تجزیه و تحلیل آزمون دروس دانشجویان کارشناسی بهداشت با استفاده از برنامه نرم‌افزاری آنالیز سؤالات چند گزینه‌ای بر اساس نظریه تستی کلاسیک (CTT)

محمد مهدی امین*، شهرام شایان، حسن هاشمی، پریناز پورصفا، افشین ابراهیمی

چکیده

مقدمه: هدف از انجام این مطالعه، استفاده از نرم‌افزار نظریه تستی کلاسیک برای تجزیه و تحلیل کمی و کیفی سؤالات چند گزینه‌ای برای دانشجویان مقطع کارشناسی رشته بهداشت می‌باشد.

روش‌ها: این مطالعه بر روی دو گروه ۵۰ نفری از دانشجویان کارشناسی رشته بهداشت عمومی و دو گروه ۵۰ و ۲۶ نفری از دانشجویان کارشناسی رشته بهداشت حرفه‌ای شرکت کننده در امتحانات تعیین پیشرفت تحصیلی به صورت آزمون چند گزینه‌ای پایان ترم در سال‌های ۸۵- ۱۳۸۴ در گروه بهداشت محیط، دانشکده بهداشت، دانشگاه علوم پزشکی اصفهان اجرا گردید. در دو مرحله، قبل و بعد از آنالیز کمی، سؤالات توسط برنامه نرم‌افزاری «نظریه تستی کلاسیک» مورد تجزیه و تحلیل قرار گرفت.

نتایج: با انجام آنالیز کیفی، مجموعه سؤالات مرحله اول اصلاح گردید. در سال بعد، گروه دیگری از دانشجویان با سؤالات اصلاح شده مرحله اول، مورد آزمون قرار گرفتند. ضریب پایایی برای آزمون زباله‌های صنعتی و آزمون بهداشت آب قبل از آنالیز کمی به ترتیب به میزان ۰/۵۷۶ و ۰/۳۵۹ به دست آمد. بعد از آنالیز کمی سؤالات، این ضرایب به ترتیب به مقادیر ۰/۶۷۱ و ۰/۶۹۴ بهبود یافت.

نتیجه‌گیری: بهره‌گیری از روش تجزیه و تحلیل سؤالات در این مطالعه منجر به بهبود ضریب پایایی حاصل شده در مرحله دوم گردید. این ضریب هنوز تا حد مطلوب (ضریب پایایی بزرگتر از ۰/۸) فاصله دارد. بنابراین، لازم است این سؤالات باز هم اصلاح شده، و با استفاده از نرم‌افزار نظریه تستی کلاسیک (**CTT**) مورد آنالیز مجدد قرار گیرد.

واژه‌های کلیدی: تجزیه و تحلیل آزمون، سؤالات چند گزینه‌ای، نظریه تستی کلاسیک

مجله ایرانی آموزش در علوم پزشکی (ویژه‌نامه توسعه آموزش) / زمستان ۱۳۸۹؛ ۱۰(۵): ۷۱۹ تا ۷۲۵

مقدمه

احتمالاً درگیر شدن در ارزشیابی دانشجو، بحرانی‌ترین وظیفه‌ای است که یک استاد با آن مواجه است. به طور کلی اساتید به شکلی جدی در این کار وارد می‌شوند ولی متأسفانه کیفیت بسیاری از روش‌های سنجش و امتحانات خیلی کمتر از حد مطلوب است. چنانچه ارزشیابی بر اساس اصول و موازین علمی صورت گیرد آن را می‌توان مهم‌ترین رکن آموزش و مؤثرترین عامل برای بهبود کیفیت یادگیری به شمار آورد(۱).

سؤالات چند گزینه‌ای معمول‌ترین نوع سؤالات برای سنجش دانش در سطوح مختلف (یادآوری، تفسیر و حل

* **نویسنده مسؤول:** محمد مهدی امین (استادیار)، مرکز تحقیقات محیط زیست، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران. amin@hlth. mui. ac. ir

شهرام شایان، کارشناس ارشد مدیریت آموزشی، دانشجوی دکترای مدیریت آموزش عالی، مرکز تحقیقات آموزش پزشکی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران؛ حسن هاشمی، دانشجوی دکترای مهندسی بهداشت محیط، مرکز تحقیقات محیط زیست، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران؛ پریناز پورصفا،کارشناس ارشد محیط زیست، مرکز تحقیقات محیط زیست، دانشگاه علوم پزشکی اصفهان؛ افشین ابراهیمی (استادیار) مرکز تحقیقات محیط زیست، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.

[Downloaded from ijme.mui.ac.ir on 2025-07-02]

مسأله) می‌باشند. گرچه بیشتر سؤالات چند گزینه‌ای معمولی برای سنجش سطح پایین دانش بکار گرفته می‌شوند ولی اگر این سؤالات خوب طرح شوند، قادرند سطح بالای دانش، فهم، ادراک، کاربرد، اطلاعات و حل مسأله را بسنجند(۲). سؤالات چند گزینه‌ای معمولاً در برنامه‌های آموزشی مورد استفاده قرار می‌گیرند اما دارای اثر ناشناخته‌ای بر روی علم دانشجویان دارند(۳).

هر سؤال چند گزینه‌ای شامل چند قسمت است: ۱) قسمت اصلی یا «تنه» یا «ساقه» (Stem) سؤال: این بخش از سؤال، متن اصلی سؤال را تشکیل می‌دهد و در بر گیرنده مسأله یا موضوعی است که سؤال باید آن را اندازه‌گیری کند. ۲) گزینه درست یا پاسخ سؤال: یکی از گزینه‌های پیشنهادی پاسخ درست سؤال است که آزمون شونده باید آن را برگزیند. به این گزینه «گزینه کلید» گفته می‌شود. ۳) گزینه‌های انحرافی: به غیر از گزینه درست سؤال، تعدادی گزینه دیگر نیز برای هر سؤال تهیه می‌شود که به آن گزینه‌های انحرافی می‌گویند. نقش گزینه‌های انحرافی منحرف کردن آزمون شدگانی است که پاسخ درست سؤال را نمی‌دانند. به طور کلی، متن اصلی یا تنه سؤال را یک جمله استفهامی یا یک جمله ناتمام تشکیل می‌دهد و گزینه‌های پیشنهادی جواب جمله استفهامی یا تکمیل‌کننده جمله ناقص هستند(٤). توانایی در نوشتن سؤالات چند گزینه‌ای خوب، نوعی هنر است(٥).

تجزیه و تحلیل سؤالات به دو روش کیفی و کمی انجام می‌گیرد. ارزیابی سؤالات باید بر اساس دو نوع پارامترهای کمی و کیفی انجام گیرد. در تجزیه و تحلیل کمی سؤالات که هدف اصلی این فرآیند می‌باشد، سؤال بر اساس سه پارامتر مورد نقد قرار می‌گیرد: ۱) سطح دشواری یا آسانی سؤال (Item Difficulty or Facility)، ۲) قدرت تمیز دادن یا تشخیص (Item Discrimination)، و ۳) گزینه‌های انحرافی (Item Distracter)

سطح دشواری (p) بیانگر درصد افرادی است که به سؤال پاسخ درست داده‌اند، و از طریق تقسیم تعداد افرادی که به آزمون جواب صحیح داده‌اند به تعداد افرادی که در آزمون شرکت کرده‌اند، به روش زیر محاسبه می‌شود.

$$P = \frac{Number\ Correctly\ Answering\ The\ Item}{Number\ Taking\ The\ Test} \times 100$$

قدرت تمیز یا تشخیص (D) نشان‌دهنده قدرت تشخیص افراد با توانایی بالا (نمره بالا) و توانایی پایین (نمره پایین) است، که به روش زیر محاسبه می‌شود.

$$D=(P\ Top\ Group)-(p\ Bottom\ Group)$$

دامنه قدرت تمیز سؤال (D) در محدوده ۱- صفر، و ۱+ می‌باشد. D برابر صفر به مفهوم آن است که سؤال فاقد قدرت تشخیص است.

هدف اصلی از استفاده از گزینه‌های انحرافی، به انحراف کشیدن افراد ضعیف است. این شاخص پس از مشخص کردن تعداد افرادی که به هر سؤال پاسخ صحیح داده‌اند، قابل محاسبه است.

ثابت پایایی (Reliability) آزمون (KR) از طریق معادله زیر محاسبه می‌شود.

$$KR = \frac{n}{n-1} \times \left(1 - \frac{\sum pq}{\delta^2}\right)$$

در این معادله n بیانگر تعداد سؤالات، و p نشان‌دهنده تعداد افرادی است که پاسخ صحیح داده‌اند.(q=1-p)(٦).

هدف از انجام این مطالعه، ارتقای کیفیت آزمون‌های پیشرفت تحصیلی مورد استفاده برای ارزشیابی دانشجو در امتحانات پایان ترم از طریق بهره‌گیری از نظریه‌های جدید در تجزیه و تحلیل سؤالات مانند نظریه تستی کلاسیک و نیز استفاده از فناوری در افزایش دقت و صحت محاسبات می‌باشد که ماحصل آن بکارگیری نرم‌افزار تجزیه و تحلیل سؤالات چندگزینه‌ای مبتنی بر نظریه تستی کلاسیک برای دانشجویان مقطع کارشناسی رشته بهداشت می‌باشد.

[Downloaded from ijme.mui.ac.ir on 2025-07-02]

تجزیه و تحلیل سؤالات به منظور کاهش منابع خطا، تعیین سؤالات مناسب، تعیین سؤالاتی که نیاز به ویرایش دارند، و تعیین شاخص افتراق بین دانشجویان قوی و ضعیف و نیز تعیین پایایی آزمون انجام می‌گیرد. طراحی سؤالاتی که توانایی دانشجویان را در کاربرد صحیح عقاید علمی اساسی، مفاهیم و اصول در پاسخ دادن به سؤالات مربوط به مسایل مطروحه در طول نیمسال تحصیلی مورد سنجش قرار می‌دهد، از اهداف دیگر تحقیق می‌باشد.

روش‌ها

این مطالعه بر روی دو گروه ۵۰ نفری از دانشجویان کارشناسی رشته بهداشت عمومی و دو گروه ۵۰ و ۲٦ نفری از دانشجویان کارشناسی رشته بهداشت حرفه‌ای شرکت‌کننده در امتحانات پیشرفت تحصیلی که به صورت آزمون چندگزینه‌ای Multiple Choice Questions: MCQ) در پایان ترم در محدوده زمانی سال‌های ۸۵-۱۳۸٤ در گروه بهداشت محیط، دانشکده بهداشت، دانشگاه علوم پزشکی اصفهان انجام شده، اجرا گردید.

در این فرآیند، نتایج آنالیز آزمون پیشرفت تحصیلی اجرا شده به روش MCQ در دو درس شامل: ۱) درس «مواد زائد جامد صنعتی (زباله‌های صنعتی)» برای دانشجویان کارشناسی بهداشت حرفه‌ای، و ۲) درس «بهداشت محیط-۱ (بهداشت آب)» برای دانشجویان کارشناسی بهداشت عمومی؛ هر کدام در دو دوره پس از اجرای آزمون، و پس از تصحیح و نمره‌دهی از طریق ماشین مارک‌خوان (O. P. Scan)، توسط برنامه نرم‌افزاری «تجزیه و تحلیل سؤالات چند گزینه‌ای بر اساس نظریه تستی کلاسیک (Classic Test Theory: CTT)» در واحد اتوماسیون امتحانات، اداره آمار و خدمات ماشینی حوزه معاونت آموزشی، دانشگاه علوم پزشکی اصفهان، در دو مرحله، قبل و بعد از آنالیز کیفی سؤالات، مورد تجزیه و تحلیل قرار گرفت.

نتایج

نتایج آنالیز کمی دو آزمون MCQ دانشجویان کارشناسی بهداشت قبل و بعد از آنالیز کمی سؤالات با استفاده از برنامه نرم‌افزاری تجزیه و تحلیل سؤالات چندگزینه‌ای بر اساس نظریه تستی کلاسیک (CTT) در جدول یک ارائه شده است.

گزارش نتایج آنالیز آزمون درس بهداشت محیط-۱ (بهداشت آب) بعد از تجزیه و تحلیل کمی سؤالات در جدول دو آورده شده است.

جدول ۱: نتایج تغییر پایایی دو آزمون MCQ دانشجویان کارشناسی بهداشت قبل و بعد از آنالیز کمی سؤالات

	ضریب پایایی (KR20- Reliability) **قبل از** تجزیه و تحلیل سؤالات	ضریب پایایی (KR20- Reliability) **بعد از** تجزیه و تحلیل کیفی سؤالات
آزمون زباله‌های صنعتی	۰.۵۷٦	۰.٦۷۱
آزمون بهداشت محیط-۱ (بهداشت آب)	۰.۳۵۹	۰.٦۹٤

[Downloaded from ijme.mui.ac.ir on 2025-07-02]

جدول ۲: گزارش نتایج آنالیز آزمون درس بهداشت محیط-۱ (بهداشت آب) بعد از تجزیه و تحلیل کمی سؤالات و میزان پایایی مرحله دوم (KR۲۰=۰/٦٩٤)

کد فایل:a. ۸۴۱۶۷۵۰ — نام درس: بهداشت محیط ا اب

تعداد دانشجو: ۴۸ — صفحه: ۲

Item	%A	%B	%C	%D	%E	%O	K	P	DIFF-INDEX	RPbis	DISC-INDI
۳۳	٤٧/٩	٨/٣	٣٥/٤	٦/٢	۰	٢	C	۰/٣٥٤	مناسب	۰/۰٢١	کم
٣٤	۰	٨٥/٤	١٢/٥	۰	۰	٢	B	۰/٨٥٣	آسان	۰/١٣٥	کم
٣٥	٩٣/٧	٢	٢	۰	۰	٢	A	۰/٩٣٦	آسان	۰/٤٢١	مناسب
٣٦	١٦/٦	۰	٦٢/٥	١٨/٧	۰	٢	C	۰/٦٢٥	مناسب	۰/١٨٦	کم
٣٧	-١٦/٦	١٠/٤	١٢/٥	٥٨/٣	۰	٢	A	۰/١٦٥	دشوار	۰/۰٥٩	کم
٣٨	٢٠/٨	٢	٧٢/٩	٢	۰	٢	C	۰/٧٢٨	آسان	۰/٣٤٧	مناسب
٣٩	٥٠	٢٥	٨/٣	١٤/٥	۰	٢	A	۰/٥	مناسب	۰/٣٧٩	مناسب
٤٠	٢٧	٤/١	٥٨/٣	٨/٣	۰	٢	C	۰/٥٨٣	مناسب	۰/٢٨٢	کم
٤١	١٦/٦	٢٩/١	٢٠/٨	٢٥	٦/٢	٢	B	۰/٢٩١	دشوار	۰/۰٧٥	کم
٤٢	٨/٣	٤٥/٨	٦/٢	٣٧/٥	۰	٢	B	۰/٤٥٨	مناسب	۰/٣۰٧	مناسب
٤٣	٥٤/١	١٤/٥	٢٧	٢	۰	٢	A	۰/٥٤١	مناسب	۰/٢٦٦	کم
٤٤	٢٧	٦/٢	١٢/٥	٥٢	۰	٢	D	۰/٥١٩	مناسب	۰/١٨٥	کم
٤٥	٢۰/٨	٢٢/٩	٣٥/٤	١٨/٧	۰	٢	C	۰/٣٥٤	مناسب	۰/١٨٥	کم
٤٦	١٦/٦	١٨/٧	٢٢/٩	٣٧/٥	۰	٤/١	D	۰/٣٧٥	مناسب	۰/٢٢٨	کم
٤٧	٢٢/٩	٣٥/٤	١٢/٥	٢٧	۰	٢	B	۰/٣٥٤	مناسب	۰/١٧٨	کم
٤٨	٣٣/٣	١٤/٥	١٨/٧	٢٩/١	۰	٤/١	A	۰/٣٣٣	مناسب	۰/١٦٥	کم
٤٩	٨/٣	٦٨/٧	٤/١	١٤/٥	۰	٤/١	B	۰/٦٨٦	مناسب	۰/٤٢٣	مناسب
٥۰	٢٧	٤/١	٣٥/٤	٢٩/١	۰	٤/١	C	۰/٣٥٤	مناسب	۰/٣۰١	مناسب

KR20 Reliability = ۰/۶۹۴۳۲۹
Standard Error = ۳/۰۹۴۰۱۳
Total Test Varianc= ۳۱/۳۱۷۷۱
Standard Deviation = ۵/۵۹۶۲۲۳
Mean P = ۰/۵۶۲۵
Mean RPBIS = ۰/۲۵۵۴۳۷۶
Total Item Varianc = ۱۰/۰۰۷۸۱
Alpha = ۰/۶۹۴۳۲۹

بحث

در مرحله اول، تعداد ۵۰ سؤال چند گزینه‌ای برای اولین بار برای دو درس مذکور طراحی گردید. سپس آزمون برگزار گردید، و پس از تصحیح و آنالیز به روشی که در بالا ذکر گردید، ضرایب پایایی مرحله اول به دست آمد (جدول ۱).

پس از دست یابی به نتایج مرحله اول گزارش ارزیابی سؤالات، نسبت به تجزیه و تحلیل کمی سؤالات اقدام شد. در گزارش نتایج آنالیز آزمون (جدول ۲)، هر یک از سؤالات بر اساس پارامترهای«ایندکس دشواری

[Downloaded from ijme.mui.ac.ir on 2025-07-02]

(Difficulty Index: DIFF. INDEX)»، «ایندکس قدرت تشخیص و افتراق آزمون شونده (Discrimination Index: DISC INDEX)» ، و چند پارامتر دیگر قابل ارزیابی است(٦). در این مطالعه نیز برای آنالیز کمی سؤالات آزمون‌های مرحله اول، با مورد توجه قرار دادن پارامترهای فوق، سؤالات دو آزمون بر اساس عواملی که باعث دشواری یا سهولت بیش از حد سؤالات و یا کاهش قدرت تشخیص می‌شوند، اصلاح گردید. برخی از این عوامل عبارتند از: ١-کلید سؤال اشتباه ثبت شده باشد، ٢- سؤال بیش از یک گزینه درست داشته باشد، ٣- متن اصلی یا تنه سؤال دارای ابهام و گنگی باشد، ٤- در گزینه درست ابهام و اشکال وجود داشته باشد، ٥- پاسخ درست آشکار باشد، ٦- سؤال بسیار دشوار بوده و در نتیجه آزمون شوندگان پاسخ را حدس زده باشند، ٧- سؤال از لحاظ محتوا با سؤال‌های دیگر همگن نباشد.

با انجام آنالیز کمی بر اساس موارد فوق الذکر، مجموعه سؤالات مرحله اول اصلاح گردید، و در سال بعد، گروه دیگری از دانشجویان همین رشته‌ها با سؤالات اصلاح شده مرحله اول، مورد آزمون قرار گرفتند (جدول ١). اگرچه مقادیر ضریب پایایی حاصل شده در مرحله دوم تا حد قابل توجهی بهبود یافته است، اما هنوز تا حد مطلوب (ضریب پایایی بزرگتر از ٠/٨) فاصله دارد. بر این اساس، لازم است این سؤالات باز هم اصلاح شده، و با استفاده از نرم‌افزار تجزیه و تحلیل سؤالات دقت و کارایی سؤالات مورد بررسی قرار گیرد.

با انجام آنالیز کمی، طراح سؤال و مدرس درس خود را از جهت گزینش یا طراحی سؤال مورد ارزیابی و داوری قرار می‌دهد. بنابراین، مدرس سعی می‌کند مرتباً سؤال طرح کند، آنها را آنالیز کند، و سؤالات خوب را جدا کند، و سؤالاتی را که نیاز به ویرایش دارند را مشخص نماید. با تجزیه و تحلیل سؤالات، مدرس و طراح سؤال می‌تواند یک بانک سؤال ارزشمند تدوین نماید، به نقاط ضعف و قوت خود و فراگیر پی ببرد. در این روش روایی و پایایی آزمون افزایش می‌یابد(٦).

از آنجا که آزمون‌های چند گزینه‌ای از سایر آزمون‌های عینی (objective) انعطاف‌پذیرتر می‌باشند (سؤالات عینی در نمره‌گذاری از اعمال نظر مصححان مصون می‌باشند). همچنین آنها علاوه بر دانش، توانایی استدلال، قضاوت، و بسیاری از بازده‌های مهم دیگر را می‌سنجند. از طرفی با استفاده از سؤال‌های چند گزینه‌ای می‌توان در یک زمان محدود، تعداد زیادی از هدف‌های آموزشی و بخش مهمی از محتوای درس را اندازه‌گیری نمود. اگر گزینه‌های انحرافی این آزمون‌ها با توجه به اشتباهات و کج فهمی‌های متداول دانشجویان تهیه شود، منبع بسیار مناسبی برای تشخیص مشکلات دانشجویان خواهد بود. نتایج آزمون‌های چند گزینه‌ای قابل تجزیه و تحلیل کردن است، و منجر به استاندارد سازی سؤالات می‌شود در نمره‌دهی آزمون عدالت رعایت می‌گردد و موجب افزایش انگیزه دانشجویان به مطالعه و یادگیری می‌شود(٤). همه این محاسن باعث گردیده است که تمایل اساتید در استفاده از این شیوه ارزیابی در آزمون‌های پیشرفت تحصیلی افزایش یابد. لذا توجه به کیفیت طراحی سؤالات چند گزینه‌ای از ضروریات نظام آموزشی می‌باشد. کیفیت یک سؤال چند گزینه‌ای 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. در طرح سؤال ابتدا باید ریشه سؤال، بعد پاسخ صحیح و در نهایت پاسخ‌های انحرافی نوشته شود(٧).

با این حال اجرای این روش مشکلاتی را نیز به همراه دارد. طراحی، تجزیه و تحلیل، اصلاح، و جایگزینی سؤالات برای مدرس و طراح سؤال وقت‌گیر و دشوار است، و نیاز به مداقه بیشتر مدرس نسبت به دیگر روش‌ها دارد. به منظور آنالیز کیفی و کمی سؤالات در مراحل مختلف انجام آزمون‌ها (قبل و بعد) و به منظور

[Downloaded from ijme.mui.ac.ir on 2025-07-02]

دست‌یابی به یک بانک سؤال با ضریب اعتبار (پایایی) بالای ۰/۸ (بالای ۸۰ درصد) لازم است مدرس با صرف زمان بیشتر علاوه بر انجام هماهنگی‌های لازم برای اجرای آزمون با استفاده از پرسشنامه‌های قابل تصحیح توسط OP-Scan، و اعمال دقت در پر کردن پرسشنامه‌ها توسط دانشجویان، با کارشناسان تعلیم و تربیت، کارشناسان طراحی و روان‌سنجی آزمون‌ها نیز مشاوره نماید.

علاوه بر این، انتخاب تعداد کم سؤالات چند گزینه‌ای موجب کاهش روایی و پایایی آزمون می‌شود، و تقلب (cheating) دانشجویان در آزمون‌های چند گزینه‌ای زیاد است. در این نوع آزمون، تمایل به طراحی سؤال در سطوح پایین یادگیری بیشتر است. نمرات به دست آمده از آزمون ممکن است دارای تورش (bias) در خصوص عواملی مثل توانایی دانشجویان در تست‌زنی یا توانایی خواندن باشد. توانایی دانشجو در حذف کردن بعضی از گزینه‌ها بدون داشتن دانش مربوطه ممکن است میزان پاسخگویی را افزون بر مقدار واقعی برآورد نماید(٦).

نتیجه‌گیری

بهره‌گیری از روش تجزیه و تحلیل کمی سؤالات در این مطالعه منجر به بهبود ضریب پایایی حاصل شده در مرحله دوم گردید. این ضریب هنوز تا حد مطلوب (ضریب پایایی بزرگتر از ۰/۸) فاصله دارد. بنابراین، لازم است این سؤالات باز هم اصلاح شده، و با استفاده از نرم‌افزار نظریه تستی کلاسیک (CTT) مورد آنالیز مجدد قرار گیرد. همچنین توجه به رعایت اصول تجزیه و تحلیل کیفی سؤالات (اصول میلمن) و نیز استفاده از جدول مشخصات آزمون (Blue print) می‌تواند باعث افزایش روایی آزمون و به دنبال آن ارتقای ضریب پایایی آزمون گردد.

منابع

1. Mahmoodi SM, Haeri A.[A Handbook for Medical Teachers]. 1st ed . Tehran: Hejrat Publication.1998. [Persian]
2 Rezai HR, Shoghi Shafagh Aria F, Dolati Y, Mohagheghi MA, Malekzadeh R. [Evaluation techniques in medicine]. 1st ed. Tehran: Amea . 1999. [Persian]
3. Henry L, Roediger III, Elizabeth J. The positive and negative consequences of multiple-choice testing. J Exp Psychol Learn Mem Cogn. 2005; 31 (5):1155-1159.
4. Saif AA. [Educational measurement, assessment and evaluation]. 4th ed .Thehran: Doran . 2007. [Persian]
5. Amin Z, Seng CY, Eng KH. Practical guide to medical student assessment. World Scientific.[cited 2011 Mar 8].Available from: http://www.worldscibooks.com/medsci/6109.html
6. Workshop on: "Student Evaluation and Methods of Question Design with emphasis on Multiple Choice Questions", Isfahan University of Medical Sciences - Education and Development Center-(EDC), 22-23 June, 2006, Isfahan, Iran.
7. Zolfaghari B, Adibi N, Derakhshanfar S, Tansaz M, Karbasi A, Niroumand P. [Academic achievement tests in medical sciences]. 1st ed. Isfahan: Modireat motaleat va amoozesh pezeshki. 2001. [Persian]

[Downloaded from ijme.mui.ac.ir on 2025-07-02]

Analysis of Multiple Choice Questions Based On Classical Test Theory

Mohammad Mehdi Amin[1], Shahram Shayan[2], Hassan Hashemi[3], Parinaz Poursafa[4], Afshin Ebrahimi[5]

Abstract

Introduction: *The aim of this study was to use Classical Test Theory (CTT) software for qualitative and quantitative analysis of Multiple Choice Questions (MCQ) for Master students of health discipline.*

Methods: *This study was carried out on two groups each including 50 undergraduate students of public health, and two groups of 50 and 26 undergraduate students of occupational health. The final exams were done to determine academic achievement of students using MCQ during the years 2004 and 2005 in the department of environmental health, school of health in Isfahan University of Medical Sciences. The questions were analyzed by CTT software in two phases before and after quantative analysis.*

Results: *Having qualitative analysis performed, the questions set for the first stage exam were modified. At the following year, another group of students were examined using those modified questions of the first stage. Reliability coefficient for two exams of industrial solid wastes and health of water, before the quantitative analysis of the questions, were obtained 0.576, and 0.359, respectively. After quantitative analysis of these questions, the coefficients values improved to 0.671 and 0.694, respectively.*

Conclusion: *The method of question analysis in this study led to improving reliability coefficient in the second stage. This coefficient is still far away from desirable coefficient (reliability coefficient greater than 0.8). Therefore, these questions still need to be modified, and analyzed again using the CTT software.*

Keywords: Exam analysis, multiple choice questions (MCQ), classical test theory (CTT).

Addresses

[1] (✉) Assistant Professor, Environment Research Center, Isfahan University of Medical Sciences, Hezarjarib St, Isfahan, Iran. E-mail: amin@hlth.mui.ac.ir

[2] PhD Student of Higher Education Management, Medical Education Research Center, Isfahan University of Medical Sciences, Isfahan, Iran. E-mail: shayan@edc.mui.ac.ir

[3] PhD Student of Environment Health, Environment Research Center, Isfahan University of Medical Sciences, Isfahan, Iran. E-mail: h2_hashemi62@yahoo.com

[4] Master in Environment, Environment Research Center, Isfahan University of Medical Sciences, Isfahan, Iran. E-mail: parinaz.poursafa@gmail.com

[5] Assistant Professor, Environment Research Center, Isfahan University of Medical Sciences, Isfahan, Iran. E-mail: a_ebrahimi@hlth.mui.ac.ir

[Downloaded from ijme.mui.ac.ir on 2025-07-02]