱، ۵۰، ۳۹	۱۸۰، ۱۸۱، ۱۸۲، ۱۸۳، ۱۸۴، ۱۸۵، ۱۸۶، ۱۸۷، ۱۸۸، ۱۸۹، ۱۹۰، ۱۹۱، ۱۹۲، ۱۹۳، ۱۹۴، ۱۹۵، ۱۹۶، ۱۹۷، ۱۹۸، ۱۹۹، ۲۰۰، ۲۰۱، ۲۰۲
آزمونهای شفاهی، ۱۱۱، ۱۲۵، ۱۲۶، ۱۳۰	۲۰۳، ۲۰۴، ۲۰۵، ۲۰۶، ۲۰۷، ۲۰۸، ۲۰۹، ۲۱۰، ۲۱۱
۱۳۱	۲۱۲، ۲۱۳، ۲۱۴، ۲۱۵، ۲۱۶، ۲۱۷، ۲۱۸، ۲۱۹، ۲۲۰، ۲۲۱، ۲۲۲، ۲۲۳
آزمونهای هوش و استعداد، ۵۱، ۴۹	۲۲۴، ۲۲۵، ۲۲۶، ۲۲۷، ۲۲۸، ۲۲۹، ۲۳۰، ۲۳۱، ۲۳۲، ۲۳۳، ۲۳۴
آزمونهای استاندارد، ۴۰، ۳۹، ۲۱	۲۳۵، ۲۳۶، ۲۳۷، ۲۳۸، ۲۳۹، ۲۴۰، ۲۴۱، ۲۴۲، ۲۴۳، ۲۴۴، ۲۴۵، ۲۴۶
آزمون اندازه بیمار، ۱۱۸، ۱۱۲، ۱۱۱، ۳۳	۲۴۷، ۲۴۸، ۲۴۹، ۲۵۰، ۲۵۱، ۲۵۲، ۲۵۳، ۲۵۴، ۲۵۵، ۲۵۶، ۲۵۷، ۲۵۸، ۲۵۹، ۲۶۰، ۲۶۱، ۲۶۲، ۲۶۳، ۲۶۴، ۲۶۵
۲۸۵، ۱۶۷، ۱۶۵	۲۶۶، ۲۶۷، ۲۶۸، ۲۶۹، ۲۷۰، ۲۷۱، ۲۷۲، ۲۷۳، ۲۷۴
	۲۸۵، ۲۸۶، ۲۸۷، ۲۸۸، ۲۸۹
	آزمون استاندارد شده، ۳۷، ۳۸، ۳۹
ارزشیابی، ۲۲، ۲۱، ۲۰، ۱۹، ۱۸، ۱۷، ۱۴، ۱۱، ۵، ۶	آزمون باز پاسخ، ۳۳
۲۵، ۲۸، ۲۹، ۳۱، ۳۳، ۳۹، ۴۷، ۴۹، ۱۰۷، ۱۳۰، ۱۴۵	آزمون تشخیصی، ۵۱، ۴۹
۱۷۱، ۱۷۷، ۱۸۰، ۲۱۷، ۲۳۱، ۲۳۲، ۲۳۳، ۲۷۵	آزمون چند گزینه‌ای، ۸۵، ۷۸
ارزشیابی تشخیصی، ۲۰، ۱۹	آزمون ذهنی، ۲۰۰، ۳۲، ۳۰
ارزشیابی تکوینی، ۱۹، ۱۸	آزمون سرعت، ۲۶۷، ۵۲
ارزشیابی مجموعی، ۱۹، ۱۸	آزمون عینی، ۳۱، ۳۰
	آزمون قدرت، ۲۵۹، ۵۲

۲۲۱، ۱۸۱، ۱۷۸

۲۸۲، ۲۳۵، ۲۳۲، ۲۳۲

ن

نمره دادن، ۱۰۳، ۱۷۶، ۲۷۱، ۲۷۶

نمره منفی، ۳۳، ۳۱، ۸۵، ۱۱۲، ۱۲۰، ۱۷۵، ۲۶۷

۲۶۸، ۲۶۹، ۲۸۵

و

ویرایش آزمون، ۳۱

ز

هم آزمایی، ۸

همبستگی دو رشته‌ای نقطه‌ای، ۲۰۹، ۲۲۹

۲۳۰، ۲۳۱، ۲۳۲

ع

عوامل سؤال، ۲۱۸، ۲۳۵

ف

فهرستهای بازبینی، ۸، ۱۲۱، ۱۶۷، ۱۷۳، ۱۷۵

۱۸۵

م

مقیاس، ۱۶، ۸۳، ۹۳، ۱۲۷، ۱۷۵، ۱۷۶، ۱۷۷، ۱۷۸

۱۷۹، ۱۸۱، ۱۸۲، ۱۸۳، ۱۸۴، ۱۸۵، ۱۸۶، ۲۲۱

مقیاس دوتایی، ۱۷۷، ۲۲۹

مقیاس درجه بندی، ۱۷۲، ۱۷۵، ۱۷۶، ۱۷۷

Academic Achievement Tests

i
n
M
e
d
i
c
a
l
S
c
i
e
n
c
e
s



Edited by

Dr. B. Zolfaghari

Dr. Gh. Asadollahi

ISBN 964-92438-1-X

شابک ۹۶۴-۹۲۴۳۸-۱-X