

دستورالعمل تحلیل آزمون‌ها

تحلیل آزمون‌ها (Test Analysis) فرآیندی است که برای ارزیابی کیفیت، دقت و تأثیر یک آزمون (تحصیلی، روان‌سنجی، استخدامی، پزشکی و ...) انجام می‌شود. هدف اصلی این تحلیل، اطمینان از این است که آزمون به طور منصفانه و دقیق، آنچه را که قرار است اندازه‌گیری کند، بسنجد و نتایج آن معتبر و قابل اعتماد باشد.

انواع اصلی تحلیل آزمون‌ها:

۱. تحلیل روان‌سنجی: (Psychometric Analysis)

○ **پایایی (Reliability):** بررسی ثبات و سازگاری نتایج آزمون. آیا اگر آزمون در شرایط مشابه تکرار شود، نتایج مشابهی به دست می‌آید؟ روش‌های محاسبه پایایی شامل آلفای کرونباخ، بازآزمایی، روش دو نیمه کردن و ... است.

○ **روایی (Validity):** بررسی این که آیا آزمون واقعاً آن چیزی را که ادعا می‌کند می‌سنجد، می‌سنجد؟ انواع روایی شامل:

▪ **روایی محتوا (Content Validity):** آیا سوالات محتوای حوزه مورد سنجش را به خوبی پوشش می‌دهند؟

▪ رعایت شدن قواعد ساختاری سوالات و رعایت قواعد نگارشی آزمون

○ **تحلیل سوالات (Item Analysis):**

▪ **شاخص دشواری (Difficulty Index):** درصد پاسخ‌دهندگان که به سوال پاسخ صحیح داده‌اند .

▪ **شاخص تمیز (Discrimination Index):** توانایی سوال در تمایز قائل شدن بین افراد قوی و ضعیف در سازه

مورد سنجش . معمولاً با مقایسه نمره سوال در گروه برتر و ضعیف (بر اساس نمره کل) محاسبه می‌شود.

▪ **عملکرد گزینه‌ها (Distractor Analysis):** بررسی جذابیت گزینه‌های نادرست. آیا گزینه‌های نادرست توسط

افراد ضعیف‌تر بیشتر انتخاب می‌شوند؟ آیا گزینه‌های بی‌اثر وجود دارند؟

۲. تحلیل آماری: (Statistical Analysis)

○ **توزیع نمرات:** بررسی شکل توزیع نمرات (نرمال، چولگی). این به انتخاب روش‌های آماری مناسب کمک می‌کند.

○ **شاخص‌های مرکزی و پراکندگی:** محاسبه میانگین، میانه، نما، دامنه، واریانس و انحراف معیار نمرات.

○ **نرم‌سازی (Norming):** ایجاد هنجارها (Norms) برای مقایسه نمره یک فرد با گروه مرجع (همسالان، همکاران و ...).

مراحل کلی تحلیل آزمون:

۱. **تعریف هدف:** هدف آزمون چیست؟ (گزینش، تشخیص، ارزشیابی یادگیری، پژوهش و ...)
۲. **انجام تحلیل‌های روان‌سنجی و آماری:** محاسبه پایایی، روایی، شاخص‌های سوالات، توزیع نمرات و ...
۳. **تفسیر نتایج:** معنادار کردن اعداد و شاخص‌ها.
 - آیا پایایی قابل قبول است؟ (مثلاً آلفای کرونباخ بالای ۰,۷ یا ۰,۸)
 - آیا روایی مورد نظر برقرار است؟
 - دشواری سوالات مناسب است؟ (معمولاً سوالات با P بین ۰,۳ تا ۰,۷ مطلوب‌ترند)
 - سوالات قدرت تمیز خوبی دارند؟ (معمولاً شاخص تمیز بالای ۰,۳ مطلوب است)
 - آیا گزینه‌های نادرست به درستی عمل می‌کنند؟
 - آیا توزیع نمرات مورد انتظار است؟
 - آیا شواهدی از سوگیری (DIF) وجود دارد؟

۴. **گزارش‌نویسی:** مستندسازی یافته‌ها، نقاط قوت و ضعف آزمون و پیشنهادات برای اصلاح.
۵. **بازنگری و اصلاح آزمون:** حذف یا اصلاح سوالات ضعیف، افزودن سوالات جدید، اصلاح دستورالعمل‌ها و ... بر اساس نتایج تحلیل.

اهمیت تحلیل آزمون‌ها:

- **افزایش اعتبار و اعتماد:** نتایج معتبر و قابل اعتماد برای تصمیم‌گیری‌های مهم (گزینش، تشخیص، ارتقاء و ...).
- **اصلاح و بهبود آزمون:** شناسایی سوالات ضعیف، مبهم یا ناعادلانه برای اصلاح در نسخه‌های بعدی.
- **انصاف (Fairness):** اطمینان از عدم سوگیری آزمون علیه گروه‌های خاص.
- **کارایی:** ساخت آزمون‌های کوتاه‌تر بدون کاهش کیفیت از طریق حذف سوالات بی‌اثر.
- **تضمین کیفیت:** بخشی ضروری از فرآیند توسعه و اجرای حرفه‌ای هر آزمون.

سخن پایانی:

تحلیل آزمون یک فرآیند فنی و تخصصی است که نیاز به دانش آمار و روان‌سنجی دارد. انجام دقیق آن برای هر آزمونی که نتایج آن در تصمیم‌گیری‌های مهم تأثیرگذار است، ضروری می‌باشد. بدون تحلیل مناسب، نمی‌توان به دقت، انصاف و اعتبار نتایج یک آزمون اطمینان داشت.

معیار سطح بندی کیفیت سؤالات

جدول ۱-۳: سطح بندی کیفیت سؤال بر اساس ضریب دشواری (۱)

میزان	تعریف	تفسیر
۰/۷ - ۰/۳	خوب	این سؤالات برای گروهی از داوطلبان چالش برانگیزتر محسوب می‌شوند و نیاز به تفکر عمیق‌تر و درک پیشرفته‌تری دارند. ممکن است برخی از شرکت کنندگان به دلیل دشواری این سؤالات نتوانند پاسخ مناسب ارائه دهند.
بالاتر از ۰/۷	آسان	سؤالاتی که به راحتی قابل پاسخ‌دهی هستند. این سؤالات معمولاً مفاهیم بنیادی یا ابتدایی را مورد سنجش قرار می‌دهند و اکثر شرکت کنندگان، از جمله کسانی که دانش کمی دارند، می‌توانند به آنها پاسخ دهند.
کمتر از ۰/۳	سخت	سؤالاتی که دشواری زیادی دارند و تنها تعداد کمی از داوطلبان قادر به پاسخ‌دهی صحیح به آنها هستند. معمولاً این نوع سؤالات به وضعیت‌های پیچیده، محاسبات پیشرفته یا جزئیات خاص می‌پردازند و برای سنجش عمق دانش یا توانایی‌های تحلیلی بسیار حائز اهمیت هستند.

جدول ۲-۳: سطح بندی کیفیت سؤال بر اساس ضریب تمیز (۱)

دامنه	تعریف	تفسیر
≥ 0.4	عالی	سؤالات در این دسته به طور مؤثر تفاوت بین دانشجویان با عملکرد بالا و پایین را نشان می‌دهند. این سؤالات عموماً به درستی دانش یا مهارت‌های لازم را اندازه‌گیری می‌کنند و افراد موفق در امتحان به راحتی به این سؤالات پاسخ می‌دهند.
۰/۳۹ - ۰/۳	خوب	این سؤالات هم چنان تمایز خوبی بین دانشجویان با عملکرد متفاوت دارند، اما ممکن است در برخی موارد به اندازه کافی قوی نباشند یا به نوعی برای برخی از سؤالات گیج‌کننده باشند. با این حال، این سؤالات نوعاً به عنوان سؤالات معتبر در نظر گرفته می‌شوند.
۰/۲۹ - ۰/۱۱	متوسط	سؤالات این دسته می‌توانند نیاز به بازنگری داشته باشند. آنها ممکن است نتوانند به خوبی تمایز بین دانشجویان با عملکرد بالا و پایین را نشان دهند. به عبارت دیگر، افراد با درک متوسط ممکن است به طور مشابه به این سؤالات پاسخ دهند و نشان‌دهنده توانایی‌های مختلف نباشند.
۰ - ۰/۱	ضعیف	سؤالات در این دسته معمولاً نمی‌توانند تمایز کافی بین دانشجویان ایجاد کنند. این سؤالات ممکن است هم برای افراد با عملکرد بالا و هم برای افراد با عملکرد پایین آسان یا دشوار باشند و مستلزم تجدید نظر یا حذف هستند.

سوالات منفی نشان می‌دهند که پاسخ‌دهندگان با عملکرد بالا به طور نامناسبی پاسخ داده‌اند، در حالی که پاسخ‌دهندگان با عملکرد پایین به طور نادرست به این سوالات پاسخ مثبت داده‌اند. این سوالات به شدت نیاز به بازنگری دارند، زیرا ممکن است نشان‌دهنده مفاهیم یا رویکردهای غیرمعتبر باشند.	منفی	۰ - ۱-
--	------	--------

جدول ۳-۳: سطح بندی کیفیت سؤال بر اساس ضریب همبستگی کودر- ریچاردسون (KR۲۰)(۲)

میزان	تعریف	تفسیر
۰ - ۰/۱	ارتباط ناچیز	نشان‌دهنده این است که سوالات به هم مربوط نیستند و ممکن است نتایج ناهمخوانی به همراه داشته باشند.
۰/۱ - ۰/۳۹	ارتباط ضعیف	این نمره نشان می‌دهد که سوالات به گونه‌ای ناچیز مرتبط هستند، ولی هنوز هم ارتباط قوی بین آنها وجود ندارد.
۰/۴ - ۰/۶۹	ارتباط متوسط	در این محدوده، سوالات می‌توانند به اندازه‌ای از یکدیگر حمایت کنند که نشان‌دهنده اندازه‌گیری یک سازه خاص باشد.
۰/۷ - ۰/۸۹	ارتباط قوی	در این سطح، سوالات به وضوح با یکدیگر مرتبط‌اند و از دقت بیشتری برخوردارند.
۰/۹ - ۱	ارتباط بسیار قوی	در این نمره، سوالات به طرز چشمگیری به هم مرتبط‌اند و این نشان‌دهنده همسانی بالا بین آنهاست.

جدول ۳-۴: سطح بندی نوع تاکسونومی سوالات (۳)

میزان	تعریف
تاکسونومی یادآوری و یا بازشناسی (سطح ۱ حیطه شناختی: دانش)	تاکسونومی سطح ۱
تاکسونومی تفسیر اطلاعات (سطوح ۲ و ۳ حیطه شناختی: درک و کاربرد)	تاکسونومی سطح ۲
تاکسونومی حل مساله (سطوح ۴، ۵ و ۶ حیطه شناختی: تجزیه و تحلیل، ترکیب و ارزشیابی)	تاکسونومی سطح ۳

رفرانس ها :

۱. Saif, A. A. (۲۰۱۷). Educational measurement, assessment and evaluation. ۷nd ed. Duran, Tehran
۲. Wong Shao Yun, V., Md Ulang, N., & Husain, S. H. (2023). Measuring the internal consistency and reliability of the hierarchy of controls in preventing infectious diseases on construction sites: The Kuder-Richardson (KR20) and Cronbach's Alpha. Journal of Advanced Research in Applied Sciences and Engineering Technology, 33(1), 392-405.

۳. جلیلی، محمد؛ خباز مافی نژاد، محبوبه؛ گندم کار، رقیه و مرتاض هجری، سارا. (۱۳۹۶). اصول و روش های ارزیابی فراگیران در

علوم پزشکی. فرهنگستان علوم پزشکی جمهوری اسلامی ایران، چاپ اول ..