

مقایسه آزمونهای ۴ و ۲ گزینه‌ای (صحیح-غلط) در ارزشیابی پیشرفت تحصیلی دانشجویان در درس فیزیک پزشکی

دکتر محسن حاجی‌زاده

گروه فیزیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد

چکیده

مقدمه

ارزشیابی پیشرفت تحصیلی دانشجویان در دروس علوم پایه اغلب توسط آزمون‌های ص-غ و یا چندگزینه‌ای انجام می‌شود. مقایسه این دو نوع آزمون از جهات مختلف با یکدیگر، می‌تواند کمک مؤثری به اساتید در طراحی سؤالات امتحانی و اطمینان از نحوه ارزشیابی پیشرفت تحصیلی دانشجویان در هر یک از موضوعات درسی باشد. برای تحلیل سؤالات و ارزشیابی آزمون می‌توان از شاخص‌های ضریب دشواری (و یا واریانس نمرات) و ضریب تمیز سؤالات و همچنین ضریب پایایی آزمون استفاده نمود.

روش کار

به منظور مقایسه دو شیوه آزمون ۴ و ۲ گزینه‌ای (ص-غ) در ارزشیابی پیشرفت تحصیلی دانشجویان در درس فیزیک پزشکی، از تحلیل سؤالات و نمرات در امتحان پایان ترم دو گروه دانشجویان پزشکی، ۷۶ نفر، و دندانپزشکی، ۷۴ نفر، که در ترم دوم سال تحصیلی ۸۶-۸۵ این درس را انتخاب نموده‌اند؛ استفاده شد. برای دانشجویان پزشکی ۴۰ سؤال ۴ گزینه‌ای از مباحث مشخصی طراحی شد. با استفاده از عبارات استفاده شده در گزینه‌های هر یک از سؤالات، یک تا ۲ سؤال ۲ گزینه‌ای (ص-غ)، جمعاً ۶۰ سؤال برای دانشجویان دندانپزشکی طراحی شد.

نتایج

نمرات امتحانی دانشجویان و نیز درصد نمرات مکتسبه آنان از هر سؤال برحسب شماره سؤال‌ها به همراه ضرائب سختی، واریانس نمرات و ضرائب تمیز سؤالات و نیز ضریب پایایی آزمون با دو روش دو نیمه کردن اسپیرمن-براون و روش KR۲۰ کودر-ریچاردسون محاسبه و ارائه شده است. به منظور مقایسه هر یک از ضرائب فوق با مقادیر بهینه آن‌ها، درصد ضرائب بهینه نیز نشان داده شده است.

بحث

در درس فیزیک پزشکی سؤالات امتحانی ۴ و ۲ گزینه‌ای ص-غ می‌تواند از نظر سختی یکسان طراحی گردد، ولی در شرایط یکسان سؤالات ۴ گزینه‌ای تا حدود ۴۰٪ دارای ضریب تمیز بالاتری بوده و بهتر از سؤالات ۲ گزینه‌ای ص-غ می‌تواند دانشجویان را ارزیابی نماید. همچنین ضریب پایایی به روش KR۲۰ کودر-ریچاردسون در امتحان با سؤالات ۴ گزینه‌ای به میزان ۱۳/۷٪ بیشتر از امتحان با سؤالات ۲ گزینه‌ای می‌باشد. با این وجود ضروری است که در هر یک از سؤالات ۴ گزینه‌ای، گزینه‌های مختلف با دقت طراحی شده و از انتخاب گزینه‌هایی که سبب اشتباه اغلب دانشجویان می‌گردد پرهیز شود.

مقدمه:

در ارزشیابی پیشرفت تحصیلی معمولاً برای ارزیابی حوزه‌های شناختی از آزمون‌های کتبی و در حوزه‌های مهارتی از آزمون‌های عملکردی استفاده می‌شود. در آزمون‌های کتبی نیز اغلب از آزمون‌های عینی استفاده می‌گردد. در این آزمون‌ها (شامل آزمون‌های ص-غ، جورکردنی و چندگزینه‌ای) سؤال و جواب در اختیار آزمون شونده قرار می‌گیرد و از او خواسته می‌شود تا در باره جواب صحیح تصمیم بگیرد. با توجه به این که در ارزشیابی پیشرفت تحصیلی دانشجویان در دروس علوم پایه بیشتر از آزمون‌های ص-غ و چندگزینه‌ای استفاده می‌شود، در نظر است این دو نوع آزمون از جهات مختلف با یکدیگر مقایسه شوند.

اگرچه استفاده از آزمون‌های چندگزینه‌ای در همه جا اثربخشی ندارد و در مواردی بهتر است از آزمون‌های دیگر استفاده شود، مثلاً برای محاسبات ریاضی و علوم از آزمون‌های پاسخ کوتاه (جای خالی) و برای سنجش واقعیت‌ها استفاده از آزمون‌های ص-غ توصیه می‌شود و یا در مواردی که تهیه گزینه‌های مناسب انحرافی دشوار است اما تعدادی مطالب متجانس وجود دارد، بهتر است از آزمون‌های جورکردنی استفاده نمود. ولی به هر حال این گونه می‌توان نتیجه‌گیری نمود که سؤال چندگزینه‌ای کاربرد زیادی دارد؛ اما همیشه بهترین نوع سؤال نیست (۲). از طرفی هیچ نوع آزمون واحدی وجود ندارد که به خوبی از عهده سنجش همه هدف‌های سطوح مختلف یادگیری برآید. بنابراین بهتر است برای ارزشیابی دقیق‌تر پیشرفت تحصیلی دانشجویان، آزمون‌های ترکیبی را که شامل سؤال‌های مختلف تشریحی، عینی و کوتاه پاسخ می‌باشند، مورد استفاده قرارگیرد یا در طول ترم چند آزمون مختلف به عمل آورده که هر یک از آنها شامل نوع بخصوصی از سؤال‌ها باشند (۱).

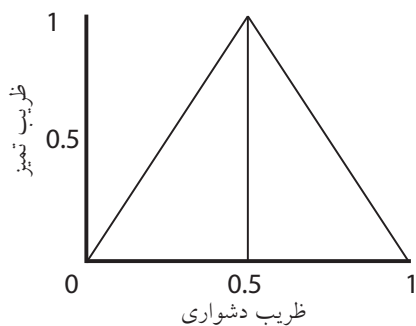
در طراحی سؤال‌ات و تعیین مدت برگزاری آزمون‌های پیشرفت تحصیلی این نکته باید همواره مورد توجه باشد که این آزمون‌ها از نوع آزمون قدرت می‌باشند و باید وقت کافی به آزمون شونده‌گان داده شود (زمانی که ۷۵٪ آزمون شونده‌گان آزمون را به پایان برسانند) تا بدون عجله و رقابت بتوانند گزینه‌های درست را انتخاب نمایند (۳).

در هر آزمون برای تحلیل سؤال‌ات می‌توان از شاخص‌های ضریب دشواری (و یا واریانس نمرات) و ضریب تمیز سؤال‌ات استفاده نمود. ضریب دشواری برای هر سؤال (p) به صورت نسبت تعداد کل افرادی که به آن سؤال جواب درست داده‌اند، به کل آزمون شونده‌گان (یا نسبت تعداد پاسخ‌های درست در گروه‌های پائین و بالا به جمع افراد گروه‌های مذکور) تعریف می‌شود. در شرایط توزیع به‌هنگار نمرات معمولاً ۲۷٪ برگه‌ها برای هر یک از گروه‌های پائین و بالا انتخاب می‌شود ولی در سایر شرایط، این گروه‌ها از ۲۵ تا ۳۳٪ آزمون شونده‌گان را تشکیل می‌دهند. مقدار این ضریب برای هر سؤال بین ۰-۱ تغییر می‌نماید. مقدار صفر نشان دهنده

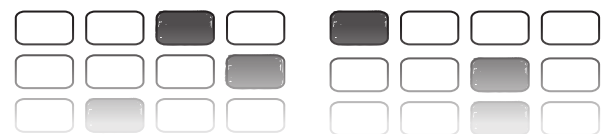


سختی سؤال بوده و بیان‌گر آن است که هیچ یک از افراد در گروه‌های بالا و پائین به سؤال پاسخ درست نداده‌اند و مقدار ۱ برعکس بیانگر سادگی سؤال می‌باشد. برای تحلیل سؤال‌ات برخی به جای این ضریب از واریانس نمرات استفاده می‌نمایند. واریانس هر سؤال از رابطه $p(1-p)$ به دست می‌آید. مقدار آن برای سؤال‌ات بسیار سخت ($p=0$) و بسیار آسان ($p=1$) برابر با صفر و برای سؤال‌ات با سختی متوسط ($p=0.5$) برابر با ۰/۲۵ می‌باشد. سطح بهینه دشواری در آزمون‌های چندگزینه‌ای اندکی کم‌تر از وسط فاصله بین ۱ و سطح موفقیت با حدس زدن تعریف می‌شود. سطح موفقیت با حدس زدن برابر با نسبت پاسخ‌های درست با حدس زدن در هر سؤال است (Allen et al). مثلاً برای سؤال‌ات ۴ گزینه‌ای با احتساب ۰/۲۵ نمره منفی سطح بهینه دشواری برابر ۰/۶۲۵ و یا برای سؤال‌ات ص-غ با ۱/۳ نمره منفی برابر ۰/۷۵ می‌باشد.

ضریب تمیز سؤال‌ات (d) به صورت نسبت تفاضل پاسخ‌های درست در گروه‌های بالا و پائین به تعداد افراد یک گروه (بالا یا پائین) تعریف می‌شود. ماکزیمم این ضریب برای هر سؤال ۱ و مقادیر کمتر از آن به معنی کاهش توان تشخیص پاسخ صحیح توسط افراد در گروه‌های ضعیف و قوی می‌باشد. مقدار صفر برای این ضریب به معنی عدم امکان تشخیص پاسخ صحیح توسط افراد در گروه‌های ضعیف و قوی و مقادیر منفی به معنی وجود اشکال در سؤال و یا آموزش غلط مطلب درسی می‌باشد. ارتباط ضرایب دشواری و تمیز سؤال‌ات و مقادیر بهینه آن‌ها را از مثلث زیر می‌توان به دست آورد.



در آزمون‌های وابسته به هنجار هرچه واریانس نمرات گروهی در آزمون و همبستگی سؤال‌ات (ضریب تمیز) بیشتر باشد، آزمون بهتر است. به عبارت دیگر سؤال‌های خوب در یک آزمون، دارای واریانس بالا (و یا ضریب دشواری متوسط) و ضریب تمیز بالایی هستند.



روش کار:

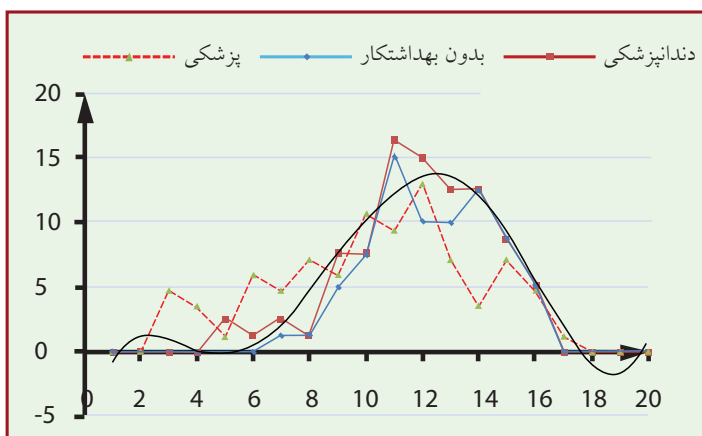
به منظور مقایسه دو شیوه آزمون ۴ و ۲ گزینه‌ای (ص-غ) در ارزشیابی پیشرفت تحصیلی دانشجویان در درس فیزیک پزشکی، از تحلیل سؤالات و نمرات امتحانی پایان ترم دانشجویان پزشکی و دندانپزشکی که در ترم دوم سال تحصیلی ۸۵-۸۶ این درس را انتخاب نموده‌اند، استفاده شد. در انتخاب دانشجویان مذکور فرض بر این است که تمامی دانشجویان از نظر استعداد تحصیلی و ضریب هوشی یکسان بوده و صرفاً بر مبنای علاقه شخصی از طریق کنکور سراسری به دو گروه پزشکی و دندانپزشکی تقسیم شده‌اند.

با سؤالات ۴ و ۲ گزینه‌ای (ص-غ) از مباحث یکسانی که توسط مدرس واحدی و به روش مشابهی به دانشجویان مذکور تدریس شده بود، امتحان کتبی به عمل آمد. برای دانشجویان پزشکی ۴۰ سؤال ۴ گزینه‌ای طراحی شد و سپس با استفاده از عبارات مشابه با عبارات استفاده شده در گزینه‌های هر یک از سؤالات، یک تا ۲ سؤال ۲ گزینه‌ای (ص-غ)، جمعاً ۶۰ سؤال برای دانشجویان دندانپزشکی طراحی گردید. امتحان دو گروه به فاصله ۲۰ روز از یکدیگر و توسط اداره آموزش دانشکده پزشکی تعیین شده بود. دانشجویان پزشکی ۷۶ نفر و دانشجویان دندانپزشکی ۷۴ نفر که ۱۳ نفر آن از بهداشتکاران دهان و دندان بودند. سؤالات ۴ جوابی با ۰/۲۵ نمره منفی و سؤالات ص-غ با ۰/۳۳ نمره منفی تصحیح گردید.

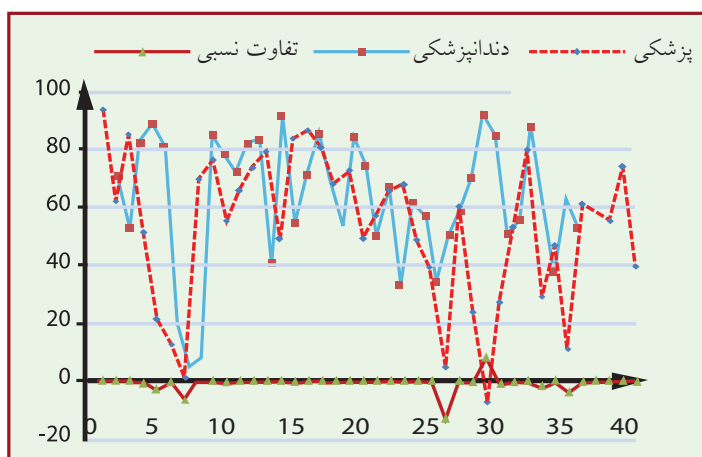
نتایج:

میانگین نمرات خام برای دانشجویان پزشکی ۱۰/۵ و برای دانشجویان دندانپزشکی ۱۲/۳ می‌باشد. میانگین نمرات دانشجویان دندانپزشکی بدون احتساب نمرات دانشجویانی که با سهمیه بهداشتکاران دهان و دندان پذیرفته شده‌اند، ۱۲/۵ می‌باشد.

توزیع نمرات دانشجویان و نیز درصد نمرات مکتسبه دانشجویان از هر سؤال برحسب شماره سؤالات ۴ گزینه‌ای در نمودارهای زیر ترسیم شده است. نمرات مکتسبه مربوط به ۶۰ سؤال دندانپزشکی با توجه به ارتباط موضوعی آن‌ها در گزینه‌های سؤالات ۴ گزینه‌ای به ۴۰ سؤال تجمیع و در نمودار شکل ۲ ترسیم شده است.



نمودار ۱ توزیع نمرات خام دانشجویان در درس فیزیک پزشکی



نمودار ۲ درصد نمرات مکتسبه دانشجویان از هر سؤال

ضرائب سختی، واریانس نمرات و ضرائب تمیز سؤالات و نیز ضریب پایایی آزمون با دو روش ۲ نیمه کردن اسپیرمن-براون و روش KR۲۰ کودر-ریچاردسون محاسبه و در جدول زیر نشان داده شده است. در این جدول به منظور مقایسه هر یک از ضرائب فوق با مقادیر بهینه آنها، درصد ضرائب بهینه نیز درج گردیده است.

ضریب پایایی به روش	متوسط ضریب سختی سؤالات	% ضریب سختی بهینه	متوسط واریانس نمرات	% واریانس بهینه نمرات	متوسط ضریب تمیز سؤالات	% ضریب تمیز بهینه	نیم کردن اسپیرمن-براون	KR۲۰ کودر-ریچاردسون	
									ضریب پایایی به روش
دندانپزشکی (ص-غ)	با بهداشتکاران	۰,۶۹ ± ۰,۲۲	۹۲	۰,۱۷ ± ۰,۰۶	۸۹	۰,۲۱ ± ۰,۱۸	۳۵	۰,۸۱۴	۰,۸۸۲
	بی بهداشتکاران	۰,۷۰ ± ۰,۲۲	۹۳	۰,۱۷ ± ۰,۰۷	۸۸	۰,۲۰ ± ۰,۱۸	۳۳	۰,۸۱۲	۰,۸۸۹
پزشکی (۴ جوابی)		۰,۵۷ ± ۰,۲۰	۹۲	۰,۲۱ ± ۰,۰۵	۸۸	۰,۳۸ ± ۰,۱۵	۴۸	۰,۸۲۲	۱,۰۰۶

ب-ح-ث:

از مقایسه نمرات دانشجویان دندانپزشکی در گروه‌های با و بدون بهداشتکاران در نمودار ۱ ملاحظه می‌شود که متوسط نمرات دانشجویانی که با سهمیه بهداشتکار دهان و دندان پذیرفته شده‌اند، کمتر از متوسط نمرات سایرین می‌باشد.

نمودار ۲ بیانگر آن است که در برخی از سؤالات، دانشجویان پزشکی (سؤال ۴ جوابی) به طور قابل توجهی نمره کمتری از دانشجویان دندانپزشکی (سؤال ص-غ) کسب نموده‌اند و این بدان معناست که در این سؤالات گزینه‌های اضافی موجب گمراهی دانشجویان شده است. در مورد سؤالات ۷ و ۲۶ و ۲۹ و ۳۵ که تفاوت نسبی نمرات کسب شده بیشتر از سایر سؤالات است، با مراجعه به ضرائب تحلیلی و متن سؤالات مشخص شد که این موضوع به جز در موردی که به دلیل سختی سؤال هر دو گروه نمره بسیار کمی کسب نموده‌اند، در سایر موارد گزینه‌های اضافی سبب تردید دانشجویان و در نهایت اشتباه در انتخاب گزینه صحیح در سؤالات ۴ گزینه‌ای شده است. این موضوع در بحث دشواری انتخاب گزینه‌های انحرافی به عنوان یکی از معایب آزمون‌های چند گزینه‌ای اشاره شده است (۴). بنابراین ضروری است که در هر یک از سؤالات ۴ گزینه‌ای، گزینه‌های مختلف با دقت طراحی شده و از انتخاب گزینه‌هایی که احتمال اشتباه دانشجویان قوی را به دلیل اشکالات و پیچیدگی‌های آن افزایش می‌دهد، پرهیز شود. با توجه به اینکه امتحان درس فیزیک پزشکی برای دانشجویان دندانپزشکی اولین امتحان و برای دانشجویان پزشکی آخرین امتحان در ترم مذکور بوده است، لذا کاهش ۲ نمره در میانگین نمرات دانشجویان پزشکی می‌تواند به دلیل برگزاری دیرتر امتحان آنان و یا امکان حدس کورکورانه در سؤالات ص-غ باشد.

بر اساس جدول فوق ملاحظه می‌شود، علی‌رغم آنکه متوسط ضریب سختی سؤالات امتحانی برای دانشجویان پزشکی کمتر است و یا به عبارت دیگر سؤالات برای این گروه از دانشجویان سخت‌تر تلقی می‌شود، ولی با در نظر گرفتن ضریب سختی بهینه برای هر نوع سؤال مشاهده می‌شود که نسبت ضریب سختی به ضریب سختی بهینه و یا درصد ضریب سختی بهینه برای دو گروه کاملاً یکسان می‌باشد. این موضوع در مورد متوسط واریانس نمرات و درصد واریانس بهینه نمرات نیز صادق است. بنابراین می‌توان نتیجه گرفت که در شرایط یکسان سختی سؤالات بستگی به نوع سؤال ۴ و یا ۲ گزینه‌ای ندارد. متوسط ضریب تمیز سؤالات امتحانی برای دانشجویان پزشکی بیشتر است و یا به عبارت دیگر سؤالات این گروه از دانشجویان بهتر توانسته است آنان را ارزشیابی نماید. بنابراین می‌توان نتیجه گرفت که در شرایط یکسان سؤالات ۴ گزینه‌ای تا حدود ۴۰٪ بهتر از سؤالات ۲ گزینه‌ای ص-غ می‌تواند دانشجویان را ارزیابی نماید.

مقایسه مقادیر ضرایب پایانی امتحان با دو روش مختلف در جدول فوق



نشان می‌دهد که مقدار این ضریب بستگی به روش منتخب دارد. در روش ۲ نیم کردن سؤالات اسپیرمن-براون تفاوتی در ضریب پایانی امتحان با سؤالات ۴ و یا ۲ گزینه‌ای وجود ندارد. ولی چنانچه از روش KR20 کودر-ریچاردسون استفاده شود، ضریب پایایی در امتحان با سؤالات ۴ گزینه‌ای به میزان ۱۳/۷٪ بیشتر از امتحان با سؤالات ۲ گزینه‌ای می‌باشد. بنابراین می‌توان نتیجه گرفت که در شرایط یکسان سؤالات ۴ و ۲ گزینه‌ای (ص-غ) می‌توانند از نظر سختی یکسان طراحی شوند، ولی به هر حال سؤالات ۴ گزینه‌ای بهتر قادر به ارزشیابی دانشجویان بوده و پایانی امتحان را نیز بالاتر می‌نماید.

نتیجه‌گیری:

در درس فیزیک پزشکی سؤالات امتحانی ۴ و ۲ گزینه‌ای ص-غ می‌تواند از نظر سختی یکسان طراحی گردد ولی در شرایط یکسان سؤالات ۴ گزینه‌ای تا حدود ۴۰٪ دارای ضریب تمیز بالاتری بوده و بهتر از سؤالات ۲ گزینه‌ای ص-غ می‌تواند دانشجویان را ارزیابی نماید. همچنین ضریب پایایی به روش KR20 کودر-ریچاردسون در امتحان با سؤالات ۴ گزینه‌ای به میزان ۱۳/۷٪ بیشتر از امتحان با سؤالات ۲ گزینه‌ای می‌باشد. با این حال ضروری است که در هر یک از سؤالات ۴ گزینه‌ای، گزینه‌های مختلف با دقت طراحی شده و از انتخاب گزینه‌هایی که سبب اشتباه اغلب دانشجویان می‌گردد، پرهیز شود.

تشکر و قدردانی:

از کارشناسان و مسئول کتابخانه واحد EDC دانشگاه علوم پزشکی مشهد که منابع مورد نیاز را در اختیار گذاشته‌اند، تشکر و قدردانی می‌شود.

منابع:

۱- علی اکبر سیف: روش‌های اندازه‌گیری و ارزشیابی آموزشی. ویرایش دوم، تهران: نشر دوران، ۱۳۷۶.

2- Allen MJ, Yen WM. Introduction to measurement theory. Monterey: Brook/Col;1977.

3- Dick W, Hagesty N. Topics in measurements. New York: McGraw-Hill;1971.

4- Ebel RL. Essentials of educational measurement. 3rd ed. New Jersey : Prentice Hall;1979.

5- Mehrens WA, Lehman IJ. Measurements and evaluation in education and psychology. 3rd ed. New York: Holt, Rinehart and Winston;1984.