

## بررسی کیفیت آزمونهای پایان ترم فراگیران دانشگاه علوم پزشکی بابل

در نیمسال دوم تحصیلی ۱۳۷۸-۷۹

منوچهر اشرفیور\*، زهرا بهشتی، سرور ملوک زاده

عضو هیأت علمی دانشکده پرستاری و مامایی آمل

**سابقه و هدف:** نتایج امتحانات نه تنها برای قضاوت درباره عملکرد فراگیران بلکه درباره اثر بخشی آموزش، روشها و مواد آموزشی نیز مفید هستند. با توجه به گستردگی استفاده از آزمونهای چند گزینه‌ای در مراکز آموزش عالی بر آن شدیم تا کیفیت آزمونهای پایان ترم را مورد مطالعه قرار دهیم.

**مواد و روشها:** در این مطالعه توصیفی - تحلیلی تعداد ۱۴۸۵ پرسش چند گزینه‌ای از ۴۰ موضوع امتحانی دروس تخصصی و پایه پرستاری در نیمسال دوم ۷۸-۷۹ مورد بررسی قرار گرفت و کیفیت هر یک از آنها با استفاده از فرمولهای مناسب مورد تجزیه و تحلیل قرار گرفت.

**یافته‌ها:** بیش از نیمی از پرسش‌ها در کل آزمونهای مورد تحلیل دارای ضریب دشواری بالای ۰/۷ بوده که در طبقه آسان و بسیار آسان قرار می‌گیرند. ۱۱/۴٪ از کل پرسش‌ها در سطح  $a=0/05$  دارای ضریب تمیز معنی‌دار بوده‌اند و ۵/۹٪ از پرسشها واقعاً به عنوان چهار گزینه‌ای عمل کرده، ولی سایر پرسشها عملاً یک یا دو گزینه‌ای بوده‌اند. ۷۶/۲٪ از آزمونها دارای ضرایب پایایی منفی بوده است. تنها ۲۳/۸٪ ضرایب پایایی مثبت داشتند که مقادیر آنها از ۰/۰۳ تا حداکثر ۰/۵ برای یک امتحان متغیر بوده است. **نتیجه‌گیری:** با توجه به نتایج این بررسی، طراحی آزمونهای پایان ترم وضعیت مناسب و راضی‌کننده نداشته و عموماً آزمونهای مورد استفاده در فرآیند ارزشیابی برآیندهای آموزشی دانشجویان دارای اشکال ساختاری اساسی بوده و این آزمون‌ها به عنوان ابزار دقیق و معتبر عمل نمی‌نمایند و ضمناً تفاوت‌های واقعی آزمودنی‌ها را نمی‌سنجند.

**واژه‌های کلیدی:** آزمون، پرسش‌های چند گزینه‌ای، کیفیت، ضریب دشواری، ضریب تمیز.

### مقدمه

هدف‌های آموزشی، روش تدریس، محتوای برنامه آموزشی و ارزشیابی را می‌توان به عنوان ارکان فرآیند تعلیم و تربیت قلمداد کرد. بدون تردید تمامی این پارامترها بر هم تأثیر متقابل دارند ولی در این میان ارزشیابی و امتحان نقش تعیین‌کننده و ممتاز خواهد داشت و به فرض مترقی بودن هدفهای آموزشی، محتوای برنامه و مؤثر بودن روش تدریس،

چنانچه نتایج حاصل از آموزش را بطور علمی ارزشیابی

نکنیم، عملاً به خطا رفته‌ایم(۱). تصمیمات یک ارزیاب در مورد اینکه چه چیزی را بررسی کند، از چه روش و

□ هزینه انجام این پژوهش در قالب طرح تحقیقاتی شماره

۱۳۷۸۲۷ از اعتبارات معاونت پژوهشی دانشگاه علوم پزشکی بابل تأمین

شده است

تمیز آنها در تشخیص دادن آزمودنی های قوی و ضعیف و قدرت هر یک از گزینه‌ها در جلب فراگیران ضعیف مورد بررسی قرار می‌گیرد. قسمت دیگر آمار آزمون است که در آن به ارزیابی اطلاعات کلی آزمونها مانند روایی، پایایی و خطای معیار اندازه گیری سؤالات میپردازد (۸). جهت محاسبه ضریب دشواری از فرمول  $pi = \frac{piU + piL}{2}$  استفاده شده است  $Pi =$  ضریب دشواری سؤال شماره  $i$ ,  $PiU =$  ضریب دشواری سؤال شماره  $i$  برای گروه قوی،  $PiL =$  ضریب دشواری سؤال شماره  $i$  برای گروه ضعیف (۹). بعد از محاسبه ضریب دشواری سؤالات بر حسب مقدار ضریب، به ۵ طبقه تقسیم شدند:

۱= بسیار آسان با ضرایب دشواری ۰/۸۶ تا ۱

۲= آسان با ضرایب دشواری ۰/۷۱ تا ۰/۸۵

۳= متوسط با ضرایب دشواری ۰/۳ تا ۰/۷

۴= دشوار با ضرایب دشواری ۰/۱۲ تا ۰/۲۹

۵= بسیار دشوار برای سؤالات با ضرایب دشواری صفر تا ۰/۱۱ (۸ و ۹).

ضریب تفکیک سؤال که معرف قدرت سؤال در تمیز آزمون شوندگان قوی و ضعیف است، متغیر دیگری است که با فرمول  $DiU = \frac{RiU}{F}$  محاسبه گردید،  $DiU =$  قدرت تفکیک سؤال شماره  $i$  در گروه قوی،  $F =$  تعداد افراد گروه قوی  $TiU =$  تعداد افراد گروه قوی که به سؤال پاسخ صحیح داده اند. قدرت تفکیک سؤال در گروه ضعیف نیز به همین ترتیب محاسبه شد.  $D, D = DiU - DiL =$  قدرت تمیز سؤال شماره  $i$  پس از محاسبه ضریب افتراق سؤالات در ۴ گروه زیر طبقه بندی شدند:

۱= خوب: سؤالات دارای ضرایب افتراق بزرگتر از ۰/۳

۲= متوسط: سؤالات دارای ضرایب افتراق ۰/۱۲ تا ۰/۲۹

۳= ضعیف: سؤالات دارای ضرایب افتراق ۰ تا ۰/۱۱

۴= منفی: سؤالات با ضرایب کوچکتر از ۰ (۸).

برای تعیین معنی دار بودن ضرایب قوه تمیز از آزمون مجذور کای موسوم به گیلفورد استفاده شده است. برای ارزیابی قدرت گزینه‌های انحرافی، پس از تهیه جدول تحلیل

ابزاری استفاده نماید و حتی سوابق شخصی و تجربیات حرفه ای او، همه بر اجرا و نتیجه ارزشیابی تاثیر میگذارند (۲). مهمترین وسیله اندازه گیری پیشرفت تحصیلی، آزمون است (۳) و معمولاً در سطوح عالی از سؤالات چند گزینه ای در آزمونها استفاده میشود (۴) که این آزمونها باید ویژگیهای اساسی مانند اعتبار قابل قبول، روایی مناسب و عملی بودن را داشته باشد و برای این منظور از سنجش پارامترهایی مثل درجه دشواری سؤال، قدرت تمیز سؤال در تشخیص دانشجوی قوی و ضعیف، درصد پاسخ به هر یک از گزینه های انحرافی و صحیح، همراه با اعتبار و خطای استاندارد استفاده میگردد (۵). این حقیقت که معلمان گاهاً در ارزشیابی یادگیری فراگیران ناکام هستند، یک مشکل علمی است. در بسیاری از موارد کلیه دانشجویان ضعیف و قوی صرف نظر از میزان آمادگی، نمرات بالایی در امتحان اخذ می نمایند و یا آنکه از گذراندن موفقیت آمیز آن باز میمانند (۶). اما بدون شک نوع و کیفیت امتحان مستقیماً بر روش تدریس و بطور کلی تر بر اعتبار معلم تأثیر 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 (۷). لذا در این مطالعه سعی شده به این سؤال که آیا امتحانات پایان ترم می‌توانند از عهده این مهم برآیند پاسخ گوئیم.

## مواد و روشها

در این مطالعه توصیفی تحلیلی کلیه آزمونهای پایان ترم دروس تخصصی و پایه پرستاری دانشگاه علوم پزشکی بابل در نیمسال دوم سال تحصیلی ۷۸-۷۹ مورد بررسی قرار گرفتند. اینگونه تحقیقات از دو جنبه یکی ارزیابی تناسب کیفیت سؤالات آزمون با جمعیت آزمودنی ها و دوم پاسخ شرکت کنندگان به تک تک سؤالات، بررسی می‌شود. ضمناً سؤالات از نظر مشخصاتی مانند دشواری سؤالات، قدرت

تابستان ۱۳۸۲

گزینه‌ها، فراوانی انتخاب هر یک محاسبه شد و گزینه‌هایی که بیش از ۵٪ آزمون شونده‌ها را به خود جلب می‌کردند و تعداد بیشتری از افراد گروه ضعیف به آن پاسخ مثبت می‌دادند دارای قدرت مطلوب، تشخیص داده می‌شدند. پایایی یا از روش a - کرانباخ، است استفاده شد. معیار اندازه گیری، انحراف معیار توزیع نمرات خطا است که با ضریب اعتبار رابطه معکوس دارد و برای محاسبه آن از فرمول زیر استفاده شده است:  $S.E=S \sqrt{1-r}$ ،  $S$  = انحراف معیار نمره ها،  $r$  = ضریب پایایی آزمون (۱۰).

### یافته ها

در این مطالعه ۴۰ عنوان از دروس تخصصی و پایه دوره‌های مختلف پرستاری مورد بررسی قرار گرفت و از میان آنها ۷۲/۵٪ آزمونها، مربوط به دروس اختصاصی و ۲۷/۵٪ نیز مربوط به درسهای پایه بودند. ۵۵/۵٪ از مدرسین دارای درجه علمی کارشناسی ارشد، ۵/۶٪ دارای PhD و ۲۷/۷٪ پزشک متخصص بودند. ۶۱/۱٪ مدرسین مرد و ۳۸/۹٪ زن بودند. ۶۱/۱٪ از اساتید، عضو هیئت علمی دانشگاه و ۳۸/۹٪ غیر هیئت علمی و حق التدریس بوده اند. ۱۲/۵٪ از آزمونها فقط متشکل از پرسشهای چند گزینه ای و

همسانی درونی آزمون، به دقت اندازه گیری و ثبات ابزار سنجش مربوط است و جهت محاسبه ضریب پایایی یا اعتبار آزمونها از فرمول شماره ۲۰ کودر- ریچاردسون که فرم خاصی

۸۷/۵٪ حاوی پرسشهای مختلط تشریحی و چند گزینه ای بوده اند. نتایج نشان داد که تنها ۳۸/۵٪ از کل پرسشها دارای درجه دشوار ببین (۰/۷-۰/۳) بودند. بیشترین فراوانی (۵۷/۲٪) به سؤالات ساده و بسیار ساده مربوط می‌شود و پرسشهای دشوار و بسیار دشوار کمترین فراوانی (۵٪) را دارا بودند (جدول ۱). در مورد قدرت تمیز سؤالات نتایج نشان می‌دهد که تنها ۱۸/۱٪ کل پرسشهای مورد مطالعه واجد ضریب تمیز بالاتر از ۰/۳ بوده و ۳۰/۱٪ از پرسشها دارای قوه تمیز متوسط بوده اند. ۵۱/۸٪ از پرسشها از قوه تمیز ضعیف و منفی برخوردار بودند. طبق فرمول ۲۰ کودر ریچاردسون، ۷۶/۲٪ از آزمونهای مورد مطالعه دارای ضرایب پایایی منفی بوده‌اند (جدول ۱). میانگین خطای معیار آزمونهای مورد مطالعه با اطمینان ۹۵٪، نمره آزمونها در فاصله ۲/۳+ مشاهده شد (جدول ۱). همچنین تنها ۵/۹٪ از پرسشها دارای سه گزینه انحرافی مناسب و ۹۴/۱٪ پرسشها در واقع کمتر از چهار گزینه داشتند (جدول ۲).

جدول ۱. میانگین ضرایب شاخصهای کیفیت آزمونها برحسب موضوع درسی در دانشگاه علوم پزشکی بابل، ۷۹-۱۳۷۸

موضوع درس	متغیر	سطح دشواری	قوه تمیز سوال	ضریب پایانی	خطای معیار
پرستاری داخلی - جراحی	۰/۶۳	۰/۱۴	-۳/۹۵	۲/۵۸	
پرستاری بهداشت جامعه	۰/۷۷	۰/۱۳	۰/۲	۱/۵۶	
پرستاری بهداشت مادر و کودک	۰/۷۸	۰/۱۴	-۰/۵۵	۲/۰۵	
پرستاری کودکان	۰/۷۲	۰/۱۴	-۱/۱۵	۲/۲۵	
روان پرستاری	۰/۷۳	۰/۲۱	-۰/۴	۲/۵	
پرستاری ویژه	۰/۵۸	۰/۱۳	-۰/۱	۲/۵	
اصول و فنون پرستاری	۰/۷۲	۰/۱۳	-۰/۸	۲/۱	

۲	-۲/۷	۰/۰۹	۰/۷۳	فوریتهای پرستاری
۱/۷	۰/۵	۰/۲۲	۰/۷	مدیریت پرستاری
۲	۰/۲	۰/۱۶	۰/۷۷	بررسی وضعیت سلامت
۲/۶	-۰/۹	۰/۱۸	۰/۷۶	ایمونولوژی
۳	۰/۹	۰/۱۶	۰/۶۴	اپیدمیولوژی
۲/۵	-۲	۰/۱۳	۰/۷۸	تغذیه

جدول ۲. فراوانی قدرت گزینه های انحرافی سوالات دروس پرستاری برحسب موضوع درسی در دانشگاه علوم پزشکی بابل، ۱۳۷۸-۷۹

موضوع درس	انحرافی قوی			متغیر
	بدون گزینه	گزینه ۱	گزینه ۲	
پرستاری داخلی - جراحی	۲۰/۸	۴۴/۱۷	۳۰/۱۸	۴/۸۵
پرستاری بهداشت جامعه	۳۹/۴	۳۷/۴۷	۱۷/۶	۱۶/۶
پرستاری بهداشت مادر و کودک	۳۱/۵۵	۴۲/۷	۲۰/۷۵	۵
پرستاری کودکان	۲۵/۶۵	۴۶/۱۵	۲۲/۱	۶/۱
روان پرستاری	۱۳/۱۵	۴۳/۶	۳۰/۵۵	۱۲/۷
پرستاری ویژه	۳۰/۳	۲۱/۲	۴۲/۴	۶/۱
اصول و فنون پرستاری	۲۵/۸	۴۲/۳	۲۸/۳	۳/۶
فوریتهای پرستاری	۲۶/۶	۴۶/۲	۲۷/۲	۰
مدیریت پرستاری	۲۶/۴	۲۶/۳	۳۱/۶	۱۵/۷
بررسی وضعیت سلامت	۲۰/۵	۴۳/۴	۳۱/۱	۵
ایمونولوژی	۲۰/۹	۴۲/۵	۳۰/۲	۴/۴
اپیدمیولوژی	۱۶	۴۸	۲۹	۷
تغذیه	۲۶/۹	۴۵/۲	۲۶/۷	۱/۲

### بحث

پایه و اساس ارزشیابی درست را اندازه گیری تشکیل میدهد. مهمترین وسیله اندازه گیری پیشرفت تحصیلی، آزمون است و هر آزمون باید از ویژگیهایی مانند عینیت، سهولت اجرا، اعتبار و پایایی برخوردار باشد (۳). با اینکه امتحانات پایان ترم دانشجویان، ملاک بوده و کسانی

که در این امتحانات قبول می شوند و به ترم تحصیلی بالاتر ارتقاء می یابند، باید حداقل دانش و مهارت لازم برای آن را دارا باشند ولی غالباً فاقد کارایی و دقت مورد نیاز آزمونهای کیفی به عنوان یک ابزار سنجش و اندازه گیری می باشند. در این مطالعه متأسفانه دروس تخصصی و پایه پرستاری در داشتن سوالات با درجه دشواری مناسب، بصورت ناموفق عمل کرده اند. یعنی از ۱۴۸۵ سؤال، تنها ۳۸/۵٪ دارای درجه

دشواری قابل قبول (۰/۷-۰/۳) بوده است که در این صورت با طرح پرسش‌های ساده فراگیران را بی‌نیاز از مطالعه و تلاش مناسب برای یادگیری جهت گذراندن موفقیت آمیز امتحان می‌سازند و استمرار بی‌رویه این شیوه، نتایج زیانباری را متوجه جامعه خواهد ساخت. آزمونهای ارتقاء رشته‌های تخصصی پزشکی در تیر ماه سال ۷۷ نیز نشان داده است که ۳۶/۳٪ از کل پرسشها دارای ضرایب دشواری آسان و ۱۷/۳٪ دشوار بوده‌اند (۸) و در آزمون دکتری پرستاری اکثر سؤالات در محدوده مناسب درجه دشواری قرار داشته‌اند (۱۱).

در مورد قدرت تمیز سؤالات که نشانه‌ای از توانایی هر سؤال در تمایز میان دانشجویان قوی و ضعیف بر اساس میزان تسلط شان به موضوع است (۹)، مطالعه نشان داد که پرسشهای امتحانات مذکور فاقد این روایی بوده است. یعنی این سؤالات طراحی مناسبی نداشته و باید کنار گذاشته شوند یا در آنها تجدید نظر صورت گیرد. در مطالعه ای که توسط دبیرخانه شورای آموزش پزشکی تخصصی کشور انجام گردیده، میانگین قدرت تفکیک سؤالات در آزمونهای ارتقاء سال ۷۷ در ۱۶ رشته، ۰/۴۵ بوده است (۸) صبوری نیز در مطالعه خود قدرت تمیز منفی نیافته است و اینرا از شیوه طراحی مناسب سؤالات میدانند (۵). ولی در این مطالعه میانگین قدرت تمیز پرسشها ۰/۱۴ بود و تنها ۱۱/۴٪ از پرسشها در سطح  $a = 0/05$  دارای ضریب تمیز معنی دار بوده و با کل آزمون همبستگی داشته است. یقیناً این پرسشها دارای معایب اساسی هستند و اگر در بانک سؤال اساتید وجود دارند، نیاز به تجدید نظر بنیادین دارند.

در مورد سؤال دیگر پژوهش که قدرت گزینه‌های انحرافی را در جلب فراگیران بررسی می‌نماید باید ذکر شود که تنها درصد اندکی از پرسشها واقعاً چهار جوابی بوده‌اند و بقیه عملاً بصورت دو یا سه گزینه‌ای عمل نموده‌اند که در این صورت احتمال انتخاب تصادفی گزینه صحیح بسیار زیاد است زیرا هر گزینه باید دست کم یک نفر از افراد

گروه ضعیف را بخود جلب نماید و اگر هر دو گروه قوی و ضعیف شامل این انتخاب غلط شدند، تعداد افراد گروه ضعیف باید بیشتر از تعداد افراد گروه قوی باشد (۹). در مطالعه دبیرخانه شورای آموزش پزشکی و تخصصی ۳۶/۳٪ از پرسشها عملاً چهار گزینه‌ای بوده‌اند و در آنها سه گزینه انحرافی در جلب داوطلبان ناآگاه بدرستی عمل می‌نمودند و ۵۲/۳٪ از سؤالات دارای دو گزینه انحرافی مناسب بودند (۸) که در مقایسه با این مطالعه، ۴۱/۹٪ از سؤالات تنها واجد یک گزینه انحرافی بوده‌اند و سایر گزینه‌ها نیز به دلیل ضعف ساختاری فاقد اعتبار بوده‌اند و این نشان از عدم دقت در طرح گزینه‌های انحرافی است.

مسئله دیگری که در این مطالعه مورد بحث قرار گرفته، اعتبار آزمون است و معنی آن این است که سؤالات آزمون تا چه اندازه با یکدیگر همبستگی متقابل دارند. اگر ضریب اعتبار درونی آزمون کم باشد بدان معناست که سؤالات مختلف آزمون متغیر واحدی را اندازه‌گیری نمی‌کنند و آزمون دارای اشکال است (۹). در این مطالعه اکثریت آزمونهای (۷۶/۲٪) دارای ضرایب پایایی منفی بودند و در آزمونهای دانشنامه تخصصی شهریور سال ۷۷ ضرایب همبستگی ۰/۸۳ تا ۰/۹۵ بوده است. آزمونهای ارتقاء تیر ماه نیز به همین سبک بوده است. اگر آزمون دارای همبستگی بالای ۰/۹ باشد از پایایی خوبی برخوردار است (۸). متأسفانه مطالعات اندکی در راستای موضوع مورد بررسی انجام شده است. ولی در پژوهشی که در سال ۱۳۸۰ در مورد آگاهی اعضاء هیئت علمی در مورد روشهای سنجش اعتبار و پایایی سؤالات امتحانی انجام شد به این نتیجه رسیدند که ۴۲/۹٪ از اعضا از پایایی به روش دو نیمه کردن ۲۴/۲٪ با روش کودر ریچاردسون آشنایی داشتند. در خصوص علت عدم سنجش آنها از روش بازآزمایی برای تعیین پایایی سؤالات امتحانی، ۳۴/۲٪ ذکر کردند که با نحوه تعیین اعتبار و پایایی آشنایی ندارند و در مجموع این مطالعه نشان داد که اگر چه آگاهی اساتید از نحوه سنجش اعتبار و پایایی سؤالات متوسط بود

ارتقاء سه رقمی نیز بوده اند ولی در امتحانهای مرسوم ما نمرات بر مبنای ۲۰ است. لذا برای قضاوت نهایی در مورد هر آزمون باید از شاخصهای دیگر نیز به همراه نتیجه امتحان استفاده نمود. صبوری در مطالعه خود با توجه به میزان اعتبار و خطای استاندارد اندازه گیری به این نتیجه رسید که آزمونهای مورد مطالعه بخوبی و بطور دقیق توانسته اند میزان دانش نظری شرکت کنندگان را ارزیابی نماید و به عبارت دیگر آزمون معتبر بوده است (۵).

از آنجائیکه طراحی آزمونهای پایان ترم دانشجویان، که بعنوان آزمونهای قدرت و ملاکی عمل می نمایند، از وضعیت مطلوبی برخوردار نبوده اند و عموماً در امتحانات از پرسشهای سطحی و کم مایه برای ارزشیابی دانشجویان استفاده می شود، لذا با افت علمی که در سالهای اخیر با آن مواجهیم، ضرورت تجدید نظر کلی در شیوه های تدریس و مواد آموزشی احساس می شود و در این رهگذر استراتژی اجتناب از پیروی قاطع از روشهای سنتی ارزشیابی و ایجاد تحول در روشهای ارزشیابی و امتحان می تواند نقشی تعیین کننده در بهبود وضع آموزش ایفا نماید، یقیناً اصلاح کیفی آموزش بدون تغییر رویه ها محقق نخواهد شد. بنابراین پیشنهاد می شود که تحقیقات بیشتری در حوزه آموزش فراگیران صورت گیرد و پس از هرامتحان آزمونهای آماری بر روی تک تک سوالات توسط طراحان سؤال انجام گیرد.

### تقدیر و تشکر

از تلاش پرسنل آموزش پرستاری دانشگاه علوم پزشکی بابل و آمل، جهت جمع آوری اوراق امتحانی و مشاوره های بی دریغ جناب آقای دکتر محمود حاج احمدی که نهایت همکاری را مبذول داشته اند، تقدیر و تشکر بعمل می آید.

اما عملکرد آنها در این مورد ضعیف است (۳). اگر پایایی را ویژگی مهم یک آزمون بدانیم با توجه به نتایج بدست آمده از آزمونهای پایان ترم چنین بر می آید که این آزمونها دقیق، قابل اتکا و معتبر نبوده و اختلافات واقعی میان آزمودنی ها را نمی سنجدند. موضوع دیگر اینکه پایایی شرط لازم برای روایی آزمون است (۹). بنابراین آزمونهای مورد تحلیل هم از نظر روایی و هم از نظر پایایی دارای اشکال هستند.

میانگین خطای معیار آزمونها همانند ضریب پایایی شاخصی از دقت اندازه گیری آزمون است و میانگین خطا در اندازه گیری با یک آزمون را نشان می دهد و هر چه بیشتر باشد اطمینان ما نسبت به دقت اندازه گیری و ارزیابی کمتر است. بعبارت دیگر اگر خطای آزمون کم باشد می توان گفت که نمره بدست آمده به نمره واقعی نزدیکتر است و چون خطای اندازه گیری از توزیع نرمال تبعیت می کند، لذا می توان انتظار داشت که با اطمینان ۶۸٪ (یا ۲ از ۳)، آزمودنی هایی که نمره واحدی را کسب نموده اند، نمره واقعی آنان در فاصله ۱- تا ۱+ خطای معیار اندازه گیری، و با اطمینان ۹۵٪ در فاصله ۲- تا ۲+ خطای معیار اندازه گیری از نمره مشاهده شده است. خطای معیار اندازه گیری در آزمونهای ارتقاء ۱۶ رشته در سال ۷۷، متوسط ۵ بوده است (۸). در صورتیکه خطای معیار بدست آمده در مطالعه حاضر ۲/۳ بوده است یعنی اگر نمره دانشجویی در یک آزمون ۱۲ باشد و نمره خطای آن ۲ باشد در این صورت با اطمینان ۹۵٪ نمره واقعی ۱۲- تا ۱۲+ خواهد بود. علت تفاوت قابل توجه میانگین خطای معیار اندازه گیری آزمونهای ارتقاء و آنچه در این مطالعه بدست آمده با توجه به اینکه نتایج، حاکی از کیفیت پایین آزمونهای مورد تحلیل بوده است، به اختلاف انحراف معیار آنها مربوط است، چرا که نمره ها در آزمونهای

\*\*\*\*\*

## منابع

۱. نادری ع، سیف نراقی م. سنجش و اندازه گیری و بنیاد های تحلیلی ابزارهای آن در علوم تربیتی و روانشناسی. تهران، انتشارات معین ۱۳۷۵؛ ص: ۴۴.
  ۲. ادیب حاج باقری م. ارزشیابی یک ارزشیابی. آموزش در علوم پزشکی ۱۳۸۱؛ ۷: ص: ۹.
  ۳. نجار ش، عابدی پ. بررسی آگاهی و عملکرد اعضاء هیئت علمی دانشگاه علوم پزشکی اهواز در مورد روشهای سنجش اعتبار و پایایی سؤالات امتحانی سال ۱۳۸۰. آموزش در علوم پزشکی، ۱۳۸۱؛ ۷: ص: ۱۱۹-۲۰.
  ۴. منصوریان ح ر و همکاران. بررسی اعتبار صوری سؤالات امتحانی اعضاء هیئت علمی دانشگاه علوم پزشکی شهید صدوقی یزد و تطابق آن با اهداف ویژه طرح درس مربوطه در سال ۱۳۸۱. آموزش در علوم پزشکی ۱۳۸۱؛ ۷: ص: ۱۱۵-۱۶.
  ۵. صبوری م و همکاران. بررسی مشخصه های روان سنجی آزمون بیست و نهمین دوره پذیرش دستیاری در سال ۱۳۸۰ بر اساس نظریه تستی کلاسیک (CTT). آموزش در علوم پزشکی، ۱۳۸۱؛ ۷: ص: ۸۲-۳.
6. Schotanus C. Student assessment and examination rules. Medical Teacher 1999;21(3):318-21.
۷. شعبانی ح. مهارتهای آموزشی و پرورشی، تهران، انتشارات سمت، ۱۳۷۳؛ ص: ۳۹۴ و ۴۱۵.
  ۸. خبرنامه های شماره ۴-۲ دبیرخانه آموزش پزشکی و تخصصی وزارت بهداشت، درمان و آموزش پزشکی ۱۳۷۷؛ ص: ۱۹-۶ و ۵-۹ و ۶-۴.
  ۹. سیف ع ا. روشهای اندازه گیری و ارزشیابی آموزش، چاپ سوم، تهران، نشر دوران، ۱۳۷۹، ص: ۲۸۱ و ۲۸۳ و ۲۸۷ و ۴۰۹ و ۴۲۱-۶.
  ۱۰. پاشا شریفی ح. اصول روان سنجی و روان آزمایی، تهران، انتشارات رشد، ۱۳۷۴، ص: ۵۰-۲۰۰.
  ۱۱. پرویزی س، پیروی ح. ارزشیابی آزمون دکترای پرستاری. آموزش در علوم پزشکی، ۱۳۸۱؛ ۷: ص: ۳۶.

---

\* آدرس نویسنده مسئول: آمل دانشکده پرستاری و مامایی دانشگاه علوم پزشکی بابل، تلفن: ۰۱۲۱-۲۲۲۱۹۱۷-۱۹