

آنالیز آزمون های چند گزینه ای دروس تخصصی رشته پرستاری دانشگاه علوم پزشکی گلستان

در سال ۱۳۸۸

* اکرم ثناگو، ** لیلیا جویباری، *** مریم فنبری گرجی

* استادیار گروه پرستاری بهداشت جامعه، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی گلستان

** استادیار گروه پرستاری بهداشت کودک و خانواده، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی گلستان

*** کارشناس مامایی، دفتر توسعه آموزش دانشکده پرستاری، دانشگاه علوم پزشکی گلستان

چکیده

مقدمه: آزمون های چند گزینه ای رایج ترین انواع آزمون ها در آموزش علوم پزشکی می باشند. بسیاری از آزمون های معلم ساخته با اشکالات متعددی همراه هستند. هدف از این مطالعه ارزیابی کمی و کیفی تعدادی از آزمون های چند گزینه ای دانشکده پرستاری و مامایی دانشگاه علوم پزشکی گلستان در سال ۱۳۸۸ بوده است.

روش ها: در این مطالعه توصیفی - کاربردی، کلیه سؤالات آزمون های کتبی چند گزینه ای نیم سال دوم سال تحصیلی ۸۷-۱۳۸۸ سه گروه آموزشی پرستاری کودک و خانواده و بهداشت جامعه و روانپرستاری شامل ۱۲ درس و ۵۲۳ سوال مورد بررسی قرار گرفت. سؤالات از نظر تاکسونومی بلوم و رعایت اصول ساختاری بر اساس چک لیست میلن ارزیابی شدند. برای تعیین قدرت تمیز و دشواری سؤالات آزمون و سایر شاخص های کمی از نرم افزار آنالیز آزمون استفاده گردید.

یافته ها: از ۵۲۳ سوال مورد بررسی ۹۴/۳ درصد از نظر ساختاری در حد خوب بودند. بیشترین اشکالات ساختاری مربوط به چیدمان خطی گزینه ها (۵۷/۴ درصد)، قرار نداشتن بخش اعظم اطلاعات در ساقه سؤال (۳۲/۷ درصد)، کلمات تکراری در گزینه ها (۳۲/۱ درصد) و همسنگ نبودن گزینه ها از نظر طول و ساختار لغوی (۲۹/۸ درصد) بوده است. ۹۶/۴ درصد سؤالات در سطح تاکسونومی یک «دانش و فهم» قرار داشتند. ۷۹/۸ درصد سؤالات از بایدهای یادگیری بودند. میانگین ضریب دشواری کل آزمون ها از ۰/۴۹ تا ۰/۸ متغیر بود. در مجموع سؤالات، ۲۱/۱ درصد مناسب، ۴۳/۷ درصد خیلی آسان و ۱۴/۵ درصد خیلی سخت بودند. تقریباً تمامی آزمون ها از قدرت تمیز کمی با دامنه قدرت تمیز ۰/۱۴ و بیشترین قدرت تمیز ۰/۳۲ برخوردار بودند و از ۱۲ درس دو درس قدرت تمیز مطلوب داشتند.

نتیجه گیری: اگرچه آزمون های چند گزینه ای مورد مطالعه از نظر ساختاری در وضع نسبتاً خوبی قرار داشتند اما از نظر توزیع تاکسونومی، قدرت دشواری و تمیز نیازمند اتخاذ تدابیر اصلاحی می باشند. که در این راستا برگزاری کارگاه های آموزشی ویژه اساتید، نظارت و ارزشیابی عملکرد اعضای هیات علمی از سوی مرکز مطالعات و توسعه آموزش علوم پزشکی دانشگاه می تواند بسیار موثر باشد.

واژه های کلیدی: آزمون چند گزینه ای، تاکسونومی، اصول میلن، دروس تخصصی پرستاری، آنالیز آزمون

مقدمه

چیزی را بررسی کند، از چه روشی و ابزاری استفاده نماید و حتی سوابق شخصی و تجربیات حرفه ای او همه بر اجرا و نتیجه ارزشیابی تاثیر می گذارند (۲). مهم ترین وسیله اندازه گیری پیشرفت تحصیلی، آزمون است (۳) و معمولاً در سطوح عالی از سؤالات چند گزینه ای در آزمون ها استفاده می شود (۴). این آزمون ها باید ویژگی های اساسی اعتبار قابل قبول، روایی مناسب و عملی بودن را داشته باشند. برای این منظور از سنجش پارامترهای مدل درجه دشواری سوال، قدرت تمیز سوال در تشخیص دانشجوی قوی و ضعیف، درصد پاسخ به هریک از گزینه های انحرافی و صحیح، همراه با اعتبار و خطای استاندارد استفاده می گردد (۵). آزمون های

هدف های آموزشی، روش تدریس، محتوی برنامه آموزشی و ارزشیابی را می توان به عنوان ارکان فرایند تعلیم و تربیت قلمداد کرد. بدون تردید تمامی این پارامترها برهم تاثیر متقابل دارند ولی در این میان ارزشیابی و امتحان نقش تعیین کننده و ممتاز خواهد داشت. به فرض مناسب بودن اهداف آموزشی، محتوی برنامه و موثر بودن روش تدریس چنانچه نتایج حاصل از آموزش را بطور علمی ارزشیابی نکنیم، عملاً به خطا رفته ایم (۱). تصمیمات یک ارزیاب در مورد این که چه

نویسنده مسئول: دکتر لیلیا جویباری، گرگان، دانشگاه علوم پزشکی گلستان، مرکز مطالعات و توسعه آموزش پزشکی، دفتر ارزشیابی.

lm_jouybari@yahoo.com

روش‌ها

در این مطالعه توصیفی- کاربردی، کلیه سؤالات آزمون‌های چند گزینه‌ای دروس تخصصی رشته پرستاری تدریس شده توسط گروه پرستاری بهداشت جامعه و روان و گروه کودک و خانواده در ترم دوم سال تحصیلی ۱۳۷۸-۱۳۸۸ در دانشگاه علوم پزشکی گلستان مورد بررسی قرار گرفت.

داده‌ها شامل سؤالات آزمون‌های چند گزینه‌ای دروس تخصصی رشته پرستاری تدریس شده توسط گروه پرستاری بهداشت جامعه و روان، کودک و خانواده در نیمسال اول تحصیلی ۸۷-۸۸ بود. پس از برگزاری آزمون پاسخ‌نامه تک تک دانشجویان به تفکیک دروس تخصصی مربوطه وارد برنامه نرم افزاری ویژه تجزیه و تحلیل سؤالات چند گزینه‌ای شد و ساختار سؤالات آزمون نیز با استفاده از چک لیست مبتنی بر اصول میلمن (millman) آنالیز شد. همچنین کلیه سؤالات از نظر تاکسونومی بلوم در سه سطح تاکسونومی یک (دانش و فهم)، تاکسونومی دو (کاربرد و آنالیز) و تاکسونومی سه (سنتز و ارزشیابی) مورد بررسی قرار گرفت.

پس از ورود داده‌ها به برنامه نرم افزاری تجزیه و تحلیل سؤالات چند گزینه‌ای، از طریق گزارشات ارزیابی، مشخصات روانسنجی آزمون اعم از ضریب دشواری تک تک سؤالات، قدرت تمیز هر یک از سؤالات (با استفاده از روش RPBS)، فراوانی درصدی پاسخ‌گویی دانشجویان به هر یک از گزینه‌های سؤال به منظور تحلیل گزینه‌های انحرافی و همچنین میزان Reliability کل آزمون و خطای استاندارد اندازه‌گیری (SD) تعیین و سپس مشخصه‌ها و آماره‌های فوق با توجه به استانداردهای آزمون‌های پیشرفت تحصیلی مورد بررسی و تحلیل قرار گرفت. ضریب دشواری بین ۰/۳ تا ۰/۷ مناسب و ۰ تا ۰/۳ خیلی سخت و ۰/۷ تا ۱ خیلی آسان در نظر گرفته شد. در خصوص تعیین روایی سازه نیز برای تک تک سؤالات یک شناسنامه مشتمل بر ساقه سؤال، گزینه انحرافی و کلید، طراحی شد. تک تک سؤالات از نظر شاخص‌های میلمن مورد بررسی قرار گرفت. در رابطه با روایی محتوایی نیز با مراجعه به جدول دو بعدی مورد استفاده برای طراحی سؤالات میزان روایی محتوایی کل آزمون تعیین گردید. کلیه ملاحظات اخلاقی در این پژوهش از جمله محرمانه نگه داشتن اطلاعات فردی رعایت شد.

یافته‌ها

در مجموع ۵۲۳ سؤال از آزمون‌های کتبی پایان ترم دو گروه مورد بررسی قرار گرفت. ۱۲ درس مورد بررسی ۴۱/۷ درصد سؤالات مربوط به گروه پرستاری کودک و خانواده و ۵۸/۳ درصد سؤالات مربوط به گروه بهداشت جامعه و روانپرستاری

چندگزینه‌ای به علت عینیت کامل، از پایایی بالا (حداکثر پایایی ممکن) برخوردار می‌باشند، اما روایی آنها به ویژه در مورد سطح دانش (تاکسونومی) و رعایت اصول طراحی ساختار از مسایل مورد توجه در مطالعات و پژوهش‌های آموزشی بوده است. بدیهی است اگر یک آزمون از طراحی مطلوب در زمینه انتخاب تاکسونومی سؤالات و رعایت قواعد ساختاری برخوردار نباشد واجد روایی محتوا و روایی ساختار مطلوب نبوده و در این صورت نه تنها نقش اصلی آزمون به عنوان جزء تکمیل کننده و حیاتی چرخه آموزش از دست می‌رود، بلکه این امر برانگیزه‌های فراگیران اثرات منفی بر جای گذاشته و موجب هدر رفتن تلاش اساتید و سیستم آموزشی خواهد شد، لذا بررسی کیفیت سؤالات و میزان استاندارد بودن آزمون‌ها ضروری است. مطالعات موجود در دانشگاه‌های مختلف در این زمینه نتایج متفاوتی را نشان داده است. ملاک‌های آنالیز در بخش قابل توجه‌ای از سؤالات آزمون‌های چند گزینه‌ای اشکال داشته و بیش از نیمی از سؤالات فاقد مشکلات ساختاری گزارش شده است (۶ و ۸). نتایج مطالعات تاثیر نامطلوب ناشی از مختل بودن ملاک‌های آنالیز آیتم در بخش قابل توجهی از سؤالات آزمون‌های چند گزینه‌ای و تاثیر نامطلوب آن را بر کیفیت آزمون نشان داده است (۶، ۹ و ۱۰). یک بررسی نشان داد اگر چه در اغلب آزمون‌های بالینی و برخی آزمون‌های فیزیوپاتولوژی، بخش اعظم سؤالات، درست و فاقد اشکالات بودند، اما عدم رعایت اصول ساختاری به هر میزان، روایی سازه را در آزمون مختل نموده و در این صورت کارکردهای آموزشی آزمون را تحت تاثیر قرار داده و این مساله با ایجاد زمینه برای کسب نمره با روش‌هایی غیر از تسلط علمی نظیر حدس زدن، گمراه شدن دانشجویان، نگرش دانشجویان را نسبت به آزمون با اشکال مواجه می‌سازد (۱۱).

این حقیقت که معلمان گاه در ارزشیابی یادگیری فراگیران ناکام هستند یک مشکل علمی است. در بسیاری از موارد کلیه دانشجویان ضعیف و قوی صرفنظر از میزان آمادگی، نمرات بالایی در امتحان اخذ می‌نمایند و یا آنکه از گذراندن موفقیت‌آمیز آن، باز می‌مانند (۱۲). بدون شک نوع و کیفیت امتحان مستقیماً بر روش تدریس و بطور کلی‌تر بر اعتبار معلم تاثیر می‌گذارد. لذا در تهیه و اجرای آزمون‌ها باید دقت کافی مبذول داشت تا برای تصمیم‌گیری آموزشی از ارزش مطلوبی برخوردار باشند. به همین جهت این آزمون‌ها باید ویژگی‌های مطلوب آزمون‌های میزان شده نظیر روایی، پایایی و عملی بودن را داشته باشند (۱۳). این مطالعه با هدف تجزیه و تحلیل کیفی آزمون‌های چندگزینه‌ای دروس تخصصی پرستاری انجام شد.

همچنین ۷۹/۸ درصد سوالات از بایدهای یادگیری (must learn) در سطح تاکسونومی یک بودند. مابقی سوالات در دسته بهتر است (better to learn) و خوب است یاد گرفته شود (nice to learn) قرار داشتند. برای تعیین روایی سازه آزمون های چندگزینه ای پایان ترم در این مطالعه از اصول میلمن استفاده شد. در خصوص «ساختار سوالات» اصول طراحی سوالات یا اصول میلمن شامل ۱۴ شاخص می باشد. امتیاز کسب شده بین حداقل ۱۴ و حداکثر ۲۸ محاسبه شد. امتیاز ۱۴ به معنی صفر و امتیاز ۲۸ به معنی ساختار کاملاً درست می باشد. نمرات کسب شده به سه گروه ساختار ضعیف ۱۴-۱۸، ساختار متوسط ۱۹-۲۳ و ساختار خوب ۲۴-۲۸ طبقه بندی شدند. ۵/۷ درصد ساختار متوسط و ۹۴/۳ درصد ساختار خوب داشتند. همچنین با استفاده از میانگین نمره از ۲۸ امتیاز بهترین ساختار آزمون ها مربوط به دروس گروه پرستاری کودک و خانواده بود. جدول شماره دو، درصد اشکال ساختاری آزمون های چهار گزینه ای را به طور کلی در آزمون های مورد بررسی نشان می دهد.

تحلیل کمی آزمون که با ضریب تمیز سوال (Discrimination Index)، ضریب دشواری سوال (Facility Index)، پایایی آزمون (Reliability) در جدول شماره سه تعیین شده است.

بود (جدول ۱). از نظر توزیع سطح تاکسونومی سوالات، ۹۶/۴ درصد سوالات در سطح تاکسونومی یک (دانش و فهم) و صرفاً ۳/۶ درصد سوالات در سطح تاکسونومی دو (کاربرد و آنالیز) قرار داشتند. هیچ سوالی در سطح تاکسونومی سه (سنتر و ارزشیابی) قرار نگرفت.

درس	تعداد سوال	درصد سوال
پرستاری بهداشت جامعه (۱)	۳۰	۵/۷
پرستاری بهداشت جامعه (۲)	۷۷	۱۴/۷
پرستاری بهداشت جامعه (۳)	۳۷	۷/۱
پرستاری بهداشت روان (۱)	۳۸	۷/۳
پرستاری بهداشت روان (۲)	۳۶	۶/۹
پرستاری بهداشت مادران و نوزادان (۲)	۳۰	۵/۷
پرستاری بهداشت مادران و نوزادان (۱)	۵۰	۹/۶
پرستاری کودک و خانواده (۲)	۱۰،۷	۵۶
پرستاری کودک و خانواده، دانشجویان کارشناسی ناپیوسته	۵۳	۱۰/۱
پرستاری کودک و خانواده (۱)	۴۱	۷/۸
بیماری های نوزادان (دانشجویان مامایی)	۴۶	۸/۸
پرستاری مادران و نوزادان (۲) بخش نوزادان	۲۹	۵/۵

جدول ۲: توزیع فراوانی اشکالات ساختاری آزمون های چهارگزینه ای

سوال	بلی (درصد)	خیر (درصد)
آیا بخش اعظم اطلاعات در ساقه سؤال گنجانده شده است؟	۶۷/۳	۳۲/۷
آیا سؤال یک هدف اختصاصی یادگیری را مورد ارزیابی قرار می دهد؟	۹۰/۲	۹/۸
آیا لغات استفاده شده در ساقه یا گزینه ها، شفاف و مستقیم بیان شده اند؟	۹۱/۸	۸/۲
آیا از کاربرد گزینه منفی برای ساقه منفی خودداری شده است؟	۹۹	۱
آیا از کاربرد گزینه های نظیر همه موارد هیچ کدام و گزینه های ترکیبی خودداری شده است؟	۹۳/۹	۶/۱
آیا از کاربرد گزینه های متضاد یکدیگر خودداری شده است؟	۹۷/۳	۲/۷
آیا از لغایت مثبت در ساقه سؤال استفاده شده است یا در صورت منفی بودن ساقه سؤال، لغایت منفی مشخص شده اند؟	۹۷/۱	۲/۹
آیا هر سؤال مستقل از سؤالات دیگر می باشد؟	۶۹/۹	۳/۱
آیا گزینه ها از نظر طول، ساختار لغوی و سبک نگارش همسنگ هستند؟	۷۰/۲	۲۹/۸
آیا تا حد امکان از کاربرد عبارت های تکراری در گزینه ها خودداری شده است.	۶۷/۷	۳۲/۱
آیا کلمات به کار رفته در ساقه و یا گزینه ها از نظر املایی صحیح هستند؟	۸۷/۴	۱۲/۶
آیا گزینه ها به طور عمودی لیست شده اند؟	۴۲/۶	۵۷/۴
آیا بطور مشخص یک گزینه صحیح می باشد؟ خودداری از طرح سوالات Controversy	۸۶	۱۴
آیا ساقه سوال و گزینه ها مثبت هستند؟	۸۷/۲	۱۲/۸

جدول ۳: نتایج آنالیز کمی آزمون های چهار گزینه ای

درس	KR20 Reliability	Standard Error	Total Test Variance	Standard Deviation	Mean P (item facility)	Mean RPBIS (Discrimination index)	Total Item Variance	Alpha
کودکان ناپیوسته	۰/۶	۲/۸	۲۳	۴/۷	۰/۵۶	۰/۲۲	۸/۶۹	۰/۶۳
کودکان بیمار	۰/۶	۲/۹	۲۳/۳	۴/۸	۰/۵۴	۰/۲۲	۸/۸۶	۰/۶
کودکان ۱	۰/۱	۱/۹	۴/۲	۲/۰۶	۰/۶	۰/۲	۳/۸۲	۰/۱
نوزادان بیمار	۰/۵۹	۱/۹	۹/۶	۳/۱	۰/۶	۰/۳۲	۴/۱	۰/۵۹
بیماری های نوزادان	۰/۷۲	۲/۴۳	۱۹/۷۵	۴/۴	۰/۶	۰/۳	۶/۳۱	۰/۷
بهداشت مادران و نوزادان ۱	۰/۷۲	۲/۸	۳۸/۲	۶/۱۸	۰/۶	۰/۲۹	۸/۷۸	۰/۷۸
بهداشت مادران ۲	۰/۴	۲/۲	۱۰/۲۳	۳/۱	۰/۵	۰/۲	۵/۳	۰/۴
بهداشت جامعه ۱	۰/۵	۲/۴۹	۱۲/۷۱	۳/۵۶	۰/۵۴	۰/۲۵	۶/۴	۰/۵
بهداشت جامعه ۲	۰/۳	۳/۲	۱۴/۹۹	۳/۸۲	۰/۴۹	۰/۱۴	۱۰/۵۵	۰/۳
بهداشت جامعه ۳	۰/۵	۲/۱	۱۰/۵۶	۳/۲۴	۰/۶۴	۰/۲۵	۴/۷۹	۰/۵۶
بهداشت روان ۱	۰/۲	۲/۱	۶/۱	۲/۴۸	۰/۸	۰/۱۵	۴/۷۹	۰/۲۳
بهداشت روان ۲	۰/۵۱	۲/۱	۹/۴۹	۳/۰۸	۰/۷۵	۰/۲۲	۴/۷۸	۰/۵۱

سوال قابل ارزیابی بودند. از این تعداد تنها ۱۲/۱۸ درصد سوالات (۹۳ مورد) مناسب بودند. عمده سوالات شامل ۱۹۲ مورد یا به عبارتی ۴۳/۷ درصد سوالات خیلی آسان و ۶۳ سوال یا ۱۴/۵ درصد سوالات نیز خیلی سخت بودند و مابقی نیز دارای اعتبار درونی اندکی بوده، به طوری که P Index برای آن‌ها محاسبه نشد.

بحث و نتیجه گیری

یافته‌های این تحقیق نشان داد اگرچه سوالات از نظر ساختار در حد خوب و متوسط قرار داشتند اما از نظر تاکسونومی در حد دانش و فهم بودند در حالی که توصیه می‌شود حدود ۳۰ تا ۴۰ درصد سوالات در حد کاربرد و آنالیز باشند. علاوه بر این سوالات از قدرت تمیز و دشواری مناسبی برخوردار نبودند. که به نظر می‌رسد این ضعف در طراحی سوالات ناشی از عدم آگاهی کافی و ندیدن آموزش های لازم اساتید باشد.

اشکالات و نقایص طراحی سوالات چهار گزینه ای در تحقیقات مختلف بررسی شده‌اند. شایع‌ترین اشکالات ساختاری: نارسا بودن ساقه سوال، مبهم بودن سوال یا گزینه‌ها، کاربرد منفی در منفی در سوال؛ عدم تطابق گزینه با ساقه سوال از نظر جمله بندی و روال منطقی سوالات می‌باشد (۱۴).

طبق ارزیابی ساختار سوالات با اصول میلمن ۷ درصد ساختار متوسط و ۹۴/۳ درصد ساختار خوب داشتند که این نشان دهنده نکته مثبت از نظر ساختاری در این مطالعه می‌باشد. در ۶۷/۳ درصد سوالات قسمت اعظم اطلاعات در ساقه سوال گنجانده شده بود. ۹۰/۲ درصد سوالات یک هدف کلی و در ۵۷/۴ درصد سوالات آزمون چیدمان گزینه ها به صورت

همانطور که جدول فوق نشان می‌دهد از ۱۲ آزمون مورد بررسی حداقل ضریب همبستگی ۰/۱ مربوط به درس کودکان ۱ و حداکثر آن ۰/۷۲ مربوط به دو درس بیماری‌های نوزادان و بهداشت مادران و نوزادان ۱ است.

ضریب همبستگی تمامی آزمون‌ها کم و یا خیلی کم بود. خطای استاندارد نیز از ۱/۹ تا ۳/۲ متغیر بود. از مهم‌ترین موارد در جدول فوق مربوط به مقادیر Mean P و Mean RPBIS می‌باشد. Mean P یا میانگین ضریب دشواری کل آزمون نیز از ۰/۴۹ تا ۰/۸ متغیر بود. مطلوب‌ترین ضرایب دشواری مربوط به دروس پرستاری کودکان ناپیوسته، کودکان بیمار، کودکان ۱، بیماری‌های نوزادان، نوزادان ۲، بهداشت مادران ۱، بهداشت جامعه ۱، بهداشت جامعه ۲، بهداشت جامعه ۳ بوده است. دروس بهداشت روان ۱ و ۲ نیز از ضریب دشواری خیلی آسان و تقریباً تمامی ۱۲ آزمون از قدرت تمیز کمی برخوردار بودند. کمترین قدرت تمیز ۰/۱۴ و بیشترین قدرت تمیز ۰/۳۲ بوده است. از مجموع ۱۲ درس صرفاً دروس پرستاری نوزادان ۲ و بیماری‌های نوزادان تقریباً قدرت تمیز مطلوبی با ۰/۳ و ۰/۳۲ داشتند. کمترین قدرت تمیز مربوط به دو درس بهداشت جامعه ۲ و بهداشت روان ۱ به ترتیب با قدرت تمیز ۰/۱۴ و ۰/۱۵ بوده است. کمترین تعداد سوال هر آزمون ۲۱ و بیشترین ۶۴ سوال بوده است. کمترین دانشجو در هر آزمون ۲۰ و بیشترین ۵۲ نفر بوده است.

بهداشت جامعه ۲ و بهداشت روان ۲ کمترین درصد سوالات مطلوب (به ترتیب ۱۰/۹۳ و ۱۱/۴ درصد) داشتند. بیشترین درصد سوالات مطلوب مربوط به بهداشت جامعه ۱ و کودکان بیمار با ۳۹ و ۳۴/۷ درصد بودند. از کل سوالات صرفاً ۴۳۹

بیشتری از سوالات به سطوح عمیق تر حوزه شناختی تاکسونومی دو و سه (اختصاص یافته اند) (۱۱). یک بررسی در یکی از دانشگاه های کشور نیز نشان داد سنجش محفوظات در ۷۷ درصد سوالات بارز است (۲۱).

براساس نتایج این تحقیق تنها ۳/۶ درصد سوالات در سطح تاکسونومی دو و هیچ سوالی در تاکسونومی ارزشیابی قرار نداشتند که نشان دهنده تمرکز طراحان سوال بر محفوظات دانشجویان و نه کاربرد و آنالیز و سنتز و ارزشیابی یا فهم عمیق مفاهیم است. چنین رویکردی موجب سوق دادن دانشجویان به محفوظات و برداشت سطحی مطالب می گردد. آموزش رشته پرستاری همانند سایر رشته های علوم پزشکی آموزش توأم تئوری و بالین برای دانشجو انجام می گیرد. سوالات چند گزینه ای به تنهایی نمی تواند اعتبار لازم را برای سنجش آموخته های دانشجویان داشته باشد. همانطور که بسیاری از دانشگاه های معتبر جهان پیشنهاد نمودند اساتید دانشگاه علاوه بر استاندارد سازی سوالات چند گزینه ای از ارزشیابی جامع تر و دقیق تری استفاده نمایند (۲۲ و ۲۳).

همچنین نتایج نشان داد که ضریب دشواری ۲۱/۱ درصد سوالات مناسب، ۴۳/۷ درصد سوالات خیلی آسان و ۱۴/۵ درصد سوالات نیز خیلی سخت بودند. تحلیل سوالات جامع آزمون نهایی پرستاری در دانشکده علوم پزشکی شاهرود نشان می دهد عمده ترین مشکلات طراحی سؤال، طراحی سوالات با ضریب دشواری خیلی پایین و یا خیلی بالا، و دارای گزینه های فاقد کارایی برای متمایز ساختن دانشجویان ضعیف از قوی است (۲۴). نتایج حاکی از آن بود که هر ۱۲ آزمون از قدرت تمیز کمی برخوردار بودند. کمترین قدرت تمیز ۰/۱۴ و بیشترین قدرت تمیز ۰/۳۲ بود که با نتایج Bachhoff.el (۲۰۰۰) همخوانی دارد (۲۵). در مطالعه شبان اکثر سوالات از قدرت تمیز ضعیفی برخوردار بودند (۲۶). تحقیقی که در دانشگاه علوم پزشکی بابل (۱۳۸۲) انجام شد، نشان داد طراحی آزمون های پایان ترم وضعیت مناسب و راضی کننده نداشته و عموماً آزمون های مورد استفاده در فرآیند ارزشیابی برآیندهای آموزشی دانشجویان، دارای اشکال ساختاری اساسی بوده و این آزمون ها به عنوان ابزار دقیق و معتبر عمل نمی نمایند و تفاوت های واقعی آموذنی ها را نمی سنجند (۲۷). در بررسی روایی و پایایی امتحان جامع در دانشکده پرستاری و مامایی حضرت فاطمه دانشگاه علوم پزشکی شیراز، ۸۶ درصد سوالات، مستلزم اصلاح تشخیص داده شدند و محققان به این نتیجه رسیدند که توجه بیشتر در آزمون سازی و

عمودی نبود. بنابراین با کمی پیگیری و آموزش و ممارست اساتید دانشکده پرستاری و مامایی در بخش ساختاری سوالات می توانند، سطح آزمون ها را ارتقاء دهند.

در تحقیق حق شناس و همکاران (۱۳۸۷) نیز ۴۶ درصد سوالات بدون اشکال بودند و بقیه یک یا چند اشکال ساختاری داشتند. عدم رعایت چیدمان عمودی گزینه ها، همسنگ بودن گزینه ها و آوردن بخش اعظم اطلاعات در ساقه سؤال، در بیش از ۱۰ درصد سوالات وجود داشت (۱۱). در مطالعه حاضر هیچ سوالی با ساختار ضعیف شناسایی نشد، تنها ۵/۷ درصد ساختار متوسط و ۹۴/۳ درصد ساختار خوب داشتند. بیشترین اشکال ساختاری چیدمان عمودی گزینه ها، نبود بخش اعظم اطلاعات در ساقه سوال، عبارات تکراری در گزینه ها، طول و ساختار لغوی، سبک نگارش و همسنگ نبودن گزینه ها را تشکیل می داد. نتایج پژوهش جوادی و همکاران (۱۳۸۶) و شکورینا و همکاران (۱۳۸۸) با این یافته ها همخوانی دارد (۱۶ و ۱۵).

نتایج تحقیق واعظی و همکاران (۱۳۸۶) نشان داد میزان رعایت قواعد تهیه سوالات چند گزینه ای در سوالات پایان ترم اعضای هیات علمی دانشکده های پزشکی، بهداشت و پرستاری مامایی تنها در ۲۴ درصد موارد خوب بوده است. بیشترین درصد نمرات خوب مربوط به دانشکده پرستاری و مامایی و کمترین درصد نمرات خوب مربوط به دانشکده پزشکی بود (۱۷).

بر اساس نتایج پژوهش حاضر ۹۶/۴ درصد سوالات در سطح تاکسونومی یک (دانش و فهم) و تنها ۳/۶ درصد سوالات در سطح دو (کاربرد و آنالیز) قرار داشتند. هیچ سوالی در تاکسونومی سه (سنتز و ارزشیابی) قرار نگرفت. که این بیانگر این است که کسب امتیاز قبولی با این آزمون ها نمی تواند نشان دهنده یادگیری دانشجو از مطالب درسی باشد چرا که این آزمون ها بیشتر در سطح تاکسونومی یک بودند. هر چند مطالعات دیگر هم در سطح کشور نقص سوالات را در این زمینه نشان می دهد. این یافته ها با مطالعه حق شناس و همکاران که نشان داد که از نظر توزیع سطح دانش ۷۷/۴ درصد سوالات در تاکسونومی یک ۱۶/۸ درصد در تاکسونومی دو و بقیه در تاکسونومی سه طراحی شده اند همخوانی دارد (۱۱). مطالعه دیگر نیز نشان دادند که بین ۷۰ تا ۷۸ درصد سوالات دانشجویان پزشکی در سطح تاکسونومی یک بودند (۱۸ و ۱۹). همچنین این نتایج با مطالعه Huxham (۱۹۸۰) همخوانی دارد (۲۰). اگرچه در سال های اخیر سهم

آزمون‌های مورد بررسی در مطالعات، تمرکز عمده طراحان سؤال بر محفوظات دانشجویان و نه فهم عمیق و تحلیل و کاربرد مفاهیم، بوده است، جای تأمل دارد، چرا که بدیهی است چنین رویکردی موجب سوق دادن دانشجویان به محفوظات و برداشت سطحی مطالب می‌گردد که در مورد مطالب علوم پزشکی این شیوه یادگیری نه تنها از ماندگاری لازم در ذهن دانشجو برخوردار نیست، بلکه از کارکرد لازم در زمینه پایه ریزی بستر دانش آینده و گسترش توانایی‌های ذهنی دانشجو کاملاً دور است.

با توجه به اهمیت نقش ارزیابی در فرآیند آموزش و تحکیم و تثبیت آموخته‌های دانشجویان در امتحانات پایان ترم، ارائه سوالات استاندارد و منطبق با قواعد طراحی سوالات چند گزینه‌ای اجتناب ناپذیر است که برای نیل به این هدف، آموزش اعضای هیات علمی برای این منظور ضروری است و شایسته است که مراکز مطالعات و توسعه آموزش پزشکی کشور در خصوص بررسی سوالات چند گزینه‌ای و ارائه بازخورد مناسب به آنان همت گمارند.

قدردانی

از مرکز مطالعات و توسعه آموزش پزشکی و معاونت پژوهشی دانشگاه علوم پزشکی گلستان جهت حمایت معنوی و مادی در انجام این پژوهش و همچنین از گروه‌های آموزشی که امکان ارزیابی آزمون‌ها را فراهم نمودند تشکر و قدردانی می‌گردد.

اصلاح گزینه‌ها هنگام استفاده مجدد از سؤالات باید مورد تأکید قرار گیرد (۲۸). کیفیت یادگیری دانشجویان را می‌توان با نتایج آزمون‌ها بررسی کرد حال اگر آزمون‌ها از کفایت لازم برخوردار نباشند در تصمیم‌گیری آموزشی خلل ایجاد می‌شود. ارزیابی آزمون‌های یک گروه از دانشجویان مامایی نشان داد که ۷۱ درصد سوالات در سطح یک طبقه‌بندی بلوم قرار داشت. فراوانی نسبی سؤالات آسان در امتحانات پایان ترم ۵۲/۳۵ درصد بود و تنها ۲۷ درصد سؤالات دارای ضریب تمیز مناسب بودند (۲۹). در یک بررسی که با هدف مقایسه روش‌های ارزیابی MCQ، MEQ، EMQ در دانشجویان پزشکی صورت گرفت. ۵۵ درصد دانشجویان اظهار داشتند بعضی از سوالات آزمون MCQ و EMQ را با حدس به درستی پاسخ داده‌اند (۳۰).

با توجه به مجموعه یافته‌های این مطالعه و شواهد فوق پژوهشگران معتقدند سؤالات امتحانی باید به گونه‌ای ساخته شوند که از نتایج آن‌ها بتوان احتمال موفقیت فراگیران را با دقت پیش‌بینی کرد. بنابراین با توجه به اهمیت ارزشیابی در نظام آموزش عالی و لزوم آگاهی و توانمندی اساتید دانشگاه‌ها نیاز به پیگیری و استمرار برنامه‌هایی در زمینه ارتقاء و کیفیت آزمون‌ها وجود دارد. جهت عملکرد بهینه اساتید در طراحی و اجرای آزمون چند گزینه‌ای راهکارهای لازم جهت بهینه نمودن دانش و نگرش اساتید در مورد ارزیابی فراگیران اتخاذ گردد که شامل مواردی چون: برگزاری کارگاه‌هایی در زمینه طراحی سوالات چندگزینه‌ای و ایجاد بانک سوالات استاندارد در دانشگاه‌ها می‌باشد. به هر حال این واقعیت که در

Reference

1. Naderi E, Seifnaraghi M. [Assessment and measurement and analytical elements of it in psychological sciences]. Tehran: Moein publication: Co.1982.[Persian]
2. Haj-Bagheri A. [Meta-evaluation]. 15th International congress of geographic medicine and 5th National congress on medical education. Shiraz: 18-20 Azar 1381 (Dec2002) [Persian]
3. Najjar Sh, Abedi P. The survey of knowledge and practice of faculty members about validity and reliability of exams. Iranian Journal of Medical Education. 5th National Congress on Medical Education. 2002; 8, page 41. Available from: http://www.ijme.ir/browse.php?a_code=A-10-4-266&slc_lang=fa... [Persian]
4. Mansourian HR, Shoja MR, Salimi T, Mojahed Sh, Karimi H, Aghili H, et al. [Face validity of exams of faculty members of Yazd University of Medical Sciences and its concordance with lesson plan in 2002]. 5th National Congress on Medical Education. Isfahan: Iranian Journal of Medical Education; 2002: 2(1) 7. [Persian]

5. Saburi Masih, Shayan Sh, Salehi A, Honarpisheh H. Review Psychometric Parameters of the 29th Residency Test (1380) According to the Classic Test Theory (CTT). IJME 2002; 2: 49-50. Available from: http://www.ijme.ir/browse.php?a_code=A-10-4-284&slc_lang=fa&sid=1 [Persian].
6. Kaveh M. [Analytical assessment of multiple choice test at medical school]. Medical Journal of Mashhad University of Medical Sciences 2002; 76 (45): 89-95. [Persian]
7. Hammond EJ, Mcindo AK, Sansome AJ, Spargo PM. Multiple choice examination: Adapting an evidence based approach to exam technique. Anesthesia 1998; 53(11): 1105-8.
8. McCourbie P. Improving the fairness of multiple choice questions: A literature review. Med Teach 2004; 26(8): 709-712.
9. Vakili Z, Sayah M, Fakhrian E, Rasouli Nejad SA, Mosayebi Z, Moniri R. [Survey on psychometric characteristics of resident examination Kashan Medical University]. The 8th National Congress of Medical Education. Kerman: Kerman University of Medical Sciences; 2007: 174. [Persian]
10. Toghyanifar N, Gorgyaraghi M, Esmaili A, Avizhgan M, Mazaheri M. [Survey on quality of Multiple Choice Questions of clinical clerkship Isfahan Medical Faculty 2005-2006]. The 8th National Congress of Medical Education. Kerman: Kerman University of Medical Sciences; 2007: 73. [Persian]
11. Haghshenas MR, Vahidshahi K, Mahmudi M, Shahbaznejad L, Parvinnejad N, Emadi A. [Evaluation of Multiple Choice Questions in the School of Medicine]. The First Semester of 2007 Mazandaran: Mazandaran university of medical sciences; 2008: 5(2): 119-126. [Persian]
12. Schotamus C. Student assessment and examination rules .Medical Teacher 1999; 21(3): 318-321.
13. Shabani H. [Educational and traing skills]. Tehran: SAMT publication:Co. 2001.[Persina]
14. Gareth H, Elzubier M. Imprecise terms in Uk Medical multiple choice questions: examiners think they mean. Med Educe, Glasgow, UK, 1998; 32: 343-350.
15. Javadi M. Abbaszadeh A, Borhani F, Abdoli R. [Quality analysis of multiple choice questions examinations of Bam School of Nursing]. 8th National Congress on Medical Education. Kerman: Kerman University of Medical Sciences ;2007:217. [Persian]
16. Shakournia AH, Khosravi Borojeni A, Mozafari A, Elhampour H. [An evaluation of exam questions designed by faculty members, emphasizing on the Multiple Choice Question structure, Ahvaz Jundishapur University of Medical Sciences 2007]; SDME 2009;6(2): 129-138 [Persian].
17. Vaezi AK, Morovati MA, Heydari MR. [Evaluation multiple choice questions in final semester exams]. 9th National Conference on Medical Education. Yazd: Shaheed Sadoughi University of Medical Sciences; 2007: 250-251. [Persian]
18. Gorg-Yarghi M, Ebrahimi E, Ashourian V, Azadbakht S, Avezhegan M, Esmaeli A, et al. [Analysis of multiple choice questions of academic progress for medical students and designing necessary intervention for improving student evaluation process]. 9th National Conference on Medical Education. Yazd: Shaheed Sadoughi University of Medical Sciences and Health Services; 2007: 27-28. [Persian]
19. Vahidshahi K, Mahmudi M, Shahbaznejad L, Shakeri S. [Evaluation of Multiple Choice Questions in the School of Medicine in 2006-7. 9th National congress of medical education]. Yazd: Shaheed Sadoughi University of Medical Sciences ; 2007: 41-42. [Persian]

20. Huxham GJ, Naeraa N. Is Bloom's taxonomy reflected in the response pattern to MCQ items? *Med Educ* 1980; 14(1): 6-23.
21. Arab M. [Evaluation defects of multiple choice questions in medical school of Hamadan University of Medical Sciences]. 4th National congress of Medical Education. Tehran: Research Deputy Publication; 1989:51[Persian].
22. Sim kin MG, Kuechler WL. Multiple choice test and student understanding: What is the connection? *Decision Scie J Innovat Educ* 2005; 3(1): 73-98.
23. Bush ME. Quality assurance of multiple choice tests. *Qual Assur Educ.* 2006; 14(4): 398-404.
24. Dadgari A, Ebrahimi H, Hassani MR, Mashayekhi Y. [Analysis of the comprehensive nursing final exam in Shahroud Faculty of Medical Sciences]. *Journal of Shahrekord University Of Medical Sciences* 2003;4(4): 10-4. [Persian]
25. Bachhoff E, Larrazolo N, Rosas M. The level of Difficulty and Discrimination Power of the Basic knowledge and skills Examination (EXHCOBA). *Electronic Journal of Educational Research* 2000; 2(1): 1-16.
26. Shaban M., Ramezani Badr F. [Effect of Test Item Analysis of Summative Exams on Quality of Test Designing]. *The Journal Of Faculty Of Nursing & Midwifery* 2007;13(1): 5-15. [In Persian]
27. Ashraf Pour M, Beheshti Z, Molook Zadeh S. [Quality of final examination in students of Babol Medical University1999-2000]. *Journal of Babol University of Medical Sciences* 2003; 5(Suppl.2): 42-7. [Persian]
28. Motahari M, Hashemi F, Parsayekta Z. [Validity and reliability of final exam in Hazrat Fatehm School of Nursing and Midwifery in Shiraz University of Medical Sciences]. 4th National Congress of Medical Education.Tehran: Research Deputy of Tehran University of Medical Sciences; 2009: 61. [Persian]
29. Kazemi A, Ehsanpour S, Hassanzadeh A. [Investigating the Academic Achievement Evaluation of Specialized Theoretical Courses of Midwifery BS]. *IJME* 2010; 9(4): 346-355. [Persian]
30. Asadi SN, Afshari R. [Comparison of results of MCQ, MEQ, EMQ for finding the most equitable written exam among students of medical school]. 10th National Congress of Medical Education. *Green Journal; Quarterly of Medical Education*;2009: 278-8. [Persian]

Quantitative and Qualitative Analysis of Academic Achievement Tests in Golestan University of Medical Sciences

* Sanagoo A, ** Jouybari L, *** Ghanbari Gorji M

Abstract

Introduction: Multiple choice exams are most common tests in medical education. So some teacher-made tests have different problems. The aim of this study was to evaluate quantitative and qualitative aspects of MCQ in nursing school in North of Iran in 2009.

Methods: In this descriptive study, 523 multiple choice questions of community health, pediatric psychology nursing groups including 12 courses at second academic semester of 2009 were studied. The structure of questions were assessed by Millman checklist. Bloom taxonomy of the questions considered as well. Difficulty, discrimination, and other quantitative indicators were evaluated by software too.

Results: From 523 questions, 94.3% had a good structure. Most structural problems related to linear order of items (57.4%), lack of major information in question stem (32.7%), repeated words in the items (32.1%) and lack of coordination among items respect to length and vocabulary structure (29.8 %). 96.4% of questions were from "knowledge and understanding" taxonomy. 79.8% of questions were about learning obligation. The average score of test difficulty was from 0.49 to 0.8. Generally, 21.1% of questions were appropriate, 43.7% were very easy questions and 14.5% were too hard. Almost all of the tests had low discrimination power from 0.14 to 0.32. Only two courses had optimum discrimination power.

Conclusion: Although the structure of multiple choice tests were relatively good, but the distribution of Taxonomy, difficulty and discrimination power of the tests should be corrected organizing educational workshops for faculty members also evaluating their activities by Educational Development Center of the Universities can be effective.

Keywords: academic achievement multiple choice test, Taxonomy, qualitative and quantitative analysis Millman principles

Corresponding Author: Dr. Leila Jouybari, Educational development center, Golestan University of medical Sciences.
lm_jouybari@yahoo.com