

## تحلیل آزمون چهارگزینه ای پایان ترم درس ژیمناستیک ۲ در دانشگاه پیام نور

محمود اکرامی<sup>۱\*</sup>، سمیه رجب زاده<sup>۲</sup>، فاطمه افزالی نسب<sup>۳</sup>، فاطمه ویژه زاده<sup>۴</sup>، عادله فرج پور<sup>۵</sup>

۱- دانشیار دانشگاه پیام نور ۲- دانشجوی دکترا ۳- کارشناس ارشد تربیت بدنی ۴- کارشناس ارشد تربیت بدنی ۵- کارشناس ارشد تربیت بدنی

تاریخ دریافت: ۱۳۹۳/۱۱/۱۰ تاریخ پذیرش: ۱۳۹۴/۰۲/۱۶

## Analysis of multiple test gymnastics final lesson 2 in Payame Noor University

Mahmoud Ekrami<sup>1\*</sup>, Somayeh Rajabzadeh<sup>2</sup>, Fatemeh Afzalinassab<sup>3</sup>, Fatemeh Vjehzade<sup>4</sup>,  
Adeleh Farajpour<sup>5</sup>

1- Associate Professor, Payame Noor University 2- Ph. D. Student 3- M. A. in Sport Management

4- M. A. in Sport Management 5- M. A. in Sport Management

Received: (2015/01/13)

Accepted: (2015/03/19)

### Abstract

The aim of this study was to analyze end of term multiple choice test gymnastics 2 PNU determine the validity and degree of difficulty of each of the articles through these tests. Calculate the degree of difficulty and the ability to distinguish each question showed that the number (4) of the gymnastics two practical tests (Articles 8, 13.19 and 24) were not enough validity and need to be removed from the set of test questions gymnastics 2. In this test, there was only one question too easy, but there was no hard question. It should be mentioned before deleting questions inadmissible subjects fell 73.92 percent and 69.10 percent of subjects after removing Questions undue loss was reduced and so will prevent the loss of 24 people.

### Keywords

Educational Evaluation; Classic Model, Degree of Difficulty of Questions, Validity

### چکیده

هدف از پژوهش حاضر تحلیل پاسخنامه‌های آزمون پایان ترم چهارگزینه‌ای ژیمناستیک ۲ دانشگاه پیام نور از طریق تعیین روایی و درجه دشواری هریک از ماده‌های این آزمون بود. در این مطالعه در دانشگاه پیام نور آزمون ژیمناستیک ۲ عملی دارای ۲۴ ماده ۴ گزینه‌ای است که روی ۴۸۷ نفر در سطح کشور اجرا گردیده و با استفاده از مدل کلاسیک مورد ارزیابی قرار گرفت. محاسبه درجه دشواری و قدرت تشخیص هر سؤال نشان داد که تعداد ۴ ماده از آزمون ژیمناستیک ۲ عملی (ماده‌های ۸، ۱۳، ۱۹ و ۲۴) از روایی کافی برخوردار نبوده و لازم است از مجموعه سؤال‌های آزمون ژیمناستیک ۲ عملی حذف شود. در این آزمون فقط یک سؤال خیلی آسان وجود داشت؛ اما سؤال خیلی سخت مشاهده نشد. لازم به توضیح است قبل از حذف سؤال‌های ناروا ۷۳/۹۲ درصد آزمودنی‌ها ریزش داشتند و پس از حذف سؤال‌های ناروا ریزش آزمودنی‌ها به ۶۹/۱۰ درصد تقلیل یافت و به‌این ترتیب از ریزش تعداد ۲۴ نفر جلوگیری می‌شود.

### واژه‌های کلیدی

ارزشیابی آموزشی؛ مدل کلاسیک؛ درجه دشواری سؤال؛ روایی

سؤال

## مقدمه

تعلیم و تربیت، فرآیندی است بسیار پیچیده و حساس که با جان و روح انسان‌ها سر و کار دارد، بنابراین همه افرادی که در این حیطه فعالیت می‌کنند، به ویژه آموزش‌دهندگان مسئولیت عظیم و خطیری بر عهده دارند. اجرای موفق و مطلوب فرآیند تعلیم و تربیت مستلزم آن است که دارای اهداف و مراحل کاملاً مشخص و برنامه‌ریزی شده باشد و نیز با توجه دقیق به اهداف و برنامه‌های تعریف شده، امکانات و تجهیزات لازم فراهم شده باشد. علاوه بر این، ارزیابی وضعیت موجود در حین اجرای فرآیند، بازخورد مطلوبی ایجاد خواهد کرد که مشخص می‌سازد برنامه اجرا شده تا چه حد در دست یابی اهداف از پیش تعیین شده، موفق بوده است. هم‌چنین ارزیابی درونی باعث خواهد شد تا شناخت صحیح و کافی از اشکالات و کمبودهای احتمالی به دست آید و زمینه لازم برای انجام اصلاحات فراهم گردد هدف اساسی ارزیابی درونی در تمام موارد و زمینه‌ها بهبود کیفیت امور است (فلاح و همکاران، ۱۳۸۲، ص ۷۴).

امروزه مقوله تربیت بدنی و ورزش به عنوان بخش عملی تعلیم و تربیت در نظام اسلامی از اهمیت ویژه‌ای برخوردار است که دستیابی به هدف‌های آن صرفاً از طریق برنامه‌ریزی صحیح و اصولی می‌تواند محقق گردد. در این راستا ضرورت همگامی با تغییرات شگرف و سریعی که در دنیای تکنولوژی صورت می‌گیرد، کاربرد روش‌های آموزشی پیشرفته و پیچیده را اجتناب‌ناپذیر می‌سازد. از این رو رسالت مراکز تربیت بدنی و علوم ورزشی نسبت به گذشته، سنگین‌تر و پیچیده‌تر شده است، از این رو برای ارتقای سطح کیفی و کمی این مراکز می‌باید برنامه‌ریزی‌های دقیق و اصولی و علمی انجام داد تا اهداف در این مراکز به نحو مطلوب تحقق یابد (کوزه‌چیان، ۷۹).

تحقق اهداف آموزشی، علاوه بر فراهم سازی بستر مناسب، تولید محتوای درسی و تأمین نیروی انسانی لازم، سنجش و ارزشیابی مداوم آموخته‌ها در قالب برگزاری آزمون و امتحانات متنوع را نیز می‌طلبد. اگرچه آزمون در وهله اول و از دید عموم، ابزاری نه چندان خوشایند در دست معلم برای سنجش میزان محفوظات متعلمان تعبیر می‌گردد، اما چنانچه یک آزمون با دقت و رعایت معیارهای لازم طراحی شود، به ابزاری مناسب برای نظارت بر فعالیتهای یادگیری و عاملی مؤثر در ارتقای کمی و کیفی فرآیند آموزش تبدیل می‌گردد (حقانی، ۱۳۸۵).

یکی از مهم‌ترین وظایفی که اعضای هیأت علمی با آن

مواجه هستند ارزیابی دانشجویان است. این وظیفه مهم نیاز به مهارت‌های قابل توجهی دارد، زیرا آزمون‌های دانشگاهی به عنوان یک شاخص کلیدی و حساس‌ترین نقطه فرآیند آموزش و یادگیری در تصمیم‌گیری برای ارتقا یا رد دانشجویان به کار می‌روند. مسلماً کار تهیه آزمون با نوشتن سؤالات آن پایان نمی‌یابد بلکه پس از آن باید به دنبال گردآوری شواهدی برای اطمینان از کیفیت مناسب سؤالات بود. برای چنین هدفی طراح آزمون باید یک فرآیند دقیق نظام دار را برای جمع‌آوری اطلاعات صحیح به انجام رسانده و بر پایه آن اطلاعات قضاوت نماید. هدف از چنین قضاوت‌هایی صرفاً شناسایی منابع احتمالی خطا و کاهش خطا در اندازه‌گیری است (شبان و همکاران، ۱۳۸۶). استفاده مؤثر از سؤال، به عنوان یک ترفند آموزشی، می‌تواند دانشجو را در بکارگیری اطلاعات در موقعیت‌های جدید و منحصر به فرد یاری دهد، سؤال کردن مؤثر، به معنی طرح سؤالاتی چالش‌برانگیز و وادارکننده دانشجو به تجزیه و تحلیل، ارزیابی، قضاوت و تصمیم‌گیری است (خداوند و همکاران، ۱۳۸۴).

در مطالعاتی که محققان تاکنون بر روی موفقیت و پیشرفت تحصیلی انجام داده‌اند، پیشرفت تحصیلی ۱ به عنوان یک شاخص مورد بررسی قرار گرفته است. در این مطالعات ویژگی‌های کیفی استاد نظیر مرتبه علمی در رشته مربوطه، آمادگی تدریس و انگیزه وی، دسترسی به اینترنت، فاکتورهای شخصیتی و سبک یادگیری فراگیران، میزان اهمیت خانواده به تعلیم و تربیت و ویژگی‌های فردی فراگیران نظیر علاقه به رشته، سلامت روان و رضایت دانشجویان از محتوای آموزشی خود، از جمله عوامل مؤثر در پیشرفت تحصیلی دانشجویان عنوان شده است (صفدری ده چشمه و همکاران، ۱۳۸۶، صص ۷۱ و ۷۲).

سیاست‌های نظام آموزش عالی کشور مبتنی بر توجه و تمرکز برنامه‌های آموزشی برای رسیدن به اهداف کیفی است. برنامه آموزشی باید در فاصله بین اهداف و حصول نتایج برنامه بطور مستمر ارزشیابی شود و بدین طریق، نقایص مشخص گردد. قصد ارزشیابی، مقایسه نتایج حاصل با هدف‌های آموزشی از پیش تعیین شده است تا درباره فعالیت‌های آموزشی و یادگیری دانشجویان برای حصول به بازده مطلوب تصمیم‌گیری شود (ناظم و همکاران، ۱۳۸۴، صص ۱۶۸ و ۱۶۹).

اعتقاد بر این است که ارزشیابی یکی از ابزارهای بسیار قدرتمند برای افزایش تاثیر آموزش است، اما مساله مهم این است که فرایند ارزشیابی به درستی انجام شود (رستگار، ۱۳۸۱).

سهولت نمره گذاری، اجرا و عینیت آزمون های چند گزینه ای سبب شده که به عنوان ابزار اصلی در سنجش های وسیع مورد استفاده قرار می گیرد. نوشتن سؤال های تستی با وجود همه تلاش هایی که در جهت مکانیزه و کامپیوتری کردن آن به عمل آمده است همچنان به عنوان یک هنر تلقی می شود. ابتکار و استادی در گنجاندن این مفاهیم در یک مساله، قاطعیت در جمله بندی و عبارات مربوط به سؤال به گونه ای که مسأله مورد نظر به وضوح بیان شده باشد و سرانجام بینش و مهارت در تولید گزینه های انحرافی به گونه ای که افراد ضعیف را به خود جلب کند مستلزم طرح و تحلیل دقیق سؤالها و تجزیه مستقیم است (ثرندیک به نقل از هومن، ۱۳۷۵).

امروزه به چند دلیل عمده از آزمون های چندگزینه ای بیش از سایر انواع آزمون ها در حوزه تعلیم و تربیت استفاده می شود: اول به علت آنکه آزمودنی قادر است در زمان معین تعداد زیادی سؤال را پاسخ دهد و به عبارتی دیگر در یک زمان محدود تعداد زیادی از هدف های آموزشی و بخش مهمی از محتوای درس را اندازه بگیرد. دوم اینکه آزمون های چند گزینه ای نسبت به آزمون های صحیح و غلط و دوگزینه ای کمتر امکان حدس زدن کورکورانه را به آزمون شونده می دهند. دلیل سوم برای استفاده بیشتر از این نوع آزمون سهولت در نمره گذاری و تصحیح و تفسیر آن و در نتیجه صرفه جویی در نیروی انسانی و وقت و هزینه می باشد (سیف، ۱۳۸۶).

اما معایبی نیز دارد از جمله اینکه ساختن این آزمونها بسیار دشوار است و در مقایسه با آزمون های صحیح - غلط خواندن این آزمون ها و پیدا کردن گزینه درست مستلزم وقت زیادتری است (شریفی، ۱۳۷۴)، به نقل از کریمی و همکاران ۱۳۹۰، ص ۳.

امروزه آزمون های چندگزینه ای استعداد و پیشرفت تحصیلی در حوزه تعلیم و تربیت بیش از سایر موقعیت ها مورد استفاده قرار می گیرد اما اینکه یک آزمون چندگزینه ای چه تعداد گزینه بایستی داشته باشد تا از حداکثر پایایی برخوردار باشد همواره مورد بحث می باشد. تعداد گزینه های سؤال های چندگزینه ای (گزینه های درست یا انحرافی) از ۲ تا ۵ و گاه تا ۶ متغیر است و به لحاظ نظری هر چه تعداد گزینه ها بیشتر باشد امکان حدس زدن کمتر است اما ابتدا باید مشخص کرد در یک آزمون، چند گزینه انحرافی بیشتر، باعث افزایش پایایی سؤال می شود به هر حال در عمل ممکن است اضافه کردن گزینه های انحرافی واقعا باعث افزایش پایایی آزمون نشود زیرا پیدا کردن گزینه های خوب مشکل است. پایایی هر

ارزشیابی موثر نه تنها باعث افزایش انگیزه در دانشجویان شده بلکه مدرس را نیز در ارزیابی فعالیت های خود کمک می کند. مهم ترین وسیله اندازه گیری پیشرفت تحصیلی، آزمون است. آزمون هایی که عملکرد تحصیلی دانشجویان را اندازه می گیرند، به انشایی و عینی تقسیم می شوند. در خصوص این تصمیم که از کدام یک از این آزمون ها برای اندازه گیری عملکرد تحصیلی دانشجویان استفاده شود بین اساتید، توافق نظر وجود ندارد (گرک یراقی و همکاران، ۱۳۸۹، صص ۵۳۳ و ۵۳۴).

به قول ثرندیک<sup>۱</sup> (۱۹۸۲) تاریخ اندازه گیری های روانی و تربیتی در قرن بیستم، در واقع تاریخ کشف و اختراع ابزارها و روش های اندازه گیری است که به طریقی استاندارد و تحت شرایط یکسان، رفتارهایی را که منعکس کننده خصیصه های افراد است، آشکار می کند و مورد سنجش قرار می دهد. ارائه نظریه های نوین اندازه گیری در قرن حاضر، به پیشرفت فنون و ابزارهای استاندارد شده ای انجامیده است که اندازه گیری و تبدیل توانش های فردی به مقیاس های قابل قبول برای توصیف، تفسیر و برآورد تفاوت های فردی را امکان پذیر می سازد (افروز و هومن ۱۳۷۵). تلاش های نخستین برای تکوین نظریه کلاسیک اندازه گیری در دهه ۱۸۹۰ آغاز شد. این نظریه، از روش های دیرینه برای ساخت و توسعه آزمون ها در حوزه علوم انسانی است که از اوایل دهه ۱۹۰۰ برای توسعه ابزارهای اندازه گیری و تعیین میزان همخوانی آزمون ها با نظریه و نمره گذاری امتحانات استفاده شده است (دلاور و همکاران، ۱۳۸۵، ص ۴۲).

اولین مدل و نظریه منسجم در حوزه سنجش و اندازه گیری ویژگی های روانی به اسپیرمن<sup>۲</sup> آماردان انگلیسی برمی گردد. وی در خلال سال های ۱۹۰۴ تا ۱۹۱۳ با استفاده از مفهوم همبستگی، مطالعاتی در خصوص نمرات به دست آمده از آزمون ها انجام داد. ایده اصلی اسپیرمن این بود که "هر نمره مشاهده شده از دو جزء نمره واقعی و نمره خطا تشکیل شده است." ارائه اولین مدل منسجم آماری برای تحلیل داده های به دست آمده از اجرای آزمون ها که به نظریه کلاسیک آزمون (Ctt) معروف است، یکی از کارهای علمی اسپیرمن بوده است (ایزنلو و همکاران، ۱۳۸۷، ص ۱۳۶).

فرایند آزمون سازی در طی روند تاریخی تکوین و تکامل خود با دو جنبه کاملا متمایز و مجزای ساخت و پرورش تست از یک سو و نظریه آماری برای تجزیه و تحلیل تست از سوی دیگر مواجه بوده است (هومن، ۱۳۷۵).

هدف از انجام این مطالعه، ارتقای کیفیت آزمون‌های پیشرفت تحصیلی مورد استفاده برای ارزشیابی دانشجویان در آزمون 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 حاصل آن بکارگیری نرم افزار تجزیه و تحلیل سؤالات چندگزینه‌ای مبتنی بر نظریه تستی کلاسیک برای دانشجویان می‌باشد. 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.

### روش‌شناسی تحقیق

مطالعه از نوع پس‌رویدادی و بازرسنجش، آزمون ۴ گزینه‌ای است که توسط اعضای هیئت علمی متخصص در رشته تحصیلی تربیت بدنی تهیه و همراه با کلید آن در اختیار بانک سؤال قرار گرفته، در آزمون‌های نهایی و پایان ترم مورد بهره‌برداری قرار می‌گیرد و برای پژوهشگر در حکم استفاده از اسناد و مدارک است. در دانشگاه پیام نور آزمون ژیمناستیک ۲ عملی (شماره درس ۱۲۱۵۱۸۰) دارای ۲۴ ماده ۴ گزینه‌ای است که روی ۴۸۷ نفر در سطح کشور اجرا گردیده و پاسخنامه ۵۰ نفر ترم اول سال تحصیلی ۹۲-۱۳۹۱ با استفاده از مدل کلاسیک مورد ارزیابی قرار گرفت.

در این پژوهش از دو شاخص روایی سؤال و درجه دشواری سؤال استفاده می‌شود. برای تعیین روایی سؤال از ضریب دورشته‌ای نقطه‌ای استفاده می‌گردد. ضریب همبستگی دورشته‌ای نقطه‌ای، شاخص رابطه دومتغییر است که در یکی از آنها فقط دوطبقه نمره (صحیح - غلط) وجود دارد و توزیع نمره‌های دیگر، فاصله‌ای و پیوسته است و مفروضه دیگری ندارد. این ضریب همبستگی نیز حالت خاصی از ضریب همبستگی گشتاوری پیرسون است و از فرمول زیر به دست می‌آید:

$$r_p = \frac{M_p - M_q}{S_E} \sqrt{p}$$

معنادار بودن تفاوت ضریب همبستگی دورشته‌ای نقطه‌ای از صفر در سطح  $\alpha = 0.10$  برابر  $\alpha = 1/64 = \sqrt{n}$ ، و در سطح  $\alpha = 0.05$  برابر  $\alpha = 1/58 = \sqrt{n}$ ، و در سطح  $\alpha = 0.01$  برابر  $\alpha = 1/96 = \sqrt{n}$  است. از سوی دیگر نسبت افرادی را که به سؤالی پاسخ درست داده اند درجه دشواری آن سؤال گویند، در این مطالعه اگر به سؤالی بیش از ۹۰٪ آزمون‌دهی‌ها پاسخ درست بدهند آن سؤال خیلی آسان و اگر به سؤالی کمتر از ۱۰٪ آزمون‌دهی‌ها پاسخ درست داده باشند آن سؤال خیلی سخت و اگر ضریب همبستگی دورشته‌ای

سؤال از طریق گزینه‌های انحرافی که هیچ وقت انتخاب نشده‌اند بالا نمی‌رود مطالعات نشان داده‌اند که به ندرت می‌توان سؤالاتی را یافت که بیشتر از ۳ یا ۴ گزینه انحرافی که به طور مؤثر عمل می‌کنند داشته باشند و گزینه‌های انحرافی خنثی ممکن است به پایایی آزمون لطمه وارد کنند به این دلیل که آنها زمان بیشتری را برای خواندن می‌گیرند و فضای آزمون را اشتغال می‌کنند (کریمی و همکاران ۱۳۹۰، صص ۳ و ۴).

شواهد نشان می‌دهد فراگیران در طی دوره تحصیل یاد می‌گیرند که همه فعالیت‌های کلاس ارزش یکسانی ندارند، بنابراین، تنها آنچه را فرا می‌گیرند که در امتحانات عامل کسب موفقیت است. چگونگی طرح این سؤالات در برگزیده اهداف شناختی آموزشی عنوان شده در طرح درس مربوطه می‌باشد که این اهداف به طور مطمئن قابل مشاهده و اندازه‌گیری است. در بسیاری از موارد کلیه دانشجویان ضعیف و قوی صرف نظر از میزان آمادگی، نمرات بالایی در امتحان کسب می‌نمایند و یا از گذراندن موفقیت آمیز آن باز می‌مانند. این واقعیت که دانشجویان بتوانند سؤالات یک مبحث را به درستی پاسخ دهند، تضمینی بر این نیست که از دانش کافی در آن موضوع برخوردارند. آزمون‌هایی که فقط بر پایه یادآوری مطالب (تاکسونومی) می‌باشند، از پایایی کمی برای سنجش فراگیری واقعی برخوردارند. اگر چه با استفاده از سؤالات چند گزینه‌ای می‌توان محتوای زیادی را مورد ارزیابی قرار داد، اما در صورتی که اصول صحیح طراحی سؤال رعایت نشود، از اعتبار این نوع سؤالات کاسته می‌گردد.

توسعه سریع تسهیلات آموزشی در کشورها ایجاب می‌کند که در امر پذیرش داوطلبان و تعیین نوع مورد نیاز مشاوره و راهنمایی افراد پس از به کارگماری آنان از آزمون‌ها استفاده شود. با صنعتی شدن این کشورها، نیاز مبرم به آزمون‌هایی جهت گزینش و به کارگماری افراد، بویژه در رشته‌های فنی و دفتری و تخصصی بوجود آمده است. لذا، برای تحقق این مهم نیازمند تفسیر صحیح نمره‌ها جهت اندازه‌گیری روانی و تربیتی شناخت فرد یعنی توصیف کامل و جامع او به گونه‌ای که بتوان درباره خصیصه‌های بدنی و روانی او اطلاعات مناسب و دقیقی فراهم کرد باشیم. برای این امر از مدل کلاسیک و IRT می‌توان بهره‌گیری کرد که در مدل کلاسیک از نمره‌های خام فرد استفاده می‌شود و اطلاعات درباره فرد با توجه به تعداد سوال‌های پاسخ داده شده کسب می‌شود اما در مدل IRT، توانایی یا سطح صفت مکنون فرد با در نظر گرفتن پارامتر شیب، دشواری و حدس مشخص می‌شود و سهمی که هر سوال در توانایی فرد دارد بسته به پارامترهای سوال متفاوت خواهد بود.

**یافته های تحقیق**

با توجه به توضیحات داده شده آزمون ژیمناستیک ۲ عملی مورد تجزیه و تحلیل قرار گرفت و نتایج برای این آزمون در جدول ۱ نشان داده شده است.

نقطه ای سؤالی کمتر از  $\sqrt{n} = 1/64 = 0/259$  باشد سؤال ناروا تلقی می شود گاهی همبستگی های ضعیف تر نیز برای از دست ندادن سوال روا تلقی شده است به همین منظور آستانه ناروایی سؤال به کمتر از  $0/1$  تنزل داده شد.

**جدول ۱.** تعیین ماده های خیلی سخت، خیلی آسان و ناروا ی آزمون ژیمناستیک ۲ عملی (۱۲۱۵۱۸۰)

ردیف	شماره سؤال	متن سؤال	Key	P	Rpbis	نتیجه
۱	۶	در تمرینات ژیمناستیک با استفاده از یک رژیم غذایی بر کربوهیدرات چند ساعت پس از برگشت به حالت اولیه، گلیکوژن عضلات به طور کامل دوباره سازی می شود؟	۲	۰/۹۲۰	۰/۲۷۶	خیلی آسان
						
۲	۸	هدف از تمرین زیر در آموزش نیم پشتک چیست؟ (۱) آموزش و تسلط در بالانس دست (۲) آموزش جفت شدن پاها در بالانس (۳) آموزش پرتاب پاها (۴) افزایش توان انفجاری کمر بند شانه ای	۳	۰/۶۴۰	۰/۰۹۳	ناروا (حذف شود)
۳	۱۳	عدم کنترل خم کردن آرنج ها، ایراد عمده کدام حرکت موازنه است؟ (۱) غلت جلو (۲) غلت عقب و نشستن روی چوب موازنه (۳) غلت عقب روی شانه (۴) بالانس دست - غلت جلو	۴	۰/۸۲۰	-۰/۰۵۳	ناروا (حذف شود)
۴	۱۹	در حرکت دور کردن بدن از میله بلند و اجرای چرخ شکم روی میله کوتاه، بعد از تماس شکم با میله چه زمانی میله بلند رها می شود؟ (۱) همزمان با تماس شکم با میله (۲) درست قبل از سطح افق (۳) بعد از سه چهارم دور چرخش (۴) بعد از زاویه شدن بدن	۴	۰/۲۴۰	۰/۰۷۲	ناروا (حذف شود)
۵	۲۴	در مقررات امتیاز گذاری خانم ها، حرکت یاماشیتا (پرش خرک) در کدام گروه قرار می گیرد؟ (۱) گروه یک (۲) گروه دو (۳) گروه سه (۴) گروه چهار	۱	۰/۱۸۰	-۰/۳۹۰	ناروا (حذف شود)

مجموعه سؤال های آزمون ژیمناستیک ۲ عملی حذف شود. در این آزمون فقط یک سؤال خیلی آسان وجود دارد ولی سؤال خیلی سخت مشاهده نشد. لازم به توضیح است قبل از حذف

همان گونه که در جدول ۱ مشاهده می شود تعداد ۴ ماده از ۵ ماده مورد اشاره از آزمون ژیمناستیک ۲ عملی (ماده های ۸، ۱۳، ۱۹ و ۲۴) از روایی کافی برخوردار نبوده و لازم است از

آزمون ژیمناستیک ۲، متوسط نمره آزمون شوندگان از کمینه نمره قبولی کمتر است، این آزمون همراه با نام و شماره درس، تعداد سؤال های آزمون و فاصله اطمینان ۹۹ درصدی برای متوسط پیشرفت تحصیلی در جدول ۲ نشان داده شده است.

سؤال های ناروا ۷۳/۹۲٪ آزمودنی ها ریزش داشتند و پس از حذف سؤال های ناروا ریزش آزمودنی ها به ۶۹/۱۰٪ تقلیل یافت و به این ترتیب از ریزش تعداد ۲۴ نفر جلوگیری می شود. در صورتی که کمینه نمره قبولی برابر نصف تعداد سؤال های (ماده های) هر آزمون باشد (مانند آزمون های خودخوان)، در

**جدول ۲.** آزمون ژیمناستیک ۲ عملی که متوسط پیشرفت تحصیلی از کمینه نمره قبولی کمتر است

نام درس	شماره درس	تعداد ماده	میانگین کل	واریانس	انحراف استاندارد	خطای استاندارد	فاصله اطمینان ۹۹ درصدی
ژیمناستیک عملی	۲	۱۲۱۵۱۸۰	۱۳/۰۲	۱۴/۱۴	۳/۷۶	۲/۱۵	۷/۴۷ - ۱۸/۵۷

های خودخوان) در نظر گرفته شد. برای رعایت اختصار آزمون ژیمناستیک ۲ عملی همراه با شماره درس، تعداد ماده ها، حجم نمونه، ماده های ناروا، درصد ریزش قبل از حذف ماده های ناروا، درصد ریزش پس از حذف ماده های ناروا، و تعداد دانشجویانی که از کمند ریزش جستند در جدول ۳ نشان داده شده است.

چنان که در جدول ۲ مشخص شده است متوسط نمره قبولی آزمون شوندگان در آزمون کمتر از کمینه نمره قبولی است، در صورتی که آزمون شوندگان از نمره کلاسی (حداکثر ۶ نمره) برخوردار باشند، کمینه نمره قبولی در آزمون نهایی به میزان ۰/۱ تا ۰/۱۳۳ (برحسب طول تست ۴۰ یا ۳۰) کاهش می یابد. ولی به جهت امکان مقایسه یکسان آزمون ها، این آزمون نیز بدون رعایت نمره کلاسی و یا آزمون میان ترم (مانند آزمون

**جدول ۳.** تعیین ماده های ناروا، درصد ریزش قبل، و بعد از حذف ماده های ناروا در آزمون نهایی ژیمناستیک ۲ عملی

نام درس	شماره درس	نمونه ماده ها	ماده های ناروا	ماده های خیلی خرابی	ماده های خیلی آسان	تعداد کل آزمودنی	درصد ریزش قبل	درصدی زش بعد	تعداد قب ولی بعد
ژیمناستیک عملی ۲	۱۲۱۵۱۸	۵۰	۲۴	۸ و ۱۳ و ۱۹ -	۶	۴۸۷	۷۳,۹۲٪	۶۹,۱٪	۲۴

داوران متخصص رشته مورد نظر ساخته و اجرایی شود. این نوع آزمون ها رمالاک- مرجع<sup>۷</sup> نیز می گویند. در پژوهش تقی زاده (۱۳۹۰) ۶ آزمون در بین کارکنان و مدیران دفتر آموزش و پژوهش استاندارد البرز که هر کدام شامل ۵۰ ماده چهار گزینه ای بود اجرا گردید و درجه دشواری و قدرت تشخیص ماده های خیلی سخت، خیلی آسان و ناروای هریک از آزمون ها مشخص گردید. در آن مطالعه متغیر اخلاق فردی دارای ۵۰ ماده چهارگزینه ای است و در آن دامنه p از ۰/۰۰۰ تا ۰/۸۸۹ و دامنه rpbis از ۰/۱۳۰- تا ۰/۴۲۲ قرار دارد و تعداد ۲۴ ماده از پرسشنامه اخلاق فردی ناروا ارزیابی گردیده و لازم است حذف شوند. دومین متغیر یعنی متغیر تربیت قرآنی و راهکارهای

## بحث و نتیجه گیری

ارزشیابی آموزشی از مهم ترین ارکان دوره آموزشی است. به گفته هومن (۱۳۷۵، ص ۱۶) فعالیت های مرتبط با ارزشیابی عبارت از: اندازه گیری<sup>۱</sup>، پژوهش<sup>۲</sup>، و ارزیابی<sup>۳</sup> یادگیرنده است. دونوع آزمون کلی در ارزشیابی آموزشی به کار می رود: آزمونی که برای ارزشیابی گام به گام پیشرفت های دانشجویان، مقایسه دانشجویان، تشخیص و بهبود نقاط ضعف، و در آخر به منظور رفع نواقص آموزشی به کار می رود و نه نمره دادن و صدور گواهینامه، به بیان هومن (۱۳۸۹، ص ۱۹۲) این نوازمون راسازنده<sup>۴</sup> و یا نرم-مرجع<sup>۵</sup> گویند، در مقابل آزمون های جامع یا نهایی<sup>۶</sup> است که معرف دانش و مهارت های دانشجویان است و توسط بهترین

5. Norm-referenced test  
6. Summative test  
7. Criterion-referenced test

1. measurement  
2. research  
3. appraisal  
4. Formative test

امیریان (۱۳۹۱) با استفاده از مدل IRT پارامترهای سه گانه a, b, c و نیز نمره پیشرفت تحصیلی ( $\theta$ ) را در آزمون های دستیاری، کارورزی و کارآموزی دانشگاه علوم پزشکی مشهد تعیین نمود. نتایج مطالعه امیریان نشان داد در آزمون یکم دستیاری (۱۰ درصد)، آزمون دوم دستیاری (۶/۰۶ درصد)، آزمون سوم کارورزی (۰)، آزمون چهارم دستیاری (۱/۳۹ درصد)، آزمون پنجم کارآموزی (۰)، آزمون ششم کارآموزی (۱/۶۷ درصد) سوال ها ناروا تشخیص داده شد و باید از مجموعه سوال ها حذف شود. نتیجه گیری: ثرندایک به نقل از هومن (۱۳۸۰) معتقد است کاربرد مدل کلاسیک در مطالعه های مشابه حاضر موجب می شود: ۱- همه سؤال های تست از لحاظ شکل و محتوا همگون باشد، به گونه ای که تمامی آنها بایک نوع خصیصه مکنون مشترک ارتباط داشته باشد. ۲- سؤال ها با استفاده از یک فرم تجربی مقدماتی سرند شود، به گونه ای که سؤال های مبهم و سؤال هایی که به هردلیل ناممیزا بوده است، از تست خارج شده باشد. ۳- سؤال ها به صورتی باشد که آزمودنی برای پاسخ دادن به آنها، موظف به تولید پاسخ شود (نه آنکه پاسخ درست را از میان گزینه های معین برگزیند). ۴- سؤال ها منعکس کننده رشد کلی یک نوع خصیصه (صفت) باشد، نه یک مهارت بخصوص.

از سویی اگر منظور از آزمون، اجرای تست جامع<sup>۲</sup> یعنی آزمونی است که معمولا در پایان ترم و گاه در پایان دوره تحصیلی برگزار می شود که می شود، به سبب تعیین میزان آموخته های دانشجویان در طول یک ترم یادوره آموزشی همراه بانمره دادن و یا صدور گواهینامه و داوری درمورد کارآمدی تدریس یا برنامه خاصی است، ضرورت توجه به تناسب تک تک ماده ها باکل تست از اهمیت ویژه برخوردار است. از سویی دیگر بازپروری ماده های آزمون های نهایی پایان ترم از طریق حذف سؤال های ناروا ضمن پیش گیری از حذف نابجای آزمودنی های مرزی، مایه صرفه جویی های مادی برای مرکز آزمون از طریق تقلیل تکرار آزمون های ریزشی است و معمولا در هر نیمسال تحصیلی توسط مرکز آزمون دانشگاه، حدود یکصد آزمون ریزشی با هزینه دانشگاه اجرا می شود.

آن است که دامنه p از ۰/۰۰۰ تا ۱/۰۰۰ و دامنه rpbis از ۰/۲۲۸- تا ۰/۶۳۸+ است و تعداد ۱۷ ماده آن از روایی کافی برخوردار نبوده، متغیر دستیابی به اطلاعات هم دارای ۵۰ ماده چهارگزینه ای است و در آن دامنه p از ۰/۰۰۰ تا ۰/۹۵۲ و دامنه rpbis از ۰/۶۰۹- تا ۰/۲۳۱+ قرار دارد و تعداد ۱۲ ماده پرسشنامه ناروا ارزیابی گردیده، در متغیر بعدی یعنی متغیر روابط اجتماعی دامنه p از ۰/۰۳۷+ تا ۱/۰۰۰ و دامنه rpbis از ۰/۲۴۷- تا ۰/۳۷۰+ قرار دارد و تعداد ۶ ماده آن ناروا ارزیابی گردید، پنجمین متغیر متغیر روابط خانوادگی بود که در آن دامنه p از ۰/۶۶۷+ تا ۱/۰۰۰ و دامنه rpbis از ۰/۱۸۴- تا ۰/۵۱۱+ قرار دارد و تعداد ۱۶ ماده آن از روایی کافی برخوردار نبود، آخرین متغیر یعنی متغیر سیری در اندیشه سیاسی غرب و اسلام که در آن دامنه P از ۰/۰۰۰+ تا ۰/۹۴۷+ بوده و دامنه Rpbis از ۰/۲۳۸- تا ۰/۶۲۰+ قرار دارد و تعداد ۳۲ ماده آن ناروا ارزیابی گردید و لازم است از مجموعه سوالات پرسشنامه حذف گردند. نتایج نشان داد آزمون سیری در اندیشه سیاسی غرب و اسلام ضعیف و فاقد اعتبار بوده است.

در پژوهش اکرامی (۱۳۹۰) در استاندارد تهران با استفاده از ضریب همبستگی دو رشته ای نقطه ای پارامترهای هر ماده از جمله روایی و سطح دشواری سؤال (نسبت افرادی که پاسخ درست به سوال داده اند) مورد تجزیه و تحلیل قرار گرفت. نتایج نشان داد آزمون دوره های غیر حضوری "اصول و مبانی ارتباط" دارای ۲ سوال ناروا و ۱۴ سوال خیلی آسان؛ دوره "علم و دین" دارای ۶ سوال ناروا، یک سوال خیلی سخت و ۴ سوال خیلی آسان؛ دوره "شناسایی و استفاده از ظرفیت ذهنی" دارای ۲ سوال ناروا، ۹ سوال خیلی آسان، دوره "سیری در اندیشه سیاسی اسلام و غرب" دارای ۴ سوال ناروا، ۴ سوال خیلی آسان؛ دوره "مدیریت و سازمان از دیدگاه قرآن و نهج البلاغه" بدون سوال ناروا و دارای دشواری مناسب؛ دوره "ارتباط موثر با محیط و دیگران" دارای ۴ سوال ناروا و ۱۳ سوال خیلی آسان بود. در میان دوره ها سوال های دوره "مدیریت و سازمان از دیدگاه قرآن و نهج البلاغه" بهترین طراحی را داشت.

## منابع

- ایزنلو، بلال، حبیبی عسگرآباد، مجتبی (۱۳۸۷). «مقدمه ای بر مبانی رویکردهای جدید اندازه گیری در حوزه روانشناسی و علوم تربیتی». *فصلنامه روانشناسی و علوم تربیتی*، ۱۱ (۲): ۳۱ - ۱۹.
- ثرندایک، ترجمه هومن، حیدر علی (۱۳۷۵). *روان سنجی کاربردی*. انتشارات دانشگاه تهران.
- حقانی، نادر. (۱۳۸۹). *تحلیل آزمونهای ارزشیابی پیشرفت یادگیری بر پایه تحلیل آزمونهای ارزشیابی پیشرفت یادگیری بر پایه آزمونهای الکترونیکی*. قابل دسترس از سایت Passwort Deutsch.

- دلاور، علی، مقدم زاده، علی، مطیعی لنگرودی، طیبه (۱۳۸۵). «مقایسه مدل اندازه گیری کلاسیک و مدل غیر پارامتریک سوال-پاسخ از نظر ویژگی های سوال». فصلنامه نوآوری های آموزشی، شماره ۱۸(۵): ۳۳-۴۶
- سیف، علی اکبر (۱۳۸۶). اندازه گیری، سنجش و ارزشیابی آموزشی. ویرایش چهارم. تهران: نشر دوران.
- شبان، مرضیه، رضائی بدر، فرهاد (۱۳۸۶). «تأثیر تجزیه و تحلیل سوالات آزمون های پایان ترم بر کیفیت طراحی سؤال». مجله دانشکده پرستاری و مامایی دانشگاه علوم پزشکی تهران (حیات). دوره ۱۳(۱): ۸۲ - ۷۱.
- صفدری ده چشمه، فرانک و همکاران (۱۳۸۶). «عوامل موثر در پیشرفت تحصیلی دانشجویان از نظر دانشجویان و اساتید دانشکده پرستاری و مامایی دانشگاه علوم پزشکی شهرکرد در سال ۱۳۸۳». مجله دانشگاه علوم پزشکی شهرکرد، دوره ۹، شماره ۳: ۹۹ - ۸۵
- فلاح، سیده حوریه، مهدی نیا، سید محمود، قائنی، غلامرضا و همکاران (۱۳۸۲). «ارزیابی درونی گروه بهداشت محیط دانشکده بهداشت، دانشگاه علوم پزشکی سمنان». مجله علمی دانشگاه علوم پزشکی سمنان، ۵(۲): ۳۹-۲۹.
- کریمی، بهنام، فلسفی نژاد، محمدرضا، درتاج، فریبرز (۱۳۹۰). «تأثیر تعداد گزینه های سوال در ویژگی های روان سنجی آزمون و توانایی برآورد شده در مدل های پرسش و پاسخ اندازه گیری». فصلنامه اندازه گیری تربیتی، ۶(۲): ۴۵-۳۳.
- کوزه چیان، هاشم (۱۳۷۹). «بررسی وضعیت آموزشی دانشجویان دانشکده های تربیت بدنی کشور». مجله حرکت، ۴: ۸۲-۷۵.
- گرک یراقی، محمد و همکاران (۱۳۸۹). «بررسی شاخص های کمی و کیفی آزمون های مقطع کارآموزی دوره پزشکی عمومی». مجله ایرانی آموزش در علوم پزشکی، ۱۰(۵): ۲۴-۱۱.
- ناظم، مسعود و همکاران (۱۳۸۴). «دیدگاه کارورزان در مورد آمادگی برای ورود به دوره کارورزی». مجله ایرانی آموزش در علوم پزشکی، ۵(۲): ۵۳-۴۱.
- تقی زاده، تهمینه (۱۳۹۰). ارزیابی دوره های کوتاه مدت دفتر آموزش و پژوهش استانداری تهران. پایان نامه کارشناسی ارشد مدیریت آموزشی. دانشگاه پیام نور تهران، گروه علوم تربیتی.
- امیریان، سحر. (۱۳۹۱). تعیین پارامترهای سه گانه آزمون های چهار گزینه ای دانشگاه علوم پزشکی مشهد بر پایه مدل IRT. پایان نامه کارشناسی ارشد مدیریت آموزشی. دانشگاه پیام نور تهران، گروه علوم تربیتی. منتشر نشده.