

بررسی روایی، پایایی و سطح سوالات چندگزینه‌ای آزمون‌های پایان ترم دروس تخصصی کارشناسی پرستاری در دانشگاه علوم پزشکی اصفهان

حبیب‌اله حسینی: عضو هیات علمی، گروه بهداشت و مدیریت، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.

مهسا شکور*: عضو هیات علمی، مرکز مطالعات و توسعه آموزش پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران.
عبداله رضایی دهاقانی: عضو هیات علمی، گروه بهداشت و مدیریت، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.

چکیده: در هر برنامه آموزشی ارزشیابی رکنی اساسی بوده که میتواند آموزش را به مسیر صحیح هدایت نماید. استفاده از آزمون خوب پیشرفت تحصیلی، مستلزم در نظر گرفتن تمامی هدفهای آموزشی و همه محتوای برنامه درسی است. در این تحقیق، با توجه به اهمیت دروس تخصصی نظری پرستاری، محققین به ارزشیابی امتحانات پایان ترم دروس تخصصی نظری پرستاری دانشگاه علوم پزشکی اصفهان پرداخته‌اند. متغیرهای مورد بررسی در سوالات امتحانی شامل نوع سوالات، سطح سوالات براساس طبقه بندی بلوم، ضریب دشواری و شاخص تمیزسوالات، روایی، پایایی آزمونها و تعداد سوالات برای محتوای هر درس بود. نتایج پژوهش حاضر نشان داد که ۲۰۹ سوال چندگزینه‌ای (۵۶ درصد) در سطح یک حیطة شناختی طبقه بندی بلوم قرار داشتند. ۱۹۷ سوال آسان (۵۸/۶ درصد) بود و ۱۲۱ سوال (۳۶ درصد) دارای ضریب تمیز مناسب بودند. بر اساس آزمون پیرسون بین تعداد سوالات به ازای هر محتوای آموزشی در امتحانات با برآورد تعداد سوالات مناسب برای هر محتوای آموزشی، همبستگی معنی دار بود ($p < 0.001$). در بررسی سوالات آزمون‌های دروس تخصصی نظری پرستاری از نظر تعداد سوالات و ضریب تمیز نامناسب بود که میتواند منجر به تمایز نادرست و نتیجه گیری اشتباه از یادگیری دانشجویان شود. این در حالی است که ارزشیابی باید چنان باشد که بتواند سنجش درستی از میزان یادگیری دانشجویان در تمام سطوح آموزشی باشد اما متاسفانه به دلایل مختلف به این مهم توجه نشده بود و بیشتر سوالات آسان و در سطح شناختی طراحی شده بودند. شناخت این نواقص ارزشیابی از سوی مسئولین آموزشی و کمیته ارزشیابی دانشکده‌ها و اقدام جهت رفع آن ضروری به نظر می‌رسد.

واژگان کلیدی: آموزش پرستاری، آزمون، ارزشیابی.

*نویسنده‌ی مسؤؤل: عضو هیات علمی، مرکز مطالعات آموزش پزشکی، دانشگاه علوم پزشکی اراک، اراک، ایران.

Email: s.shakour@gmail.com

مقدمه

نمره آنها بین ۰/۳ تا ۰/۷ هست. برای ارزیابی سطح سوالات از سطح بندی بلوم در یادگیری شناختی استفاده می شود و به این ترتیب سوالات می توانند از سطوح پایین به بالا در شش سطح شناختی دانش، درک، کاربرد، تحلیل، ترکیب و ارزشیابی قرار گیرد (سیف ۲۰۰۴). برای ارزیابی روایی نیز می توان روایی پیشگو، محتوا، ملاکی و صوری را ارزیابی کرد که بستگی به هدف ارزیابی سوالات دارد. ارزیابی پایایی نیز به روش های مختلف انجام می شود که معمول ترین آنها ضریب الفای کرونباخ و دو نیمه کردن است (سیف ۲۰۰۰).

با توجه به اهمیت غیر قابل انکار دروس نظری اختصاصی و ارزشیابی آن برای ورود به آموزش بالینی در دانشجویان پرستاری و نبود اطلاعات لازم در مورد نحوه انجام ارزشیابی تحصیلی در دانشگاه علوم پزشکی اصفهان، این تحقیق سعی بر آن داشت تا با شناسایی وضعیت موجود در زمینه ارزشیابی امتحانات، شرایط را برای ارتقای وضعیت ارزشیابی دانشجویان فراهم کند. بدین منظور مطالعه حاضر با هدف ارزشیابی امتحانات پایان ترم دروس تخصصی نظری پرستاری و درک جایگاه واقعی امتحانات بر اساس شاخص های استاندارد در مدل کلاسیک ارزشیابی سوالات امتحانی انجام گردید.

روش کار

این پژوهش توصیفی به صورت مقطعی بر روی سوالات آزمون دروس نظری تخصصی پایان ترم دانشجویان کارشناسی پیوسته پرستاری در نیمسال اول تحصیلی در دانشگاه علوم پزشکی اصفهان انجام شد. روش نمونه گیری سوالات آزمون به روش نمونه گیری در دسترس بود و عدم دسترسی به اساتید و یا عدم تمایل اساتید درس برای در اختیار قرار دادن سوالات مربوط به درس خود، از معیارهای خروج این پژوهش بود.

سوالات آزمون های مورد ارزیابی ۸ درس، شامل پرستاری نوزادان، آموزش به بیمار، فوریتهای بهداشتی جامعه ۱، بهداشت روان ۱، بهداشت روان ۲، بهداشت جامعه ۳ و داخلی جراحی ۴ بود. ابزار جمع آوری اطلاعات، پرسشنامه بود. پرسشنامه قبلاً توسط

ارزشیابی پیشرفت تحصیلی سنجش عملکرد یادگیرنده گان و مقایسه نتایج حاصل با هدف های آموزشی است. که می تواند به منظور تصمیم گیری در این باره که آیا فعالیتهای آموزشی استناد و کوششهای یادگیری دانشجویان به نتایج مطلوب رسیده است باشد. همچنین از اهداف مهم امتحان تعیین شواهد پیشرفت برای شروع مراحل بعدی تدریس و طبقه بندی فراگیران و نیز اصلاح روش تدریس معلم است (وینک ۲۰۰۸). از طرفی شواهد متعددی نشان داده است که کیفیت آزمونها بطور اساسی کیفیت آموزش و روند یاددهی - یادگیری را تحت تاثیر قرار می دهد (لارسن و همکاران ۲۰۰۸؛ رید و همکاران ۲۰۰۷؛ بردی ۲۰۰۵) بنابراین کیفیت سوالهای طرح شده در این آزمونها از اهمیت بالایی برخوردار است.

یک آزمون مناسب پیشرفت تحصیلی، آزمونی است که به بهترین شکل منعکس کننده تمامی هدف های آموزشی و همه محتواهای برنامه درسی باشد و توزیع و نوع سوالات بر اساس حجم و اهمیت محتوای آموزشی و ارزش هر محتوی تعیین شود، به طوری که کلیه اهداف آموزشی را پوشش دهد (حیدری و همکاران ۲۰۰۴). از روش های ارزشیابی امتحانات، استفاده از مقیاسهایی همچون ضریب تمیز، ضریب دشواری، سطح سوالات، روایی و پایایی سوالات است. ضریب تمیز به معنای قدرت تمیز سوال بین دانشجوی ضعیف و قوی است که مقیاس آن بین یک تا منفی یک متغیر است. بهترین سوالات، سوالاتی هستند که نمره آنها بین ۰/۵ تا ۱ هست. سوالات با ضریب تمیز منفی نشان می دهد که در آن سوال دانشجوی گروه قوی بدتر از گروه ضعیف عمل کرده است. این گونه سوال ها دارای معایب اساسی هستند که باید به طور کلی کنار گذاشته شوند و یا در آنها تجدید نظر اساسی صورت پذیرد (سیف ۲۰۰۰). ضریب دشواری نیز جهت ارزیابی آسان بودن سوالات برای دانشجویان است که هرچه نمره ضریب کمتر شود سوال سخت تر و برعکس هرچه نمره بیشتر شود سوال آسان است و لذا سوالاتی مناسب هستند که

بین تعداد سوالات هر محتوای آموزشی و تعداد سوالات برآورد شده برای هر محتوای آموزشی، آزمون ضریب همبستگی اسپیرمن و پیرسون مورد استفاده قرار گرفت. برای حفظ امانت، نتایج امتحانات با نام قراردادی جدید دروس گزارش شده است.

یافته ها

در مطالعه حاضر هشت درس از دروس دوره کارشناسی پرستاری جهت بررسی وضعیت ارزشیابی امتحانات پایان ترم دروس مورد ارزیابی قرار گرفت. در ابتدا ارتباط بین تعداد جلسات دروس بر حسب سرفصل های طرح درس در هر جلسه و تعداد سوالات طراحی شده در امتحان پایان ترم با کمک آزمون پیرسون بررسی گردید که نتایج نشان داد همبستگی بین تعداد جلسات و تعداد سوالات طراحی شده در مجموع کلیه دروس از نظر آماری معنی دار بود ($r=0/34$; $P\text{-value}<0/01$) (جدول شماره ۱).

جدول شماره ۱. رابطه همبستگی بین تعداد جلسات و تعداد سوالات

P-value	ضریب همبستگی اسپیرمن	[دامنه میان چارگی] میانه	میانگین تعداد سوالات به ازای هر جلسه آموزش ± انحراف معیار
۰/۰۰۱	۰/۳۴	۱ [۰/۵-۱]	۰/۹۴ ± ۰/۶۷

در مجموع سوالات تمامی دروس ۳۳۶ سوال بود که تک تک سوالات از جنبه های مختلف بررسی شدند. در بررسی سطح سوالات طراحی شده در حوزه شناختی مشخص گردید ۲۰۹ سوال (۵۶/۲ درصد) از کل سوالات در سطح دانش، ۱۰۸ سوال (۲۹ درصد) در سطح درک، ۳۸ سوال (۱۰/۲ درصد) در سطح کاربرد، ۱۴ سوال (۳/۸ درصد) در سطح تحلیل و ۳ سوال (۰/۸ درصد) در سطح ترکیب بوده اند. که بیشترین سوالات طراحی شده به تفکیک دروس نیز در سطح دانش بوده است، که در این نتایج به تفکیک دروس ارائه شده است. در خصوص ضریب دشواری ۱۹۷ سوال (۵۸/۶ درصد) آسان، ۱۰۸ سوال (۳۱/۹ درصد) مناسب یعنی با ضریب دشواری

کاملی در پژوهش بررسی وضعیت ارزشیابی تحصیلی دروس تخصصی نظری دوره کارشناسی پیوسته مامایی در سال ۸۶ ارزیابی و استفاده شده بود که ضریب همبستگی ۰/۷۵. پایایی داشت. (کاملی و همکاران ۲۰۱۰). این پرسشنامه شامل قسمت هایی برای ثبت مشخصات کلی هر درس (نام درس، تعداد واحد، نحوه ارزشیابی، تعداد سوالات به ازای زمان اختصاص داده شده به تدریس هر محتوا در طرح درس، سطح عنوان شده در اهداف درس) (در طرح درس) و مشخصات سوالات (سطح سوالات، نوع سوالات، ضریب دشواری و شاخص تمیز هر سوال) بود. نمره ضریب دشواری بالای ۰/۷. سوال آسان، زیر ۰/۳. سوال سخت و بین ۰/۳ تا ۰/۷. سوال مناسب در نظر گرفته شد و نمره شاخص تمیز بین ۰/۵ تا ۱. نشان دهنده سوال مناسب، کمتر از ۰/۵ تا ۱. نشان دهنده سوال با شاخص تمیز کم و کمتر از ۰/۵ تا ۱- سوال نامناسب گزارش شد.

روش انجام کار به این صورت بود که کلیه سوالات امتحانات پایان ترم و طرح درس دروس تخصصی از اساتید مربوط به هر درس (که تمایل داشتند همکاری کنند و سوالات و طرح درس های مربوطه را در اختیار محقق قرار دادند) دریافت شد. ابتدا برای هر درس یک پرسشنامه در نظر گرفته شد. سپس سوالات و طرح درس برای تعیین سطح سوالات و سطح عنوان شده در اهداف درس براساس طبقه بندی بلوم در اختیار دو تن از متخصصین آموزش پزشکی قرار گرفت و اطلاعات مربوطه و همچنین اطلاعات مربوط به تعداد سوالات بوسیله محقق در پرسشنامه ثبت گردید. روایی صوری و روایی محتوای سوالات نیز ضمن مراجعه به اهداف آموزشی، با کمک متخصصین موضوع درسی بررسی شد. ضریب دشواری و شاخص تمیز سوالات چند گزینه ای نیز به تفکیک هر درس و با هماهنگی از مرکز خدمات ماشینی دانشگاه علوم پزشکی دریافت و بوسیله محقق در پرسشنامه ثبت گردید.

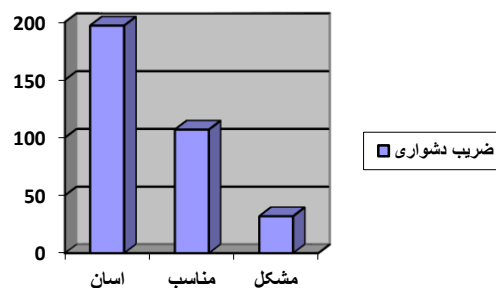
برای تجزیه و تحلیل داده ها از نرم افزار SPSS نسخه ۱۶ استفاده شد. برای ارزیابی همبستگی بین سطوح سوالات امتحانات و سطوح محتوای آموزشی هر سوال در طرح درس و همچنین ارزیابی میزان همبستگی

این در حالی است که کسب توانایی استدلال و استنتاج از اهداف آموزشی گروه های پزشکی است و کسب مهارت تشخیص و به کارگیری آموخته های دروس تخصصی لازمه فعالیت در حیطه پرستاری است و برخورداری پرستاران از دانش لازم و کافی و افزایش توانایی آنان در تصمیم گیری منجر به بهبود عملکرد آنان می شود (استونر ۱۹۹۹). براساس نتایج به دست آمده از ارزشیابی امتحانات پایان ترم دروس تخصصی پرستاری بیش از آن که به ارزیابی قدرت استدلال و ارزشیابی آنها پرداخته شود به اندازه گیری سطوح دانش پرداخته شده است. متاسفانه آزمون توان شناختی دروس تخصصی دانشجویان پرستاری در این سطح از سوالات ممکن است نتواند دانشجویان را در راستای وظایفی که در آینده از آنها انتظار می رود ارزیابی کند.

در رابطه با ضریب دشواری و شاخص تمیز گفته می شود سوال هایی مناسب هستند که ضریب دشواری متوسط و ضریب تمیز بالا دارند و توصیه می شود پس از برگزاری هر آزمون ضریب دشواری و شاخص تمیز سوالات تعیین شود و سپس طراح سوال، سوالات را مطابق با ضریب دشواری و شاخص تمیز اصلاح کند (سیف ۲۰۰۰) و سطح دشواری سوالات باید با هدف های تدریس و محتوای دروس و روش تدریس و توانایی فراگیران مطابقت داشته باشد (نادری و همکاران ۱۹۹۷). در این مطالعه ارزیابی نتایج به دست آمده در مورد ضریب دشواری سوالات نشان می دهد در ۵۸/۶ درصد امتحانات، سوالات آسان بیشترین فراوانی نسبی را به خود اختصاص داده اند و فقط ۳۱/۹ درصد از سوالات از نظر ضریب دشواری مناسب بوده اند. در تحقیق کاظمی سوالات ۷۳/۳۳ درصد امتحانات آسان بوده و بیشترین فراوانی نسبی را به خود اختصاص داده بود و تنها در دو مورد سوالات مناسب از نظر ضریب دشواری، بیشترین فراوانی را دارا بوده است (کاظمی و همکاران ۲۰۱۰). در تحقیق دادگری و همکاران (۲۰۰۳) نیز در تحلیل سوالات جامع آزمون نهایی پرستاری ۵۷/۷۸ درصد سوالات آسان و ۲۶/۹۴ درصد مناسب و ۱۵/۲۸ درصد سوالات مشکل بود (دادگری ۲۰۰۳). در مطالعه حیدری نیز

متوسط و ۳۲ سوال (۹/۵ درصد) دشوار بوده اند (نمودار شماره ۱). همچنین ارزیابی شاخص تمیز نشان داد ۱۲۱ سوال (۳۶ درصد) از کل سوالات دارای ضریب تمیز مناسب، ۱۷۰ سوال (۵۰/۶ درصد) دارای ضریب تمیز کم و ۴۵ سوال (۱۳/۴ درصد) دارای ضریب تمیز نامناسب بودند.

در بررسی روایی محتوی و روایی صوری کلیه سوالات امتحانی، نتایج نشان داد ۹۷/۳ درصد از سوالات، روایی محتوا و ۹۳/۳ درصد از آنها روایی صوری داشتند.



نمودار ۱. نمودار فراوانی سوالات بر اساس ضریب دشواری

بحث

نتایج ارزیابی هشت درس بررسی شده نشان داد بر اساس طبقه بندی بلوم سوالات در سطح یک یادگیری (دانشی) بیشترین فراوانی (۵۶/۲ درصد) را در کلیه دروس به خود اختصاص داده و سوالات در سطوح بالا درصد کمی را به خود اختصاص داده اند و سطوح تحلیل و ترکیب جمعاً فقط ۶/۴ درصد را شامل می شدند. این نتایج با نتایج مطالعات پژوهش خدام و همکاران (۲۰۰۵) همخوانی دارد در بررسی او نیز سطوح مورد ارزیابی در سوالات طراحی شده مربیان پرستاری و مامایی نشان دهنده پایین بودن سطوح ارزیابی شده یادگیری است چنانچه ۶۵ درصد از سوالات در سه سطح پایین شناختی قرار دارند (کاظمی ۲۰۱۰). همچنین در تحقیق کاظمی (۲۰۱۰) تحت عنوان بررسی وضعیت ارزشیابی دروس نظری دوره کارشناسی مامایی در اصفهان در ۸۳/۳۳ درصد امتحانات بیش از ۵۰ درصد سوالات در سطح یک قرار داشته است (خدام و همکاران ۲۰۰۵).

صوری مناسبی بوده اند. برخلاف نتایج گفته شده در مطالعه مک کوربی و همکاران در دانشگاه بریستول (کوربی و همکاران ۲۰۰۴) و شکورنیا (شکورنیا و همکاران ۲۰۱۰) در اهواز در درصد قابل توجهی از سوالات، وجود اشکالاتی در این زمینه را گزارش کرده اند و نتایج تحقیق کاظمی واحسان پور نشان داده است در کلیه امتحانات بیش از ۷۵ درصد از سوالات دارای روایی محتوی بوده است اما در ارزیابی روایی صوری تنها در ۲۸/۷۸ درصد از امتحانات بیش از ۹۰ درصد روایی صوری رعایت شده است (کاظمی و همکاران ۲۰۱۱).

در مطالعه حاضر تنها ۳۰ درصد از آزمون ها دارای روایی ملاکی بودند. وجود فراوانی نسبی ۷۰ درصد ارزشیابی بدون روایی ملاکی نشان دهنده ضعف سیستم ارزشیابی از نظر روایی ملاکی در درس نظری تخصصی پرستاری است. در تحقیق کاظمی و احسان پور آزمون های ۶۱/۱۱ درصد از درس نظری تخصصی مامایی دارای روایی ملاکی بوده اند که در مقایسه با تحقیق حاضر، درس نظری تخصصی پرستاری از روایی ملاکی پایین تری برخوردار بوده اند (کاظمی و همکاران ۲۰۱۱).

هر چند سوالات چندگزینه ای بدلیل پایایی زیادی که دارند محبوبیت بیشتری دارند ولی کارایی کافی در ارزیابی یادگیری در سطوح بالای شناختی ندارند، لذا اغلب صاحب نظران توصیه می کنند برای استفاده بهینه از ویژگی های انواع سوال، بهتر است در یک آزمون، از انواع مختلف سوال استفاده شود (سیف ۲۰۰۴) ولی در آزمون های ارزشیابی مورد بررسی در این تحقیق، اساتید به شکل گسترده ای از سوالات چندگزینه ای استفاده نموده اند. که وجود نواقصی در تدوین این گونه سوالات با بروز کاهش روایی ملاکی همراه شده است. باید توجه داشت که در سوالات چند گزینه ای فراگیران قوی بیشتر از فراگیران معمولی قادر به پیدا کردن اشکالات و پیچیدگی ها و نظرات غیر معمولی در سوال ها هستند، اما از آنجا که تنها یک پاسخ صحیح در سوال وجود دارد عدم دقت طراح سوال در طراحی سوالات چند گزینه ای باعث می شود تیزبینی دانشجویان قوی نه تنها باعث تشویق آنان نشود بلکه با دادن جوابی که مورد نظر

که بر روی سوالات دانشجویان پرستاری و مامایی دانشگاه شاهد انجام شده است ۶۰ درصد از سوالات دارای ضریب دشواری بالای ۰/۷ و ۱۰ درصد از سوالات دارای ضریب دشواری زیر ۰/۳ داشته اند (حیدری ۲۰۰۴). در حالی که در بررسی آزمونهای پرستاری در کشور مالزی ۴۰ درصد از سوالات ضریب دشواری بالای ۸۰ درصد داشته اند (میترا و همکاران ۲۰۰۹).

ارزیابی شاخص تمیز سوالات نیز نشان داد در بین مجموع سوالات ارزیابی شده فقط ۳۶ درصد از سوالات شاخص تمیز بالا داشتند. در تحقیق کاظمی نیز از مجموع سوالات ارزیابی شده ۴۴/۲۷ درصد از سوالات دارای شاخص تمیز مناسب بوده اند (۸). که کمی بیشتر است ولی هنوز تا حد مطلوب فاصله زیادی دارد. در تحقیق دادگری فقط ۵/۴ درصد سوالات جامع پرستاری شاهرود دارای شاخص تمیز بالا هستند (دادگری و همکاران ۲۰۰۳). و باید گفت ضریب تمیز کم سوالات در همه این مطالعات نشان دهنده ضعف طراحی سوالات در دستیابی به شاخص تمیز مناسب است. اما در مقاله میترا و همکاران در مالزی که میزان ضریب تمیز مطلوب را بالای ۰/۲ در نظر گرفته اند نتایج نشان می دهد که وضعیت کمی بهتر است ولی هنوز مناسب نیست و فقط در حدود نیمی از سوالات توانایی کافی برای تمایز گذاشتن بین دانشجوی قوی و ضعیف را داشته اند (میترا و همکاران ۲۰۰۹).

سوالات طراحی شده بایستی به دنبال آن باشد تا رفتار یادگیری شده مورد ادعای خود را به گونه ای مناسب اندازه گیری نماید (۱۱). و برای این که سوالی بتواند ابزار مفیدی برای هدایت دانشجو باشد، ابتدا باید معلم رابطه بین هدف تدریس و ارزشیابی را دریابد و سپس سوالات را طرح کند چرا که کیفیت ارزشیابی را با توجه به مناسبت داشتن و صحت آن می توان تشخیص داد (سبزواری ۲۰۰۵). در این تحقیق نیز طراحان سوالات امتحانات ارزیابی شده به روایی محتوای و روایی صوری سوالات توجه داشته اند که در بررسی سوالات آزمون ها بیش از ۹۷/۳ درصد سوالات دارای روایی محتوی و در ۹۳/۳ درصد از سوالات دارای روایی صوری بودند که نتیجه گرفته می شود این آزمونها دارای روایی محتوا و روایی

ارزشیابی و طراحی سوال آشنایی دارند، آموزشهای لازم برنامه ریزی شود.

قدردانی

نویسندگان مقاله تشکر خود را از معاونت پژوهشی دانشگاه علوم پزشکی اصفهان که از این تحقیق (باشماره طرح ۲۸۸۰۸۸) حمایت مالی نمودند و نیز اساتید دروسی که سوالات امتحانی، طرح درس و آنالیز سوالات مربوط به خود را در اختیار محققین قرار دادند اعلام می نمایند.

References

Brady, AM 2005, *Assessment of Learning with multiple choice questions*. Nurse Educ pract, vol.5, no.4, Pp.238-42.

Dadgari, A Ebrahimi, H Hassani, MR Mashyekhi, Y 2003, *Analysis of the comprehensive nursing final exam in Shahrod faculty of medical sciences*, Shahrekord University of Medical Sciences Journal, vol.4, No.4, Pp.4-10. [persian]

Heidari, M 2004, *The analysis of the achievement test performed in shahed nursing and midwifery faculty*, Iranian Journal of Medical Education. (suppl 10), P.120. [persian]

Kazemi, A Ehsanpour, S Hassanzadeh, A. 2010, *Investigating the Academic Achievement Evaluation of Specialized Theoretical Courses of Midwifery BS*. Iranian Journal of Medical Education, Vol.9 No.4, Pp.346-354. [persian]

Kazemi, A Ehsanpour, S 2011, *Item Analysis of Core Theoretical Courses Exams for Midwifery Students in Isfahan University of Medical Sciences*. Iranian Journal of Medical Education, Vol.10, No.5, Pp.643-650. [persian]

Khoddam, H Jouybari, L Sanagoo, A 2005, *Meharathaye tarahiye soal*

طراح نیست تنبیه هم بشوند (سیف ۲۰۰۰) و نمره کمتری را کسب نمایند. لذا لازم است طراحان سوال با ارزیابی روایی ملاکی سوالات امتحان خود این نقایص را بر طرف نمایند. و لازمه آن داشتن ارزیابی سوالات امتحانی ضمن داشتن علم کافی برای این ارزیابی است. در حالی که نجار گزارش کرده است که تنها ۱۰٪ اساتید روایی ملاکی و ۷/۵ درصد از آنان روایی صوری سوالات خود را مورد بازنگری قرار می دهند و کمتر از نیمی از آنها از روایی ملاکی و روایی صوری اطلاعات کافی داشته اند (نجار و همکاران ۲۰۰۴).

در نهایت ضمن اعلام بسیاری از مشکلات در ارزشیابی امتحانات، لازم به ذکر است که در پژوهش حاضر نیز محدودیت هایی وجود داشت مثلا فقط تعداد محدودی از امتحانات مورد بررسی قرار گرفت در حالی که برای قضاوت لازم است تحقیقات کاملتری که ارزشیابی مستمر دروس تخصصی نظری پرستاری (امتحانات میان ترم، کوئیزها و...) را نیز مورد بررسی قرار دهد، صورت گیرد و پیشنهاد می شود اقدامات عملی در جهت رفع مشکلات گفته شده انجام شود و نتایج آنها در بهبود فرایند ارزشیابی بررسی شود.

نتیجه گیری

در بررسی سوالات آزمون های دروس تخصصی نظری پرستاری از نظر تعداد سوالات و ضریب تمیز نامناسب بود که میتواند منجر به تمایز نادرست و نتیجه گیری اشتباه از یادگیری دانشجویان شود. این در حالی است که ارزشیابی باید چنان باشد که بتواند سنجش درستی از میزان یادگیری دانشجویان در تمام سطوح آموزشی باشد اما متاسفانه به دلایل مختلف به این مهم توجه نشده بود و بیشتر سوالات آسان و در سطح شناختی طراحی شده بودند. بطور کلی سوالات آزمون های پرستاری در این مطالعه و در بسیاری از مطالعات دیگر کیفیت لازم را ندارند و با توجه به اینکه سوالات امتحان نقشی حیاتی در موفقیت دانشجویان خصوصا در آینده حرفه ایشان ایفا می کند، پیشنهاد می شود کمیته ارزشیابی دانشکده ها بر ارزشیابی تحصیلی دروس تخصصی نظارت مستمر داشته باشند و برای طراحان سوال، که کمتر به اصول

- studying in year Two medical students. Med Educ. Aug, Vol.41, No.8, Pp.754-62.*
- Sabzevari, S 2005, *Metaevaluation and its Standards*, Strides in Development of Medical Education, Vol.2, No.2, Pp. 116-125. [persian]
- Saif, A 2000, *Education psychology (Psychology of learning and instruction)*. 3rd ed. Tehran:Agah [persian]
- Seif, A 2004, *Educational measurement, assessment and evaluation*. 4th ed. Tehran: Doran, Pp.426 - 8 [Persian]
- Shakurnia, A Khosravi, Borojeni, A Mozaffari, A et al 2010, *An Evaluation of Exam Questions Designed by Faculty Members, Emphasizing on the Multiple Choice Question Structure*, Ahvaz Jundishapur University of Medical Sciences 2007, Strides in Development of Medical Education, Vol.6, No.2, Pp.129-38. [persian]
- Stoner, M 1999, *Critical thinking for nursing*. 4th ed. St. Louis: WB Saunders Co.
- Wink, DM 2008, *Using questioning as a teaching strategy*. Nurse Edu, Vol.18, No.5, Pp.11-15.
- tavasote morabiyani baliniye parastari, *Iranian Journal of Medical Education*, Vol.5, No.2, Pp.191-2.
- Larsen, DP Butler, AC Ruediger HL 2008, Test- enhanced learning in medical education. *Med Educ, Vol.42, No.10, Pp.959-66*
- Mc Courbie, P 2004, *Improving the fairness of multiple choice questions: A literature review*. *Med Teach, Vol.26, No.8, Pp.709-712.*
- Mitra, N Nagaraja, H Ponnudurai, G et al 2009, *The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester I, multidisciplinary summative tests*, *IeJSME, Vol.3, No.1, Pp.2-7.*
- Naderi, E Seyf Naraghi, M 1997, *Evaluation and measurement and basic analysis tools in cultural sciences and psychology*, 1st ed. Tehran: Misagh. [persian]
- Najar, SH Abedi, P 2004, *Knowledge and practice of faculty members in assessing the validity and reliability of examinations*. *JQUMS, Vol.8, No.1, Pp.46-50.* [persian]
- Reid, WA Duvall, E Evans, P 2007, *Relationship between assessment results and approaches to learning and*

Validity, Reliability and Difficulty of Multiple-Choice Questions in Specialized Courses of Final Exams in Bachelor Nursing at Isfahan University of Medical Sciences

Habibollah Hoseini¹

Mahsa Shakour^{2*}

Abdollah Rezaei Dehaghani³

¹: Faculty Member, Health & management Department, Nursing and Midwifery faculty, Isfahan University of Medical Sciences, Isfahan, Iran.

^{2*}: Assistant professor, EDC, Arak University of Medical Sciences, Arak, Iran.

³: Faculty Member, Health and management Department, Nursing and Midwifery faculty, Isfahan University of Medical Sciences, Isfahan, Iran.

Abstract: Different evidences have revealed that the quality of exams significantly affects the quality of training and teaching-learning procedures. Hence, the present study with the purpose of evaluating final exams of specialized theoretical courses of Nursing and also showing the true position of exams, was conducted based on standard indicators of classical model of test evaluation. It was a descriptive cross-sectional study which was carried out on final exam' questions of nursing Bachelor' specialized theoretical courses. Collecting data was done by a questionnaire. Some variables in exams were evaluated like: reliability, validity, difficulty, discriminate index, discrepancies, type and level of questions. Accordingly, 209 MC items (56%) was as cognitive level as the Blooms' level, 197 questions were considered as easy one (58.6%), and 121 questions (36%) had the ideal discriminate index. Moreover, the correlation between sessions and the overall number of designed questions for all courses was significant ($r= 0.34$, $P\text{-value} < 0.001$). For bloom's level, 56.2% of all the questions were at knowledge level and 3 (0.8%) at synthesis level. At last, exam questions of nursing major in the present study and many other studies were not of high quality. Therefore, it is recommended that evaluation committees of universities undertake continuous supervision over educational and training plans to be offered for question designers.

Keywords: Nursing Education, Meta Evaluation, Assessment.

***Corresponding author:** Assistant professor, Arak University of Medical Sciences, Arak, Iran.

Email: ms.shakour@gmail.com