



راهنمای عملی آزمون اسکریپت کانکوردنس

در ارتباط با TOP TIPS

مرکز مطالعات و توسعه آموزشی علوم پزشکی، با هدف توانمندسازی اساتید، بر آن است تا نکات کلیدی و کاربردی در ارتباط با مباحث مختلف آموزش علوم پزشکی را در قالب برگه‌های آموزشی الکترونیکی تحت عنوان **Top Tips** در اختیار اعضای هیات علمی قرار دهد. هر یک از این برگه‌ها حاوی اطلاعات مختصر و نکات کلیدی و کاربردی در ارتباط با یک موضوع خاص در آموزش پزشکی می‌باشند که می‌تواند در امر آموزش مورد استفاده اساتید محترم قرار گیرد. در این شماره ترجمه AMEE Guide شماره ۷۵ که یک راهنمای عملی در ارتباط با طراحی و اجرای آزمون توافق شرحنامه (اسکریپت کانکوردنس) می‌باشد و در اولین جلسه برنامه ساعت توسعه آموزش EDC در تاریخ ۲۴ مرداد توسط آقای دکتر قدیر پوربایرامیان به صورت مجازی ارائه گردید، آورده شده است.

شناسه محتوا

عنوان: آزمون **Script Concordance**: از تئوری تا عمل (امی گاید شماره ۷۵)

ترجمه و گردآوری: دکتر قدیر پوربایرامیان، عضو هیات علمی مرکز مطالعات و توسعه آموزش پزشکی

عنوان دپارتمان: کمیته ارزیابی و سنجش دانشجو، مرکز مطالعات و توسعه آموزش پزشکی

مخاطبان: اعضای هیات علمی و دانشجویان دوره دکتری حرفه ای و تحصیلات تکمیلی

Reference: Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE guide no. 75. Medical teacher. 2013 Mar 1;35(3):184-93.

آزمون Script Concordance^۱ (SCT) در آموزش علوم پزشکی برای ارزیابی جنبه خاصی از شایستگی استدلال بالینی استفاده می شود. توانایی تفسیر اطلاعات پزشکی در شرایط عدم قطعیت. SCT مبتنی بر مدل های نظری تثبیت شده سازمان دهی دانش و استدلال بالینی، دارای ۳ ویژگی کلیدی طراحی است: (۱) پاسخ دهندگان با موقعیت های بالینی بد تعریف شده ای مواجه می شوند و باید بین چندین گزینه واقعی انتخاب کنند. (۲) قالب پاسخ منعکس کننده نحوه پردازش اطلاعات در موقعیت های چالش برانگیز حل مسئله است. و (۳) امتیاز دهی تنوع پاسخ های کارشناسان به موقعیت های بالینی را در نظر می گیرد. نمرات SCT به این معنی است که نشان دهد توانایی پاسخ دهندگان در تفسیر داده های بالینی چقدر با پزشکان با تجربه در یک حوزه دانشی خاص مقایسه می شود. مجموعه قابل توجهی از تحقیقات از اعتبار ساختاری، قابلیت اطمینان و امکان سنجی SCT در انواع رشته های علوم پزشکی و در سراسر طیف آموزش علوم پزشکی از آموزش پیش بالینی تا توسعه حرفه ای مستمر پشتیبانی می کند. در عمل، عملکرد آن به عنوان یک ابزار ارزیابی به توسعه دقیق آیتم و انتخاب دقیق پانل بستگی دارد. این راهنما که به عنوان آغازگری برای افراد ناآشنا در SCT در نظر گرفته شده است، اصول اساسی، مبانی نظری و اصول ساختاری حاکم بر آزمون اسکرپت کانکوردانس را پوشش می دهد.

نکات کاربردی

- ❖ آزمون Script Concordance برای ارزیابی جنبه خاصی از شایستگی استدلال بالینی استفاده می شود. توانایی تفسیر داده های بالینی در شرایط عدم قطعیت.
- ❖ Script Concordance دارای ۳ ویژگی کلیدی طراحی است: (۱) آزمون دهندگان با موقعیت های بالینی نامشخص مواجه می شوند و باید بین چندین گزینه واقع بینانه انتخاب کنند. (۲) قالب پاسخ منعکس کننده نحوه پردازش اطلاعات در موقعیت های پیچیده حل مسئله است. و (۳) امتیاز دهی تنوع پاسخ های متخصصان به موقعیت های بالینی را در نظر می گیرد.
- ❖ Script Concordance بر اساس اصول نظریه اسکرپت بیماری است، که از متون روانشناسی شناختی از یک بحث بزرگتر در مورد ماهیت و توسعه تخصص پدید آمده است.
- ❖ شواهدی که از اعتبار، قابلیت اطمینان و امکان سنجی SCT حمایت می کنند از تحقیقات انجام شده در بسیاری از شاخه های علوم پزشکی و در سراسر طیف آموزش علوم پزشکی از آموزش پیش بالینی تا توسعه حرفه ای مستمر به دست آمده است.
- ❖ توصیه های عملی و مبتنی بر شواهد برای راهنمایی ساخت آزمون Script Concordance وجود دارد.

مقدمه

آزمون تطابق اسکرپت (SCT) در آموزش حرفه های بهداشتی برای ارزیابی جنبه خاصی از شایستگی استدلال بالینی استفاده می شود. توانایی تفسیر اطلاعات پزشکی در شرایط عدم قطعیت. این ویژگی های روانسنجی مطلوب (روایی سازه، قابلیت اطمینان، امکان سنجی) را در تحقیقات انجام شده در انواع رشته های علوم پزشکی نشان داده است، و در سراسر طیف آموزش علوم پزشکی از مقطع کارشناسی تا تحصیلات تکمیلی و توسعه حرفه ای مستمر. زیربنای نظری آن، که ریشه در نظریه اسکرپت از روانشناسی شناختی دارد، دائماً موضوع تحقیقات علمی بوده است. شیوه هایی برای ساخت دقیق SCT ها ایجاد شده است و به طور سیستماتیک بررسی شده است.

^۱ به فارسی آزمون توافق متن یا آزمون همخوانی شرحنامه

این راهنما برای مخاطبانی از اساتید علوم پزشکی در نظر گرفته شده است که آشنایی کمی با SCT یا منطق زیربنایی آن ندارند. هدف آن سوق دادن خواننده به سمت اصول اولیه، مفاهیم نظری و اصول ساختاری حاکم بر SCT است. در بخش اول مروری کلی بر رویکرد SCT ارائه خواهد شد. در بخش دوم، مبانی نظری قالب آزمون مورد بحث قرار خواهد گرفت. در بخش سوم، توصیه‌های کاربردی و مبتنی بر شواهد برای طراحی آزمایشی ارائه خواهد شد.

نمای کلی: اصول آزمون

مشخصات طراحی

SCT یک آزمون کتبی برای ارزیابی استدلال در شرایط عدم قطعیت است. در یک SCT، سناریوهای بالینی مختصری به آزمون شوندگان ارائه می‌شود، و به دنبال آن یک سری سوالات قضاوت در مورد احتمالات تشخیصی یا گزینه‌های مدیریتی در هنگام ارائه عناصر جدید اطلاعات را به دنبال دارد. اگرچه زمینه بالینی کافی برای اجازه دادن به تصمیم‌گیری‌های معنادار داده شده است، مقدار مشخصی از عدم قطعیت، عدم دقت یا ناقص بودن عمداً در هر مورد تعبیه شده است تا شرایط مبهمی را که اغلب مشخصه برخوردهای بالینی در موقعیت‌های واقعی است، شبیه‌سازی کند.

SCT علاوه بر تکیه بر مشکلات بالینی نامشخص، دو ویژگی کلیدی دیگر در طراحی دارد. اولین مورد این است که قالب پاسخ منعکس‌کننده روشی است که اطلاعات پزشکی اغلب در موقعیت‌های چالش برانگیز حل مسئله پردازش می‌شود، طبق مدل‌های نظری تثبیت شده سازماندهی دانش و استدلال بالینی که از تحقیقات روانشناسی شناختی و آموزش پزشکی به دست آمده است. دوم این است که برخلاف اکثر اشکال مرسوم ارزیابی، هیچ پاسخ صحیحی برای سوالات SCT وجود ندارد. در عوض، چندین پاسخ به هر یک از سوالات آزمون ممکن است قابل قبول در نظر گرفته شود، همانطور که به طور مستقل توسط اعضای یک پانل مرجع متشکل از پزشکان مجرب که از یک رشته یا حوزه دانش انتخاب شده‌اند برای تنظیم کلید نمره‌گذاری آزمون تعیین می‌شوند.

فرمت محرک و پاسخ

محرک آزمون شامل یک سناریوی بالینی کوتاه و به دنبال آن مجموعه‌ای از سوالات شامل سه قسمت است. بخش اول («اگر به فکر...»²) فرضیه‌ای را در قالب یک امکان تشخیصی، یک گزینه تحقیقاتی یا یک جایگزین درمانی ارائه می‌دهد. بخش دوم («...و سپس پیدا می‌کنید...»³) اطلاعات جدیدی را ارائه می‌کند، مانند یک علامت معاینه فیزیکی، یک وضعیت از قبل موجود، یک مطالعه تصویربرداری، یا یک نتیجه تست آزمایشگاهی، که ممکن است (یا ممکن است نداشته باشد) بر فرضیه داده شده این سوال در بخش سوم پاسخ داده شده است («...این فرضیه می‌شود»⁴)، که شامل یک مقیاس پاسخ از نوع لیکرت است (معمولاً از ۲- تا ۲+ متغیر است). آزمودنی‌ها در این مقیاس تأثیری را که فکر می‌کنند اطلاعات جدید (قسمت ۲) احتمالاً بر فرضیه پیشنهادی (قسمت ۱) خواهد داشت را نشان می‌دهند. نمونه‌هایی از موارد SCT در شکل ۱ ارائه شده است.

² "If you were thinking of..."

³ "...and then you find..."

⁴ "...this hypothesis becomes:"

شکل ۱. نمونه‌هایی از آیتم SCT

(a) نوع قضاوت: تشخیصی

خانمی ۵۸ ساله با سابقه دو هفته ای سرگیجه متناوب به اورژانس مراجعه می کند. او بین عارضه ها احساس خوبی دارد.

این تشخیص می شود:					و سپس شما پیدا می کنید:	اگر فکر می کنید که:
-۲	-۱	۰	+۱	+۲	عارضه سرگیجه ۳۰ دقیقه طول می کشد. Episodes of vertigo last 30 minutes	Q1. سرگیجه وضعیتی حمله‌ای خوش خیم Benign paroxysmal positional vertigo
-۲	-۱	۰	+۱	+۲	سابقه فشار خون بالا History of Hypertension	Q2. حمله اسکیمی گذرا Transient ischemic attack
-۲	-۱	۰	+۱	+۲	جراحی اخیر برداشت ضایعه پوستی Recent surgical removal of skin lesion	Q3. سندروم منیر Meniere's syndrome

(b1) نوع قضاوت: بررسی

خانم ۳۳ ساله ای با سندرم تخمدان پلی کیستیک و فشار خون قبلی مرتبط با بارداری برای ارزیابی سردردهای پس از زایمان، اختلالات بینایی و پارستزی بازوها^۵ مراجعه کرده است. فشار خون او در مطب شما ۱۰۰/۱۸۰ است.

این تشخیص می شود:					و سپس شما پیدا می کنید:	اگر فکر می کنید که:
-۲	-۱	۰	+۱	+۲	سردردهای بیمار وقتی که او به صورت صاف دراز می کشد بیشتر می شود. The patient's headaches worsen when she lies flat	Q4. دستور ونوگرافی رزونانس مغناطیسی Ordering magnetic resonance venography (MRV)
-۲	-۱	۰	+۱	+۲	بیمار ۴ هفته پیش تحت زایمان طبیعی خود به خود قرار گرفت The patient underwent spontaneous vaginal delivery 4 weeks ago	Q5. دستور جمع آوری پروتئین ادرار ۲۴ ساعته Ordering a 24-hour urinary protein collection
-۲	-۱	۰	+۱	+۲	بیمار یک هفته پیش سزارین شد. The patient underwent Caesarean section 1 week ago	Q6. انجام پونکسیون کمری Performing a lumbar puncture

(C) نوع قضاوت: درمان

از شما خواسته شده است که خانم ۷۴ ساله پرفشاری خونی را ببینید که هیدروکلروتیازید و آسپرین ۸۰ میلی گرم در روز دریافت می کند که یک عارضه ۱۵ دقیقه ای از گفتار نامفهوم و کلافگی دست چپ^۶ را تجربه کرده است. داپلر کاروتید تنگی ۹۰٪ شریان کاروتید داخلی راست را نشان می دهد.

این تشخیص می شود:					و سپس شما پیدا می کنید:	اگر فکر می کنید که:
-۲	-۱	۰	+۱	+۲	تنگی ۷۰٪ شریان کاروتید داخلی چپ 70% stenosis of the left internal carotid artery	Q7. فرستادن او برای اندارترکتومی کاروتید راست Sending her for right carotid endarterectomy
-۲	-۱	۰	+۱	+۲	LDL 1.97 mmol/L (محدوده نرمال: ۲.۰۰-۳.۴۰ mmol/L) LDL 1.97 mmol/L (normal range: 2.00-3.40 mmol/L)	Q8. شروع درمان با استاتین Initiating statin therapy
-۲	-۱	۰	+۱	+۲	بیمار سابقه بیماری زخم پپتیک دارد Patient has a history of peptic ulcer disease	Q9. جایگزینی آسپرین با کلودیوگرل ۷۵ میلی گرم در روز Replacing aspirin with clopidogrel 75 mg daily

⁵ Paresthesias of the arms

⁶ slurred speech and clumsiness of the left hand

سوالات SCT برای اجتناب از داشتن پاسخ‌های «درست» یا «جماع»⁷ طراحی شده‌اند. در عوض، امتیازدهی SCT بر اساس یک روش انبوه⁸ است که تنوع مشاهده شده در پاسخ‌های متخصصان به موقعیت‌های بالینی خاص را در نظر می‌گیرد. طرح امتیازدهی SCT فرض می‌کند که برای هر سؤال، پاسخ ارائه شده توسط بیشترین تعداد اعضای پانل (یعنی پاسخ مودال)⁹ ممکن است تحت شرایط داده شده، استدلال «استاندارد طلایی» در نظر گرفته شود، در حالی که پاسخ‌های سایر اعضای پانل نشان‌دهنده تفاوت است. تفسیری که ممکن است هنوز از نظر بالینی ارزشمند و شایسته اعتبار جزئی¹⁰ باشد. بنابراین، برخلاف اکثر ابزارهای ارزیابی مرسوم، SCT از یک سیستم امتیازدهی استفاده می‌کند که واقعیت مهمی را در عمل بالینی تصدیق می‌کند: اینکه حتی پزشکان مجرب اغلب داده‌ها را تفسیر می‌کنند، قضاوت می‌کنند و به موقعیت‌های بالینی نامشخص به روش‌هایی که متفاوت هستند پاسخ می‌دهند (در محدوده قابل قبولی از عملکرد پزشکی).

زیربنای آزمون: منطق آزمون

"اسکرپیت" چیست؟

نظریه اسکرپیت که ریشه در روانشناسی شناختی دارد، توضیحی را برای چگونگی ذخیره و بازیابی اطلاعات در ذهن انسان برای تأثیرگذاری بر تفسیر افراد از اشیاء و رویدادهای جهان پیشنهاد می‌کند. تئوری اسکرپیت که در علوم پزشکی به کار می‌رود، نشان می‌دهد که دانش پزشکی در ساختارهای دانش تخصصی به نام «اسکرپیت‌ها» سازماندهی می‌شود که اطلاعات بالینی و پاتوفیزیولوژیک مرتبط را در مورد دسته بندی‌های تشخیصی وسیع (مانند بیماری‌های قلبی عروقی)، بیماری‌های خاص (مانند انفارکتوس میوکارد)، یا حتی بیماران فردی مرتبط می‌کند (به عنوان مثال مورد آقای جونز). اسکرپیت‌های پزشکی، که به آن «اسکرپیت‌های بیماری» گفته می‌شود، در اولین برخورد بالینی شروع به شکل‌گیری می‌کنند و با تجربه به‌روزرسانی، بازسازی، متناسب، هرس و پالایش می‌شوند. اسکرپیت‌های بیماری کامل را نه باید به‌عنوان مجموعه‌های گسترده‌ای از حقایق «مستقل»، بلکه به‌عنوان شبکه‌های دانش سازمان‌یافته‌ای که امکان تفسیر سریع و اقدام کارآمد را در مواجهه با مشکلات بالینی فراهم می‌کند، مفهوم‌سازی کرد.

اسکرپیت‌ها و استدلال بالینی

بر اساس تئوری اسکرپیت، در طول هر برخورد بالینی، سیگنال‌های اولیه مشتق شده از بیمار و محیط بالینی به طور خودکار مجموعه کوچکی از اسکرپیت‌های مربوط به بیماری را در ذهن پزشک فعال می‌کند. اسکرپیت‌های بیماری، انتظارات پزشک را در مورد علائم، علائم و ویژگی‌های زمینه‌ای که بیمار احتمالاً (یا نه) نشان می‌دهد، تنظیم می‌کند. انتظارات پزشک در اصل «فرضیه‌هایی» هستند که می‌توانند از طریق جمع‌آوری و تفسیر بیشتر داده‌های متمرکز ارزیابی شوند. هر بخش جدید از اطلاعات جمع‌آوری شده (مانند اطلاعات اخذ تاریخچه، یافته‌های معاینه

⁷ consensus

⁸ aggregate method

⁹ the modal answer

¹⁰ partial credit

فیزیکی، نتایج آزمایش) را می‌توان به‌عنوان حمایت‌کننده، مخالف یا بدون تأثیر بر یک فرضیه تفسیر کرد. بنابراین، تفسیر داده‌های بالینی مستلزم ارزیابی پزشک از «تناسب» بین اطلاعات بالینی مورد انتظار و واقعی با هدف پذیرش یا رد فرضیه‌های بالینی است.

جستجوی پزشک برای «تناسب» بین ویژگی‌های بالینی مورد انتظار و واقعی یک مورد می‌تواند به دو روش مکمل و اغلب تعاملی ادامه یابد: (۱) از طریق استفاده از یک نوع فرآیند استدلال «غیر تحلیلی» که متکی بر شناخت الگوهای تداعی، قضاوت‌های سریع است. و قدردانی از «گشتالت» کلی پرونده. و (۲) از طریق استفاده از یک فرآیند استدلال کندتر و «تحلیلی» که بر آزمون فرضیه آگاهانه و تفکر قیاسی متکی است. در هر لحظه در طول یک برخورد بالینی، استراتژی شناختی درگیر برای تفسیر داده‌های بالینی تا حدی به کیفیت و ارتباط اسکرپیت‌های فعال شده پزشک بستگی دارد (که به نوبه خود به دانش و تجربه قبلی پزشک بستگی دارد) و تا حدی به پیچیدگی و ابهام مشکل بالینی موجود. صرف نظر از استراتژی استدلالی که به کار گرفته شده است، زمانی که پزشک تشخیص دهد که داده‌های جمع‌آوری شده به اندازه کافی با یک اسکرپیت خاص مطابقت دارد تا وضعیت بالینی در حال آشکار شدن را توضیح دهد و/یا امکان اقدام مناسب را فراهم کند، تشخیص موقت و طرح مدیریت ظاهر می‌شود. کادر ۱ نمونه‌ای گویا از فعال‌سازی و پردازش اسکرپیت را ارائه می‌دهد.

کادر ۱. نمونه‌ای از فعال‌سازی و پردازش اسکرپیت

فرض کنید از شما خواسته شده است که بیمار مبتلا به سردرد را ارزیابی کنید. هنگامی که بیمار وارد مطب شما می‌شود، شما به سرعت – شاید حتی ناخودآگاه – متوجه می‌شوید که او یک زن جوان است که به نظر می‌رسد کمی ناراحت است. وقتی از او می‌خواهید که سردردش را توصیف کند، به شما اطلاع می‌دهد که «سمت چپ سر او را تحت تأثیر قرار می‌دهد» و «بسیار دردناک» است. این نشانه‌های زمینه‌ای اولیه، هم کلامی («سمت چپ»، «دردناک») و هم غیرکلامی (زن جوان، ظاهر ناراحتی، محیط اداری)، فوراً اسکرپیت می‌گرن شما را به یاد می‌آورد: شبکه دانش به هم پیوسته‌ای که از طریق گذشته انباشته‌اید. تجربه و یادگیری در مورد تشخیص و درمان بیماران مبتلا به می‌گرن.

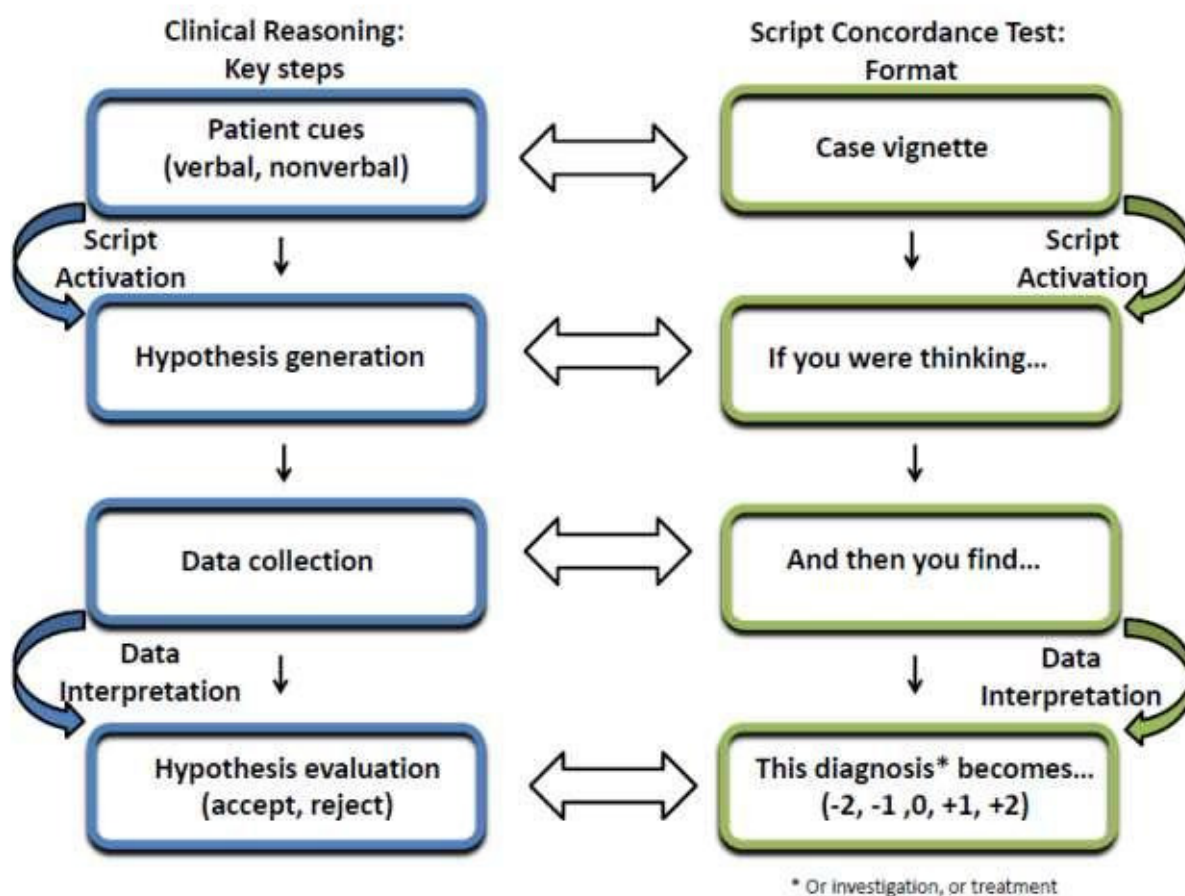
در این مورد، شما پیش‌بینی می‌کنید که بیمار «سردردهایی» را گزارش کند که «شدید» هستند (ویژگی‌های ثابت خط می‌گرن شما)، که با «تهوع» و «حساسیت به نور» (ویژگی‌های بسیار معمولی) همراه است. «یک طرفه» (ویژگی معمولی). بر اساس تجربه شما، این ویژگی‌ها همه ویژگی‌های مرتبط قوی اسکرپیت می‌گرن شما هستند و شما به راحتی می‌توانید «الگوی می‌گرن» را تشخیص دهید. فرضیه تشخیصی اولیه شما («این یک مورد می‌گرن را نشان می‌دهد») فوراً با کشف این چند قطعه اطلاعات بالینی پشتیبانی می‌شود، زیرا آنها به خوبی با انتظارات پیشینی شما در مورد نحوه ارائه بیماران مبتلا به می‌گرن مطابقت دارند.

با این حال، یافته‌های غیرمنتظره «تب» و «سفتی گردن» به‌طور خودکار باعث بسیج یک ساختار دانش جایگزین در ذهن شما می‌شود – اسکرپیت مننژیت شما. داده‌های بالینی در اختیار شما اکنون باید در پرتو حداقل دو متن رقیب تفسیر شوند. در مواجهه با این مشکل بالینی، به جمع‌آوری و سنجش اطلاعات ادامه می‌دهید تا زمانی که قضاوت کنید ویژگی‌های مورد واقعی با ویژگی‌های یکی از این اسکرپیت‌ها به اندازه کافی مطابقت دارد تا شما را قادر به انجام تحقیقات، مداخلات درمانی و مشاوره مناسب کند. هنگامی که بیمار بعدی شما وارد اتاق می‌شود، اسکرپیت‌های فعال قبلی شما از حافظه کاری حذف می‌شوند و اسکرپیت‌هایی که مربوط به مورد جدید هستند بلافاصله در ذهن شما جاری می‌شود.

SCT چه چیزی را اندازه گیری می کند؟

همانطور که در شکل ۲ نشان داده شده است، SCT دارای سه ستون است که مربوط هستند به مراحل تولید فرضیه ("اگر فکر می کردید...")، جمع آوری داده ها ("...و سپس پیدا می کنید...") و تفسیر داده ها / ارزیابی فرضیه ("...این فرضیه می شود...")، به ترتیب، فرآیند استدلال بالینی. سناریوهای بالینی و فرضیه‌های ستون اول برای فعال‌سازی اسکرپت‌های بیماری مربوطه از پایگاه داده ذهنی آزمون شونده طراحی شده‌اند. وظیفه آزمون‌شونده تعیین میزانی است که برای هر سؤال، اطلاعات بالینی جدید ارائه شده با ویژگی‌های اسکرپتی که توسط آن سؤال ارائه شده است یا مطابقت دارد (یا نیست). تطابق اسکرپت منوط به استنباط است که آزمون شونده‌گان با اسکرپت‌های بیماری تکامل یافته‌تر، داده‌ها را تفسیر می‌کنند و در موقعیت‌های نامشخص قضاوت می‌کنند که به طور فزاینده‌ای با شرایط پزشکان با تجربه با توجه به سناریوهای بالینی یکسان مطابقت دارد، و عملکرد این مهارت‌ها را می‌توان با استفاده از مقیاس لیکرت دریافت کرد. مشاهده این که نمرات SCT به طور مداوم با افزایش سطح آموزش افزایش می‌یابد، از اعتبار این استنباط حمایت می‌کند.

شکل ۲. رابطه بین مراحل کلیدی در فرآیند استدلال بالینی و قالب آیتم‌های آزمون تطابق اسکرپت



SCT اندازه‌گیری نمی‌کند؟

SCT برای بررسی توانایی فرد آزمون‌شونده در به خاطر آوردن حقایق جداسازی‌شده و جدا شده از حافظه، به روشی که سایر آزمون‌های نوشتاری، مانند پرسش‌نامه‌های چند گزینه‌ای یا سوالات پاسخ کوتاه، معمولاً انجام می‌دهند، طراحی نشده است. این نوع آزمون‌ها منابع خوبی برای بررسی وسعت و عمق پایگاه دانش یک آزمون‌شونده (یعنی «میزان» دانش در اختیار هستند). تمرکز SCT، از سوی دیگر، کشف ساختار و سازماندهی پایگاه دانش است (یعنی چگونه حقایق در حافظه با هم مرتبط می‌شوند و برای مشکلات بالینی غنی و معتبر به کار می‌روند). تحت این پارادایم، SCT مستلزم استدلال فراتر از بازیابی محض دانش حفظ شده است: دانش واقعی مرتبط باید برای پاسخ به سوالات SCT لازم است - اما کافی نیست.

به یک معنا گمراه کننده است که ادعا کنیم SCT برای ارزیابی استدلال بالینی استفاده می‌شود. در واقع، SCT تنها به یک نتیجه خاص از فرآیند استدلال بالینی مربوط می‌شود. برای هر سوال SCT، هم یک فرضیه اولیه (ستون ۱) و هم یک قطعه جدید از اطلاعات بالینی که حاکی از فرآیند جمع‌آوری داده‌ها در طول زمان است (ستون ۲) ارائه می‌شود. بنابراین SCT توانایی فرد آزمون‌شونده را برای ایجاد فرضیه‌های مناسب یا جمع‌آوری اطلاعات پزشکی مهم در یک زمینه بالینی مشخص ارزیابی نمی‌کند. در عوض، SCT بر مرحله تفسیر داده‌ها/ارزیابی فرضیه استدلال بالینی تمرکز می‌کند، که در آن فرض می‌شود که آزمودنی در مورد تناسب داده‌های جدید با فرضیه ارائه شده تصمیم بگیرد. بنابراین SCT برای بررسی یک نشانه کلیدی در امتداد یک مسیر نظری پذیرفته شده استدلال بالینی تحت عدم قطعیت طراحی شده است.

زیربنای آزمون: منطق آزمون SCT در عمل: طراحی و ساخت آزمایشی

اصول کلی

همانند فرمت سایر آزمون‌ها، ساخت SCT با در نظر گرفتن دقیق هدف مورد نظر از آزمون (ارزیابی تکوینی، آزمون‌ها با ریسک بالا، حفظ گواهینامه، و غیره)، گروه هدف (دانشجویان، دستیاران یا کارورزان، متخصصان بهداشتی دارای مجوز، و غیره) و حوزه دانش (جراحی قفسه سینه، سالمندان، دامپزشکی، پرستاری، اخلاق و غیره) آغاز می‌شود. توسعه آزمون بعدی توسط این نگرانی‌های مهم هدایت می‌شود.

برای تقویت اعتبار محتوایی نتایج آزمون، توصیه می‌شود قبل از توسعه آیتم‌های آزمون، بلوپرینت آزمون را ایجاد کنید. بلوپرینت‌های آزمون برای اطمینان از اینکه دامنه دانش مورد نظر به طور جامع توسط سوالات آزمون پوشش داده می‌شود مفید است. به عنوان مثال، یک بلوپرینت آزمون در عصب‌شناسی عمومی ممکن است برای اطمینان از نمونه‌گیری گسترده از کمپلکس‌های مختلف علائم (ضعف کانونی، تغییرات وضعیت ذهنی، اختلال راه رفتن و غیره)، بیماری‌های خاص (صرع، مولتیپل اسکلروزیس، بیماری پارکینسون و غیره)، انواع قضاوت (تشخیص، بررسی، درمان، و غیره) و عرصه‌های پزشکی (کلینیک مراقبت‌های سرپایی، بخش اورژانس، بخش مراقبت‌های ویژه عصبی و غیره) در مورد آیتم‌های آزمون مورد استفاده قرار گیرد.

طراحی سوالات SCT در عمل: طراحی و ساخت آزمایشی

طراحی آیتم

در یک SCT، یک "آیتم" به یک سناریوی بالینی (که "آیتم"، "وینیت^{۱۱}" یا "وینیت آیتم^{۱۲}" نامیده می‌شود) و مجموعه‌ای از سوالات جایگزین شده در درون آن اشاره دارد. به عنوان مثال، شکل ۱ شامل ۳ آیتم SCT است که هر کدام با ۳ سوال همراه است. مهم است که توجه داشته باشید (و به صراحت در دستورالعمل به آزمون شوندگان ذکر کنید) که هر سؤال مرتبط با یک آیتم خاص باید مستقل از سوالات دیگر در مجموعه در نظر گرفته شود. آیتم‌های SCT را می‌توان بر اساس ملاحظات تشخیصی (شکل a1)، بررسی (شکل b1)، یا درمانی (شکل c1) طراحی کرد.

آیتم‌های SCT از تجربیات بالینی روزمره سرچشمه می‌گیرند. فورنیه و همکاران (۲۰۰۸) اتخاذ رویکرد "ویژگی‌های کلیدی" زیر را برای ساخت یک آیتم SCT پیشنهاد می‌کند (شکل a1):

۱. وضعیت بالینی رایجی را که اخیراً در عمل بالینی خود با آن مواجه شده‌اید، ثبت کنید. مثال: «زنی ۵۸ ساله با سابقه دو هفته‌ای سرگیجه متناوب به بخش اورژانس مراجعه می‌کند. او بین عارضه‌ها احساس خوبی دارد.» این مرحله محتوای وینیت آیتم را فراهم می‌کند. توجه داشته باشید که شرح آیتم باید مختصر، ناقص تعریف شده (یعنی تمام اطلاعات مورد نیاز برای حل مشکل در دسترس نیست) و واقع بینانه باشد.

۲. مشخص کنید که در این شرایط چه فرضیه‌های تشخیصی مرتبط یا گزینه‌های مدیریتی را در نظر می‌گیرید. مثال: سرگیجه موضعی حمله‌ای خوش‌خیم، حملات ایسکمیک گذرا، سندرم منیر، و غیره. این مرحله محتوای ستون ۱ ("اگر فکر می‌کردید:") را فراهم می‌کند و برای فعال‌سازی اسکرپت‌های بیماری خاص در ذهن آزمون‌شونده طراحی شده است. توجه داشته باشید که فرضیه‌های تشخیصی (یا بررسی یا درمانی) همه باید قابل قبول باشند (یعنی آزمون‌شوندگان باید احساس کنند که فرضیه‌ها در واقع ملاحظات معقولی در زمینه وینیت مورد نظر هستند).

۳. مشخص کنید که چه داده‌های بالینی ممکن است به شما کمک کند تا در این موقعیت به یک تصمیم یا مسیر اقدام مناسب برسید و چه اطلاعاتی بر استدلال شما تأثیر کمی دارد یا هیچ تأثیری ندارد. مثال: مدت زمان سرگیجه، سابقه فشار خون بالا، علائم شنوایی همراه، برداشتن اخیر ضایعه پوستی با جراحی، و غیره. این مرحله محتوای ستون ۲ را ارائه می‌کند («و سپس پیدا می‌کنید:»)، و فرآیند جمع‌آوری داده‌ها را شبیه‌سازی می‌کند. توجه داشته باشید که محتوای ابداع شده برای این ستون باید طیفی از پاسخ‌های مثبت («۲+» یا «۱+»)، منفی («۲-» یا «۱-») و خنثی («۰») را ایجاد کند در مقیاس لیکرت در سراسر آیتم‌های آزمون.

بر اساس تجربه ما، توصیه می‌کنیم که دو نویسنده مسئولیت طراحی آیتم‌ها برای SCT را بر عهده بگیرند. طراحی کنندگان آزمون باید با هدف، مخاطبان هدف و حوزه محتوای آزمون آشنا باشند. قبل از اجرای آزمون، یک پیش‌نویس اولیه باید برای ۲-۳ بازبین مستقل برای بازخورد در مورد وضوح و مرتبط بودن آیتم‌های آزمون ارسال شود. فورنیه و همکاران (۲۰۰۸) یک پرسشنامه مفید (به زبان انگلیسی) برای ارزیابان برای مشاوره تا در هنگام ارزیابی کیفیت اقلام SCT ابداع کرده‌اند.

¹¹ Vignette

¹² Case vignette

افزایش اعتبار

هدف رویکرد SCT شبیه سازی شرایط عمل پزشکی واقعی است، که در آن دوره های عمل یا خطوط تفکر در مورد مشکلات بالینی خاص به ندرت غیرقابل انکار است حتی در میان متخصصان. اگرچه وینیت ها هرگز نمی توانند پیچیدگی کامل مواجهه با بیماران واقعی را منعکس کنند، طراحان SCT تشویق می شوند آیتمی را از آیتم های نمونه ای که در عملکرد روزانه دیده می شود ایجاد کنند. مواد سمعی و بصری، از جمله بخش های ویدیویی، می توانند برای افزایش اعتبار تجربه آزمون استفاده شوند.

چند آیتم، چند سوال؟

نشان داده شده است که SCT ها با زمان تست ۶۰ تا ۹۰ دقیقه قابلیت روایی نمره کافی را به دست می آورند. مطالعاتی با استفاده از نظریه آزمون کلاسیک یا تئوری تعمیم پذیری در چندین حوزه دانش انجام شده است تا تعداد بهینه آیتم ها و سوالات در آزمون را تعیین کند. این مطالعات نشان می دهد که برای به دست آوردن قابلیت اعتبار قابل قبول (یعنی تخمین آلفای کرونباخ در محدوده ۰.۷۵-۰.۸۰)، SCT ها باید شامل تقریباً ۲۵ آیتم با ۳ سؤال در هر آیتم باشد. استفاده از ۳ سؤال در هر آیتم در SCT ناشی از نگرانی های نظری و روان سنجی است: نشان داده شده است که با توجه به محدودیت های حافظه فعال، تنها مجموعه کوچکی از فرضیه ها در ذهن پزشک در هر زمان معین فعال هستند.

مقیاس لیکرت

مقیاس های پنج نقطه ای از نوع لیکرت معمولاً در SCT استفاده می شوند، اگرچه این موضوع مورد بحث قرار گرفته است. دامنه معمولاً از ۲+ تا ۲- متغیر هستند و شامل یک نقطه خنثی (۰) هستند. با این حال، زمانی که آزمون برای استفاده به عنوان یک محرک یادگیری در نظر گرفته شده است تا به عنوان یک ابزار ارزیابی، منطقی است که از مقیاس لیکرت ۳ درجه ای (۱-، ۰، ۱+) استفاده شود. به عنوان مثال، سوالات SCT با استفاده از مقیاس های ۳ نقطه ای به عنوان سکوی پرشی برای بحث در زمینه فعالیت های حرفه ای علم پزشکی مسمتر استفاده شده است. همچنین ممکن است معقول باشد که از یک مقیاس ۳ درجه ای برای ارزیابی فراگیران تازه کار استفاده شود، که انتظار می رود اسکرپیت آنها در مراحل اولیه توسعه باشد.

برای هر آیتم SCT، توصیفگرهای حد^{۱۳} مقیاس لیکرت بر اساس نوع قضاوت مورد نیاز آزمون شونده متفاوت است: تشخیص، بررسی، یا درمان. جدول ۱ توصیفگرهای حد خاصی را برای این وظایف مختلف پیشنهاد می کند. برای تشویق به انتخاب گزینه ها در محدوده مقیاس لیکرت، توصیه می کنیم از حدهایی در حد فاصل (۲+، ۲-) استفاده کنید که بیش از حد قطعی یا صریح^{۱۴} نیستند (مثلاً حد "مطمئن یا تقریباً مطمئن"^{۱۵} به حد "کاملاً مطمئن"^{۱۶} یا به سادگی "حتمی"^{۱۷} ترجیح داده می شود).

¹³ anchors

¹⁴ categorical or unequivocal

¹⁵ certain or nearly certain

¹⁶ absolutely certain

¹⁷ certain

جدول ۱. سرعنوان های ستون توصیه شده (1a) و توصیفگرهای حد در مقیاس لیکرت (1b) برای موارد تشخیص، بررسی و درمان در SCT (اقتباس از Fournier و همکاران، ۲۰۰۸)

(a1) سرعنوان های ستون

نوع داوری	ستون ۱	ستون ۲	ستون ۳
تشخیص Diagnosis	"اگر به این فکر می کردید:" 'If you were thinking of:'	"و سپس پیدا کردید:" 'And then you find:'	"تشخیص می شود این:" 'This diagnosis becomes:'
بررسی Investigation	"اگر به این فکر می کردید:" 'If you were thinking of:'	"و سپس پیدا کردید:" 'And then you find:'	"بررسی می شود این:" 'This diagnosis becomes:'
درمان Treatment	"اگر به این فکر می کردید:" 'If you were thinking of:'	"و سپس پیدا کردید:" 'And then you find:'	"درمان می شود این:" 'This diagnosis becomes:'

(b1) توصیفگر های حد های مقیاس لیکرت

	+2	+1	0	-1	-2
تشخیص	"قطعی یا تقریباً قطعی" Certain or almost 'certain	"به احتمال زیاد" 'More likely'	"نه بیشتر و نه کمتر محتمل" Neither more 'nor less likely	"کمتر احتمال دارد" 'Less likely'	"منتفی یا تقریباً منتفی است." 'Ruled out or almost ruled out'
بررسی (با در نظر گرفتن سودمند)	"کاملاً یا تقریباً کاملاً ضروری" Completely or 'almost completely necessary	"مفیدتر" 'More useful'	"نه بیشتر و نه کمتر مفید" Neither more 'nor less useful	"کمتر مفید" 'Less useful'	"کاملاً یا تقریباً کاملاً غیر ضروری" Completely or almost 'completely unnecessary
بررسی (با در نظر گرفتن ریسک-فایده)	"کاملاً یا تقریباً به طور کامل نشان داده شده است" Completely or 'almost completely indicated	"بیشتر نشان داده شده است" More 'indicated	"نه بیشتر و نه کمتر نشان داده شده است" Neither more 'nor less indicated	"کمتر نشان داده شده است" Less 'indicated	"منع مصرف یا تقریباً منع مصرف" Contraindicated or almost 'contraindicated
درمان (با در نظر گرفتن سودمند)	"کاملاً یا تقریباً کاملاً ضروری" Completely or 'almost completely necessary	"مفیدتر" 'More useful'	"نه بیشتر و نه کمتر مفید" Neither more 'nor less useful	"کمتر مفید" 'Less useful'	"کاملاً یا تقریباً کاملاً غیر ضروری" Completely or almost 'completely unnecessary
درمان (با در نظر گرفتن ریسک-فایده)	"کاملاً یا تقریباً به طور کامل نشان داده شده است" Completely or 'almost completely indicated	"بیشتر نشان داده شده است" More 'indicated	"نه بیشتر و نه کمتر نشان داده شده است" Neither more 'nor less indicated	"کمتر نشان داده شده است" Less 'indicated	"منع مصرف یا تقریباً منع مصرف" Contraindicated or almost 'contraindicated

* همانطور که Fournier و همکاران (۲۰۰۸) اشاره کرد، حد ۰ در مقیاس لیکرت SCT به معنای جایگاهی برای داوطلبان بدون نظر واضح نیست، برخلاف حد ۰ در یک پرسشنامه که اغلب نشان دهنده پاسخ "نمی دانم" است.

تشکیل پنل مرجع

اندازه پنل

تشکیل یک پنل مرجع برای تنظیم شبکه امتیازدهی آزمون یکی از ویژگی های منحصر به فرد SCT است. گانپون و همکاران (۲۰۰۵) نشان داده‌اند که برای آزمون‌های پرمخاطره، حداقل ۱۵ عضو پنل برای به دست آوردن تخمین‌های کافی از قابلیت اطمینان نمرات مورد نیاز هستند و تنها سود حداقل با داشتن بیش از ۲۰ عضو پنل به دست می‌آید. برای آزمون‌های با ریسک پایین‌تر، تعداد اعضای پنل کمتری مورد نیاز است، اما زمانی که پنل‌ها از کمتر از ۱۰ عضو تشکیل شده باشند، ممکن است قابلیت اعتبار آزمون به خطر بیفتد.

ترکیب پنل

ترکیب پنل یک ملاحظه کلیدی در طول طراحی SCT است. به یاد بیاورید که هدف SCT مقایسه مهارت‌های استدلال آزمون‌دهی‌ها با نمایندگان «متخصص» حرفه یا آزمون‌کنندگان تخصصی است که علاقمند به مشارکت هستند. ایجاد یک پنل مرجع SCT به دلیل فقدان قابل توجه اجماع در مورد آنچه در واقع "تخصص" در یک حوزه را تشکیل می‌دهد، پیچیده است. در غیاب دستورالعمل‌های استاندارد شده و مبتنی بر شواهد برای انتخاب «متخصصان»، توصیه می‌کنیم که تصمیم‌گیری‌های انتخاب منعکس‌کننده استانداردهای تایید شده جامعه در زمینه تخصص باشد. برای مثال، معیارهای عضویت در پنل SCT ممکن است شامل گواهینامه رسمی در یک زمینه، تعداد از پیش مشخص شده سال تجربه عملی در حوزه مورد علاقه، شهرت تثبیت شده برای خبرگی بالینی خوب و غیره باشد. در برخی موارد - برای مثال، زمانی که دامنه محتوای آزمون شامل چندین رشته می‌شود - تشکیل پنل‌های مبتنی بر رشته مجزا برای تنظیم کلید پاسخ برای آیتم‌های مربوطه در آزمون قابل قبول است.

استخدام اعضای پنل

استخدام ۱۵-۲۰ عضو پنل برای شرکت در SCT کمتر از آن چیزی که به نظر می‌رسد دلپره آور است. برحسب تجربه، اعضای پنل انواع مشکلات بالینی آیتم‌های SCT را جذاب می‌دانند، زیرا این مشکلات نشان دهنده چالش‌های رایجی است که در طول تمرین روزانه با آن‌ها مواجه می‌شوند. علاوه بر این، هیچ مطالعه یا آمادگی قبلی برای تکمیل SCT مورد نیاز نیست. اگر ناشناس بودن کامل پاسخ‌هایشان تضمین شود، داوطلبان بالقوه پنل احتمال بیشتری برای شرکت دارند. پس از انتخاب، از اعضای پنل خواسته می‌شود تا آزمون را به طور مستقل، در شرایط راحتی، تحت محدودیت‌های زمانی که بر آزمون‌شوندگان تحمیل می‌شود، تکمیل کنند.

اجرای آزمون

SCT ها را می‌توان برای انجام روی کاغذ یا آنلاین طراحی کرد. SCTها حاوی ۶۰ تا ۹۰ سوال (در ۲۰ تا ۲۵ آیتم برای پایایی بهینه قرار گرفته‌اند) را می‌توان در حدود ۱ ساعت تکمیل کرد. از آنجایی که قالب آزمون ممکن است برای بسیاری از آزمون‌دهندگان ناآشنا باشد، آزمون‌های SCT باید با دستورالعمل‌های واضح و چند مثال عملی شروع شوند. برای سهولت در پردازش، می‌توان برگه‌های امتیاز آزمون‌های کاغذی را به صورت نوری خواند.

آنالیز POST HOC آزمون

یکی از راه‌های تأیید کیفیت مجموعه‌ای از سؤالات SCT، اجرای آزمایشی آن بر روی یک گروه از پاسخ‌دهندگان، ارزیابی ویژگی‌های روان‌سنجی آن، و سپس اجرای یک نسخه بهینه از آزمون برای گروه هدف دیگری از پاسخ‌دهندگان است. با این حال، این استراتژی با مشاهده این که معیارهای دشواری آزمون و تمایز آیت‌ها می‌تواند بسته به مشخصات پاسخ‌دهندگان آزمایش شده به طور قابل توجهی متفاوت باشد، محدود می‌شود. برای شناسایی آیت‌های SCT با عملکرد ضعیف، ما انجام تجزیه و تحلیل آیت‌ها (پاسخ‌های ۱) اعضای پانل و ۲) آزمون شوندگان را توصیه می‌کنیم. این رویکرد تعقیبی برای تضمین کیفیت، روشی عملی و قابل توجیه برای ارزیابی دقت روان‌سنجی آزمون است.

تجزیه و تحلیل پاسخ‌های اعضای پانل

با وجود محدودیت‌ها، تغییرپذیری پاسخ در میان اعضای یک پانل مرجع SCT یک عامل کلیدی تعیین‌کننده قدرت افتراق دهنده SCT است. سؤالاتی که باعث ایجاد اتفاق آرا در میان اعضای پانل می‌شوند با سؤالات چند گزینه‌ای تک پاسخ صحیح تفاوتی ندارند (شکل 3a)، و سؤالاتی که توزیع بسیار گسترده‌ای از پاسخ‌ها را بدست می‌آورند ممکن است بیش از حد مبهم باشند (شکل 3b). از دیدگاه روان‌سنجی، سؤالات SCT ایده‌آل آن‌هایی هستند که طیفی از پاسخ‌های متخصص را که حول یک پاسخ مودال خوشه‌بندی شده‌اند تولید می‌کنند (شکل 3c). بنابراین سؤالات SCT با کیفیت بالا را می‌توان به راحتی و به طور عینی شناسایی کرد.

برخی از SCT ها حاوی سؤالاتی هستند که پاسخ‌های h یک یا چند نفر از اعضای پانل را برمی‌انگیزند (شکل 3d). یک مطالعه اخیر تأثیر حذف پاسخ‌های کلیدی امتیازدهی آزمون را که بیش از ۲ موقعیت pn از پاسخ مودال فاصله دارد، بر روان‌سنجی آزمون بررسی کرد^{۱۸}. برای پانل‌های متشکل از بیش از ۱۵ عضو، به استثنای این پاسخ‌های "انحرافی" تأثیر قابل توجهی بر قابلیت اطمینان نتایج آزمون نداشت. به طور مشابه، حذف کامل پاسخ‌دهندگان با نمره کل آزمون نابجا از پانل مرجع (یعنی کسانی که نمره کل آزمون بیش از ۲ انحراف استاندارد از میانگین پانل دارند) بر ویژگی‌های روان‌سنجی آزمون تأثیری نداشت. بنابراین تصور می‌شود که خطای اندازه‌گیری ناشی از پانل‌های انحرافی یا پاسخ‌های انحرافی در یک SCT ناچیز است، مشروط بر اینکه اندازه پانل به اندازه کافی بزرگ باشد.

شکل ۳. بررسی تغییرپذیری پاسخ یک سوال SCT با استفاده از پانل ۱۵ نفره.

3a) نمونه‌ای از اتفاق آرا پاسخ‌ها در پانل. عملکرد این آیت‌ها متفاوت از یک MCQ تک پاسخ صحیح نیست.

3b) نمونه‌ای از واگرایی یکنواخت پاسخ‌ها در پانل. این مورد اعتبار صوری ضعیفی دارد و بدون قدرت افتراق است.

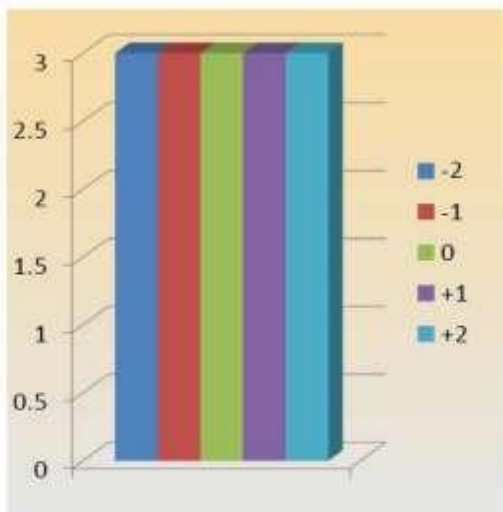
3c) نمونه‌ای از تنوع ایده‌آل پاسخ‌ها در پانل. تغییرپذیری پاسخ یک عامل تعیین‌کننده قدرت افتراق SCT است.

3d) مثالی از پاسخ «انحرافی» (در این مورد، پاسخ -۲). حذف چنین پاسخ‌هایی از کلید پاسخ احتمالاً بر قابلیت اطمینان نمره تأثیر نمی‌گذارد.

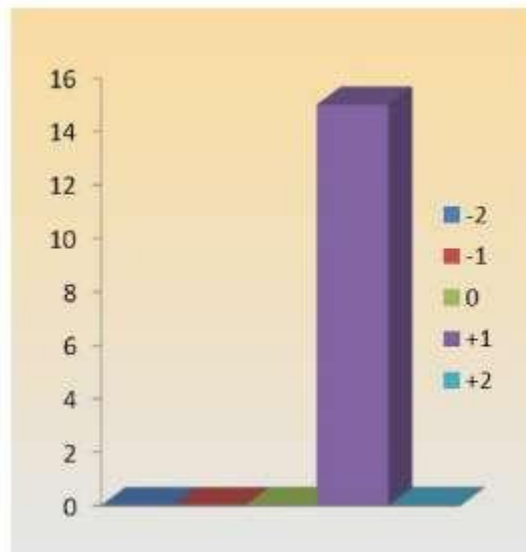
^{۱۸} به عنوان مثال، اگر برای یک سوال معین، پاسخ مودال «+۱» بود، پاسخ‌های «-۲» از کلید پاسخ حذف می‌شوند.



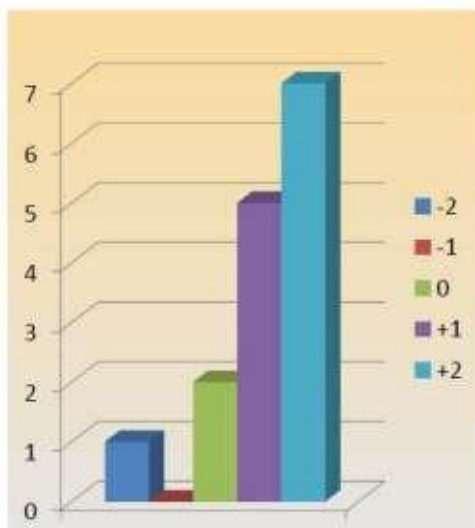
3b) نمونه ای از واگرایی یکنواخت پاسخ ها در پانل



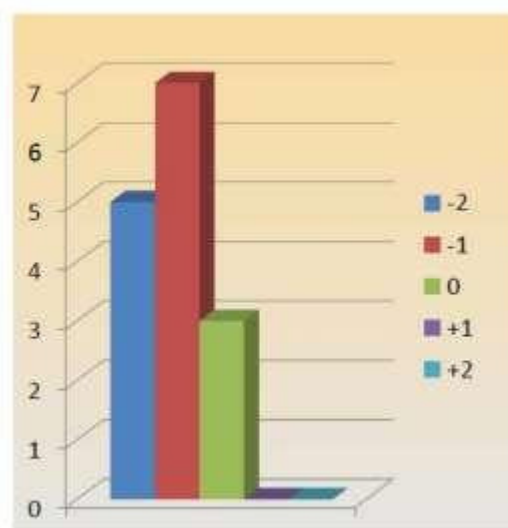
3a) نمونه ای از اتفاق آرا پاسخ ها در پانل



3d) مثالی از پاسخ «انحرافی» (در این مورد، پاسخ -۲)



3c) نمونه ای از تنوع ایده آل پاسخ ها در پانل



تجزیه و تحلیل پاسخ های آزمون شوندگان

پاسخ های آزمون شوندگان به سؤالات آزمون را نیز می توان برای شناسایی آیتم هایی که عملکرد ضعیفی دارند مورد بررسی قرار داد. محاسبه همبستگی آیتم-کل، تخمینی از ظرفیت تمایز یک آیتم، در این مرحله برای علامت گذاری آیتم های مشکل دار استفاده می شود. به طور کلی، آیتم هایی که همبستگی منفی با کل آیتم دارند، و همچنین آیتم هایی که همبستگی مثبت و کمتر از ۰.۰۵ را دارند، به میزان حداقل یا اصلاً به پایایی نمرات آزمون کمک نمی کنند و می توان آنها را حذف کرد. با این حال، همبستگی ضعیف آیتم-کل ممکن است به جای نقص در خود آیتم، به سادگی نشان دهنده ناهمگونی صلاحیت بالینی اعضای پانل یا حوزه آزمایش شده باشد. در این مواقع، آزمون سازان باید بهترین قضاوت خود را برای تصمیم گیری در مورد کنار گذاشتن یا عدم کنار گذاشتن آیتم به کار گیرند.

ایجاد نسخه نهایی

در مطالعات منتشر شده SCT، بررسی پاسخ های پانل و آزمون شوندگان با استفاده از راهبردهای ذکر شده در بالا منجر به حذف پس از آن تا یک چهارم موارد آزمون شده است. اما طراحان آزمون باید مراقب باشند: تلاش‌ها برای بهینه‌سازی قابلیت اطمینان نمره با حذف سؤالات «از نظر روان‌سنجی نامطلوب» ریسک به خطر انداختن اعتبار محتوایی آزمون را به همراه دارد. در طول بررسی آیتم‌ها، قضاوت درستی از سوی آزمون دهندگان باید اعمال شود تا تعادل قابل دفاعی بین این دو همتای روان‌سنجی مهم ایجاد شود. برای کاهش تنش بین قابلیت اطمینان و اعتبار محتوا، به طراحان SCT توصیه می‌شود که مکملی از سؤالات آزمون را به اندازه کافی بزرگ^{۱۹} ایجاد و مدیریت کنند تا دامنه محتوای مورد نظر را پوشش دهد، ضمن اینکه در صورت لزوم، امکان حذف موارد خاص از نسخه نهایی آزمون را فراهم می‌کند.

نمره‌دهی به آزمون

کلید امتیاز

با استفاده از نسخه "بهینه" آزمون، اعتبار به هر پاسخ بر اساس تعداد متخصصانی که در پانل آن پاسخ را انتخاب کرده‌اند، اختصاص می‌یابد. حداکثر امتیاز ۱ برای پاسخ انتخاب شده توسط اکثر متخصصان (یعنی پاسخ مودال) داده می‌شود. به سایر پاسخ‌ها بسته به کسری از متخصصانی که آنها را انتخاب می‌کنند اعتبار جزئی داده می‌شود. پاسخ‌هایی که توسط متخصصان انتخاب نشده‌اند صفر دریافت می‌کنند. یک نمونه از سیستم امتیازدهی SCT در شکل ۴ نشان داده شده است.

شکل ۴. سیستم نمره‌دهی

فرض کنید از پانلی متشکل از ۱۵ عضو خواسته شد که به سوال اول در مثال ارائه شده در شکل ۱ پاسخ دهند، و ۸ عضو پاسخ +۱ را انتخاب کردند، ۵ عضو پاسخ +۲ را انتخاب کردند و ۲ عضو پاسخ ۰ را انتخاب کردند. امتیاز برای این آیتم به این شکل خواهد بود. اگر پاسخ ۰ باشد، ۰.۲۵ امتیاز (۲/۸)؛ پاسخ +۱، ۱ امتیاز (۸/۸)؛ پاسخ +۲، ۰.۶۲۵ امتیاز (۵/۸)؛ پاسخ -۱ و -۲، هر دو ۰ امتیاز. نمره کل آزمون شونده برای آزمون، مجموع امتیاز کسب شده برای هر یک از سؤالات، تقسیم بر کل امتیاز قابل کسب آزمون، و ضرب در ۱۰۰ برای به دست آوردن درصد نمره است.

پاسخ	+2	+1	0	-1	-2
تعداد اعضای پانل که این پاسخ را انتخاب کردند.	5	8	2	0	0
تعداد اعضای پانل که این پاسخ را انتخاب کردند تقسیم بر پاسخ ارائه شده بوسیله بیشترین تعداد اعضای پانل (یعنی پاسخ مودال)	5/8	8/8	2/8	0/8	0/8
امتیاز برای این سوال	0.625	1	0.25	0	0

^{۱۹} به عنوان یک قانون سرانگشتی، طراحان آزمون باید سعی کنند یک و نیم (۱۵۰٪) برابر تعداد سؤالاتی را که قصد دارند در نسخه نهایی «بهینه» آزمون خود استفاده کنند (یعنی حدود ۹۰ تا ۱۲۰ سؤال) طراحی نمایند.

تنظیمات استاندارد

برای هر روش ارزیابی، تعیین نمره برش قبولی/شکست ۲۰ امری قراردادی است که باید هدف آزمون را در نظر بگیرد. برای SCT، پیشنهاد شده است که نمره قبولی باید در رابطه با عملکرد متخصصانی که در همان آزمون شرکت کرده‌اند، تعیین شود. به عنوان مثال در $2SD$ - از میانگین امتیاز پانل مرجع خبرگان. برای آزمون شوندگان در سطوح اولیه آموزش (مانند دانشجویان پزشکی پیش بالینی)، ممکن است معقول باشد که نمره قبولی را بیشتر به چپ منتقل کنند، به عنوان مثال. نمره قبولی را از میانگین امتیاز خبرگان $4SD$ -تعیین کنید. یک رویکرد جایگزین استفاده از یک پانل مرجع متفاوت برای دانشجویان پزشکی است (به عنوان مثال، گروهی از دانشجویان سال ششم ممکن است پانل مرجع ترجیحی برای دانشجویان سال پنجم باشد).

انتقاد منطقی از استفاده از رویکرد فوق برای هدف تنظیم استاندارد در SCT این است که نمره قبولی شایستگی را به روشی واضح و مشخص منعکس نمی‌کند. یکی از راه‌های حل این مسئله محاسبه امتیاز مرزی با استفاده از روش ندلسکی (۱۹۵۴) است. برای هر سوال، اعضای پانل مرجع پاسخ خود را از محدوده ۵ لنگر لیکرت موجود (۲- تا ۲+) انتخاب کرده‌اند. طبق روش ندلسکی، یک نامزد مرزی با اطمینان هر گزینه ای را که توسط هیچ یک از اعضای پانل مرجع انتخاب نشده بود حذف می‌کند و به طور تصادفی از بین گزینه‌های باقی مانده انتخاب می‌کند. تحقیقات بیشتر در مورد این و سایر روشهای تنظیم استاندارد در حال انجام است.

نتیجه‌گیری

ابزارهای استاندارد کمی برای استفاده در آموزش علوم پزشکی برای ارزیابی شایستگی استدلال بالینی موجود است. آزمون‌های موجود، مانند آزمون‌های شفاهی مبتنی بر کیس طولانی و آزمون‌های OSCE، اغلب منابع محدود دارند، در اجرا یا نمره‌دهی دست و پاگیر هستند، و یا استانداردهای آن‌ها دشوار است. آزمون تطابق اسکریپت (SCT) در تلاشی برای رفع این کمبود ایجاد شد. هدف آن بررسی یک جزء خاص از فرآیند استدلال بالینی است: توانایی تفسیر داده‌های بالینی، به ویژه در شرایط عدم قطعیت که در آن استدلال اغلب در محیط بالینی رخ می‌دهد. شواهد اعتبار و قابلیت اطمینان مربوط به SCT در حال افزایش است و SCTها ثابت کرده‌اند که ایجاد، مدیریت و امتیاز دهی نسبتاً آسانی دارند. اساتید علوم پزشکی ممکن است بخواهند SCTها را در برنامه‌های ارزیابی خود به عنوان مکمل‌های مفید برای سایر اقدامات سنتی در نظر بگیرند. ما امیدواریم که این راهنما برای کسانی که مایل به انجام این کار هستند یک منطق و یک نقشه راه ارائه دهد.