

## تحلیل علمی آزمونها

دکتر محمدعلی احمدوند\*

این آزمونها معلم ساخته هستند و در پایان هر نیمسال آموزشی به صورت متمرکز و همزمان در مراکز مختلف آموزشی اجرا می‌شوند. در آزمونهای معلم ساخته که غالباً برای تعیین میزان توفیق یادگیرندگان در رسیدن به هدفهای دقیق آموزشی طرح‌ریزی می‌شود، معمولاً از ملاک مطلق اندازه‌گیری استفاده می‌گردد (۲۰۱).

طراحان سؤالات آزمونهای پیشرفت تحصیلی باید با هدفهای آموزشی آشنا باشند و بتوانند هدفها را در قالب هدفهای شناختی، عاطفی و روانی - حرکتی با توجه به طبقات دانش، فهمیدن، کارستن، تحلیل، ترکیب و ارزشیابی تدوین کنند.

آزمونهای چندگزینه‌ای متداولترین آزمونهای عینی هستند که در بخش یادگیری در زمینه‌های گوناگون تحصیلی کاربرد دارند. این آزمونها هم از لحاظ یکنواختی، هم از لحاظ حساسیت کم در مقابل حدس زدن کورکورانه و هم از حیث سهولت تصحیح پاسخنامه‌ها بهترین نوع آزمونهای عینی هستند. آزمونهای چهارگزینه‌ای یک گزینه درست و سه گزینه انحرافی دارند، نقش گزینه‌های انحرافی، مورد نظر آزمون شوندگانی است که پاسخ درست سؤال را نمی‌دانند.

چکیده: آزمونهای چهارگزینه‌ای از متداولترین ابزارهای عینی اندازه‌گیری هستند که در دانشگاه پیام‌نور اجرا می‌شوند. برای بررسی کیفیت ارزشیابی دانشگاه، ده آزمون از دو گروه زمین‌شناسی و روان‌شناسی از دو دانشکده علوم و ادبیات و علوم انسانی به طور تصادفی انتخاب شدند. پاسخنامه دانشجویان، در ده آزمون مذکور، مورد تحلیل علمی قرار گرفت.

هدف پژوهش بررسی روایی و پایایی آزمونها، ضریب دشواری، ضریب تمیز و واریانس سؤالات آزمونها بود. خارج بودن ضرایب آماری از سطح بهینه و روایی و پایایی بیشتر آزمونهای گروه زمین‌شناسی به عنوان فرضیه‌های پژوهش تعیین شدند. نتایج پژوهش نشان داد که ۳۸ درصد از سؤالات گروه روان‌شناسی و ۴۲ درصد از سؤالات گروه زمین‌شناسی نامناسب هستند. ضریب دشواری و ضریب تمیز آزمونها خارج از سطح بهینه و واریانس آزمونها داخل سطح بهینه هستند و پایایی آزمونهای گروه زمین‌شناسی بیشتر از پایایی آزمونهای گروه روان‌شناسی و روایی آزمونهای گروه روان‌شناسی بیشتر است.

کلیدواژه: تحلیل سؤال، ضریب دشواری، ضریب تمیز، واریانس، پایایی و روایی.

### مقدمه

برای اندازه‌گیری میزان یادگیری دانشجویان، می‌توان از وسایل متعددی استفاده کرد. در دانشگاه پیام‌نور، از آزمونهای چهارگزینه‌ای و تشریحی استفاده می‌شود.

برای تهیه سؤالی چندگزینه‌ای قواعدی تدوین شده است که لازم است طراحان سؤالات امتحانی از آنها سود جویند (۳۱).

هر نوع وسیله اندازه‌گیری باید ویژگیهایی را دارا باشد تا آن وسیله برای منظوری که ساخته شده است، مفید باشد. بهترین این ویژگیها روایی<sup>۱</sup> و پایایی<sup>۲</sup> هستند. روایی اصطلاحی است که به هدفی که آزمون برای تحقق بخشیدن به آن درست شده است، اشاره می‌کند. آزمونی روایی است که برای اندازه‌گیری آنچه موردنظر، است مناسب باشد.

روایی محتوایی<sup>۳</sup> نشان می‌دهد که نمونه سؤالی مورد استفاده در یک آزمون تا چه حد معرف کل جامع سؤالی ممکن است که می‌توان از محتوا یا موضوع مورد نظر تهیه کرد. برای ساختن آزمونهای پیشرفت تحصیلی رواج باید سؤالی آزمون، نمونه کاملی از هدفها و محتوای درسی باشند.

یکی از عواملی که بر روایی آزمون اثر می‌گذارد کیفیت سؤالی آزمون است. سؤالی که طبق قواعد درست سؤال‌نویسی تهیه شده باشند، به روایی آزمون می‌افزایند. هر کدام از سؤالی آزمون که از هدف اصلی آزمون جدا می‌افتد نیز به سهم خود از روایی کل آزمون می‌کاهد. آزمونهایی که دارای مجموعه سؤالی بسیار ساده‌اند یا از سؤالی بسیار دشوار تشکیل شده‌اند از سطح روایی خوبی برخوردار نیستند (۵۴).

پایایی یک وسیله اندازه‌گیری، به دقت آن وسیله اشاره می‌کند. یک آزمون در صورتی دارای پایایی است که اگر آن را در یک فاصله زمانی کوتاه چندین بار به گروه واحدی از افراد بدهیم، نتایج حاصل نزدیک به هم باشند. روشهای تعیین پایایی متعدد هستند که یکی از آنها روش دو نیمه کردن آزمون می‌باشد. در این روش همه سؤالی فرد را در یک آزمون و همه سؤالی زوج را آزمون دیگری می‌پنداریم. ضریب همبستگی حاصل از نمرات دو نیمه آزمون، ضریب پایایی هر یک از دو نیمه خواهد بود. برای محاسبه پایایی کل آزمون از فرمول اسپیرمن - براون استفاده می‌شود (۷۶).

### تحلیل سؤالی آزمون

هدف از تحلیل سؤالی آزمون، روایی تک تک سؤالی و تعیین میزان دقت و نارساییهای آنهاست. طراحان سؤالی امتحانی می‌توانند پس از اجرای آزمون نتایج آن را تحلیل کنند و به بهبود کیفیت سؤالی برای استفاده‌های بعدی اقدام کنند.

متداولترین مورد استفاده اطلاعات مربوط به تجزیه و تحلیل سؤالی، انتخاب سؤالی بهتر و مناسبتر برای تدوین فرم نهایی آزمون است. درگزینش هر سؤال باید از طریق تجزیه و تحلیل دقیق، نشان داد که سؤال انتخابی تا چه میزانی روایی و پایایی آزمون را افزایش خواهد داد (۸۱).

### مراحل تحلیل سؤال

داده‌های موردنیاز برای تحلیل سؤالی یک آزمون، پاسخهایی هستند که آزمون‌شوندگان به هر سؤال داده‌اند. بنابراین باید تعیین شود که در هر سؤال چند نفر را به خود جلب کرده‌اند و چند نفر آن را بی‌جواب گذاشته‌اند.

برای تحلیل هر سؤال و محاسبه شاخصهای آماری باید دو گروه قوی و ضعیف آزمون‌شوندگان تعیین شوند و نحوه پاسخ‌دهی آنان به گزینه‌های مختلف یک سؤال به دست آیند. برگه‌های آزمون از بالاترین نمره به پایینترین نمره تنظیم می‌شوند و بر اساس تعداد اوراق درصدی از بالا و پایین به عنوان گروه قوی و ضعیف انتخاب می‌گردند و چگونگی جواب دادن افراد در کارت تحلیل سؤال درج می‌گردند.

درصد کل آزمونیهایی که به یک سؤال جواب مثبت می‌دهند، ضریب دشواری آن سؤال را به دست می‌دهد. هر اندازه ضریب دشواری یک سؤال بزرگتر باشد، آن سؤال آسانتر است و هر اندازه کوچکتر باشد، سؤال دشوارتر است. برخی ضریب دشواری را مانند ضریب تمیز<sup>۴</sup> به صورت عدد اعشاری و دو رقمی نشان می‌دهند.

به طور کلی ضرایب دشواری بین ۳۰٪ تا ۷۰٪ حداکثر اطلاع را درباره تفاوت بین آزمونیها به دست می‌دهند. البته در شرایط ویژه مثل انتخاب بهترین داوطلبان برای تحصیلات تکمیلی یا انتخاب کودکان برای طی یک دوره

1. validity

2. reliability

3. content validity

4. discriminative index

نظریه سؤال - پاسخ بر الگوی صفت مکنون<sup>۳</sup> استوار است. نظریه صفت مکنون بر این فرض تکیه کرده است که یک ویژگی زیربنایی وجود دارد که به شخص امکان می‌دهد تا در یک تکلیف شناختی معین، موفقیت کسب کند. افزون بر این چنین فرض می‌شود که هرچه شخص از این صفت بیشتر برخوردار باشد، در آزمون مربوط عملکرد بهتری خواهد داشت.

در نظریه سؤال - پاسخ، هرچه توانایی فرد از میزان دشواری سؤال بیشتر باشد، احتمال پاسخ صحیح فرد به سؤال نیز بیشتر می‌شود و هرچه جایگاه فرد بر روی محور تواناییها از میزان دشواری سؤال کمتر بشود احتمال پاسخ صحیح این فرد به سؤال نیز کمتر و کمتر می‌شود.

با استفاده از نظریه سؤال - پاسخ می‌توان برخی ویژگیهای سؤال را تعیین کرد. این ویژگی سؤال برای بسیاری مقاصد آزمون‌سازی از جمله تحلیل سؤالهای آزمون مفیدند. ویژگیهای هر سؤال را به صورت نمودار نشان می‌دهند و آن را منحنی ویژگی سؤال<sup>۴</sup> (آی، سی، سی) می‌نامند.

منحنی ویژگی سؤال، احتمال پاسخ درست دادن به هر سؤال را به توانایی آزمون شونده ربط می‌دهد. برای تهیه منحنی ویژگی سؤال نسبت یا درصد آزمون شوندگانی که آن سؤال را درست جواب داده‌اند در رابطه با نوعی ملاک، مثلاً نمره کل آزمون آنها، رسم می‌شود. بر روی محور افقی نمره کل آزمون و بر روی محور عمودی نسبت آزمون شوندگانی که به سؤال پاسخ درست داده‌اند، مشخص می‌شود. در این نظریه می‌توان ضریب دشواری سؤال را مستقیماً به سطح زیرمنحنی ویژگی سؤال ربط داد.

- هر چه سطح زیرمنحنی ویژگی سؤال بیشتر باشد، ضریب دشواری سؤال بزرگتر است. هر چه منحنی ویژگی سؤال، حالت یکسانی بیشتری داشته باشد همبستگی بین آن سؤال و کل آزمون بیشتر است (۱، ۸، ۱۰ و ۹).

آموزشی، مقدار ضریب دشواری، متفاوت خواهد بود. در سازمان سنجش برای آزمونهای کارشناسی ارشد، ضریب دشواری تا ۰/۶۹ قابل قبول تلقی شده است (۱۰ و ۹).

ضریب تمیز، قدرت سؤال را در تمایزگذاری یا تشخیص بین گروه قوی و گروه ضعیف مشخص می‌کند. انتخابهای درست گروه بالا منهای انتخابهای درست گروه پایین بخش بر تعداد افراد یک گروه، ضریب تمیز سؤال را تعیین می‌کند. هر قدر ضریب تمیز بزرگتر باشد، قوه تمیز آن سؤال بیشتر است.

سؤالهای خوب یک آزمون آنهایی هستند که دارای ضریب دشواری متوسط و ضریب تمیز بالایی باشند. گزینه‌های انحرافی باید طوری تهیه شوند که بتوانند افراد ضعیف را به خود جذب کنند و گرنه گزینه‌های بیهوده‌ای هستند که بودن یا نبودن آنها در سؤال، تفاوتی ایجاد نمی‌کنند. هر گزینه انحرافی باید دست‌کم یک نفر از افراد گروه ضعیف را به خود جلب کند و اگر هر دو گروه قوی و ضعیف را به خود جلب می‌کند، تعداد افراد گروه ضعیف بیشتر باشد. بنابراین، در صورتی یک سؤال به خوبی عمل می‌کند که افراد ضعیف بیشتر از افراد قوی گزینه‌های انحرافی آن سؤال را انتخاب نمایند.

روشهای تحلیل سؤال با استفاده از نظریه سؤال - پاسخ<sup>۵</sup> روشهای کلاسیک تحلیل سؤالات آزمونها دارای معایبی هستند که به برخی از آنها اشاره می‌شود:

- وابسته بودن خصوصیات سؤالات به گروه آزمون و برعکس، به طور مثال اگر سؤالها ساده باشند آزمودنیها قویتر و اگر سؤالها دشوارتر باشند آزمودنیها ضعیفتر ارزشیابی می‌شوند. اگر آزمون برای افراد قوی اجرا شود، دشواری کمتری را نشان می‌دهد تا زمانی که همین آزمون برای افراد ضعیف اجرا شود.
- نمره‌گذاری در روشهای مبتنی بر نظریه کلاسیک غیرواقعی است. به هر سؤال ارزش یکسانی تعلق می‌گیرد؛ درحالی که سؤالات مختلف از نظر محتوایی که مورد پرسش قرار می‌دهند و از نظر دشواری یکسان نیستند که بتوان برای آنها ارزشی برابر قایل شد.

### هدفها و فرضیه‌های پژوهش

سؤالهای مطروح برای پژوهش عبارت‌اند از:

- روایی و پایایی آزمونها چگونه است؟
- دشواری و قدرت تمیز و واریانس هر سؤال چقدر است؟
- شاخصهای هر آزمون چگونه‌اند؟
- میانگین نمرات آزمونها چقدر است؟

هدفهای پژوهش، پاسخگویی علمی به سؤالات

مذکور می‌باشد تا نتایج پژوهش در اختیار طراحان آزمونها گذاشته شود و طراحان در بهبود بخشیدن به کیفیت سؤالات خود و نگهداری سؤالات مناسب در بانک سؤال اقدام کنند.

فرضیه‌های پژوهش عبارت‌اند از:

- ضریب دشواری آزمونها خارج از سطح بهینه است.
- ضریب تمیز آزمونها خارج از سطح بهینه است.
- واریانس آزمونها خارج از سطح بهینه است.
- پایایی آزمونهای زمین‌شناسی بیشتر از پایایی آزمونهای روان‌شناسی است.
- روایی آزمونهای زمین‌شناسی بیشتر از روایی آزمونهای روان‌شناسی است.

### روش

آزمونهای مختلف گروههای آموزشی متعدد دانشگاه که در نیمسال اول ۸۲-۸۱ در مراکز آموزشی تهران اجرا شد، جمعیت آماری پژوهش را تشکیل می‌دهند. برای نمونه‌گیری از هر دانشکده یک گروه آموزشی به قید قرعه انتخاب شد. از دانشکده علوم، گروه آموزشی زمین‌شناسی و از دانشکده

ادبیات و علوم انسانی، گروه آموزشی روان‌شناسی و از هر گروه آزمونهای پنج درس به طور تصادفی تعیین گردید تا تحلیل علمی روی آنها انجام شود.

آزمونهای گروه زمین‌شناسی، دیرینه‌شناسی ۱، دیرینه‌شناسی ۲، زمین‌شناسی اقتصادی، رسوب‌شناسی و سنگ‌شناسی رسوبی و آزمونهای گروه روان‌شناسی، اعتیاد، علم‌الذکر، احساس و ادراک، تفکر و زبان و مقدمات نوروسیکولوژی به قید قرعه انتخاب شدند.

ابزار مورد استفاده، پاسخنامه‌های دانشجویان در آزمونهای ۱۰ درس مذکور است. برای تحلیل آزمونها از شاخصهای دشواری سؤال، قدرت تمیز و واریانس استفاده می‌شود و سپس گزینه‌های انحرافی تحلیل می‌شوند. توزیع نمرات هر آزمون و میانگین نمرات، ضریب همبستگی دو نیمه آزمون و پایایی هر آزمون محاسبه می‌شود.

### نتایج

برای تعداد ۳۷۰ سؤال امتحانی از ۱۰ آزمون گروه‌های زمین‌شناسی و روان‌شناسی، کارت تحلیل سؤال تهیه شد. پاسخهای گروه بالا و پاسخهای گروه پایین به چهار گزینه هر سؤال ثبت گردید. ضریب دشواری، قدرت تمیز و واریانس هر سؤال محاسبه، در کارت درج شد و سؤالات از حیث گزینه‌های انحرافی و شاخصها بر مبنای سؤالات مناسب، سؤالات مناسب و قابل اصلاح و سؤالات نامناسب تفکیک شدند.

میانگین نمرات دانشجویان در هر آزمون، ضریب همبستگی بین دو نیمه آزمونها و ضریب پایایی کلی هر آزمون به دست آمد.

جدول ۱. رتبه‌بندی آزمونهای گروه زمین‌شناسی به ترتیب درصد تعداد سؤالات مناسب و قابل اصلاح آزمون

تعداد سؤال	نامناسب	قابل اصلاح	مناسب	مناسب و قابل اصلاح	آزمون
۴۰	۲۶/۶	۲۳/۳	۵۰	۷۳/۳	سنگ‌شناسی رسوبی
۴۰	۳۲/۵	۲۷/۵	۴۰	۶۷/۵	زمین‌شناسی اقتصادی
۴۰	۴۰	۲۲/۵	۳۷/۵	۶۰	رسوب‌شناسی
۳۰	۴۵	۲۷/۵	۲۷/۵	۵۵	دیرینه‌شناسی ۱
۴۰	۶۷/۵	۱۰	۲۲/۵	۳۲/۵	دیرینه‌شناسی ۲

۷۳/۳ درصد سؤالات آزمون سنگ‌شناسی رسوبی مناسب و قابل زمین‌شناسی، سنگ‌شناسی رسوبی و سپس زمین‌شناسی اصلاح هستند و ۳۲/۵ درصد سؤالات آزمون دیرینه‌شناسی ۲ اقتصادی، رسوب‌شناسی ۱ و دیرینه‌شناسی ۲ هستند. مناسب و قابل اصلاح می‌باشند. بهترین آزمون در گروه

جدول ۲. رتبه‌بندی آزمونهای گروه زمین‌شناسی به ترتیب درصد تعداد سؤالات

مناسب و قابل اصلاح آزمون

تعداد سؤال	نامناسب	قابل اصلاح	مناسب	مناسب و قابل اصلاح	آزمون
۴۰	۱۷/۵	۲۲/۵	۶۰	۸۲/۵	نوروسیکولوژی
۳۰	۳۶/۶	۳۰	۳۳/۳	۶۳/۳	تفکر و زبان
۴۰	۴۲/۵	۳۷/۵	۲۰	۵۷/۵	احساس و ادراک
۳۰	۴۶/۶	۳۳/۳	۲۰	۵۳/۳	اعتیاد
۴۰	۴۷/۵	۳۰	۲۲/۵	۵۲/۵	علم‌النفس

۸۲/۵ درصد سؤالات آزمون نوروسیکولوژی مناسب و نوروسیکولوژی و سپس تفکر و زبان، احساس و ادراک، قابل اصلاح هستند. بهترین آزمون در گروه روان‌شناسی اعتیاد و علم‌النفس هستند.

جدول ۳. مقایسه درصد تعداد سؤالات مناسب و نامناسب در دو گروه آموزشی

سؤالات نامناسب	سؤالات مناسب و قابل اصلاح	گروه آموزشی
۳۸/۱۸	۶۱/۸۲	روان‌شناسی
۴۲/۳	۵۷/۷	زمین‌شناسی

۶۱/۸۲ درصد سؤالات گروه روان‌شناسی مناسب و قابل اصلاح است که این درصد بیشتر از درصد گروه زمین‌شناسی است.

جدول ۴. یافته‌های آماری مربوط به آزمونهای گروه زمین‌شناسی به ترتیب میانگین آزمون

پایه کل آزمون	همبستگی دو نیمه آزمون	واریانس	ضریب تمیز	ضریب دشواری	میانگین	آزمون
۰/۷۴	۰/۵۹	۰/۱۵	۰/۳۱	۰/۷۵	۱۵	دیرینه‌شناسی ۱
۰/۶۹	۰/۶۱	۰/۱۴	۰/۲۳	۰/۷۴	۱۴/۳	دیرینه‌شناسی ۲
۰/۷۸	۰/۶۵	۰/۱۸	۰/۳۶	۰/۶۶	۱۳	سنگ‌شناسی رسوبی
۰/۸۵	۰/۷۵	۰/۱۷	۰/۱۸	۰/۵۸	۱۱/۲	رسوب‌شناسی
۰/۶۰	۰/۴۳	۰/۱۹	۰/۲۳	۰/۵۶	۱۰/۳	زمین‌شناسی اقتصادی

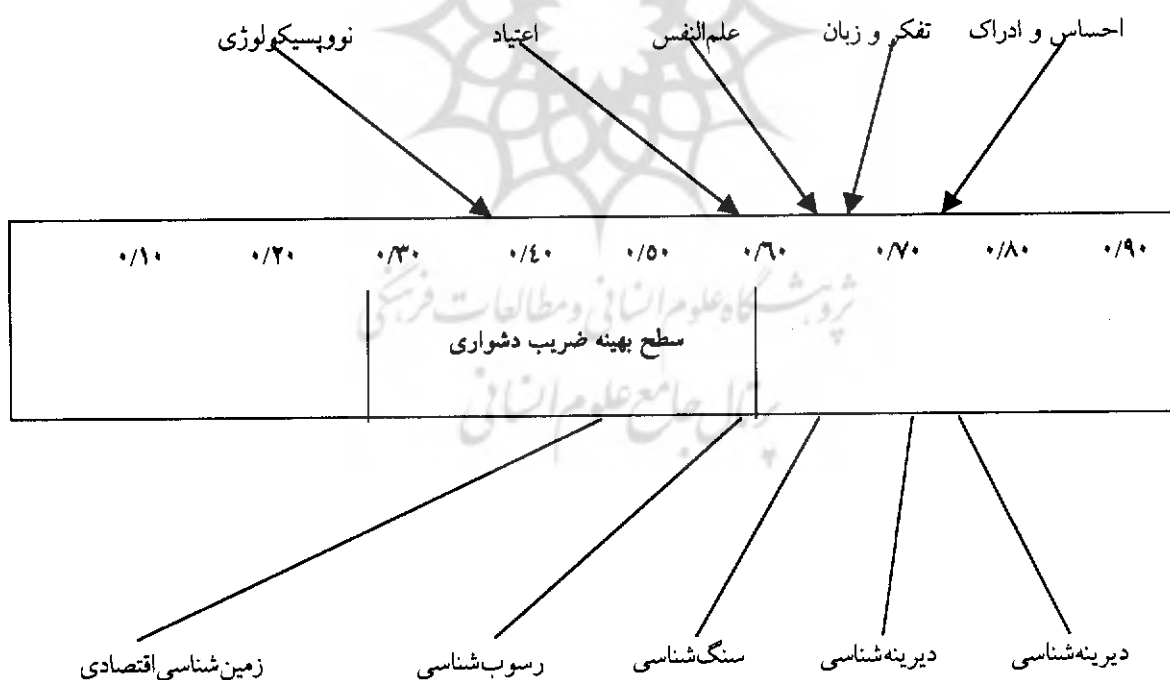
آزمون‌شوندگان در دیرینه‌شناسی، بالاترین میانگین و در زمین‌شناسی اقتصادی پایینترین میانگین را به دست آورده‌اند.

جدول ۵. یافته‌های آماری مربوط به آزمونهای گروه روان‌شناسی به ترتیب میانگین آزمون

پایه کلی آزمون	همبستگی در نیمه آزمون	واریانس	ضریب تمیز	ضریب دشواری	میانگین	آزمون
۰/۶۶	۰/۵۰	۰/۱۷	۰/۲۸	۰/۷۲	۱۴/۴	احساس و ادراک
۰/۷۶	۰/۶۲	۰/۱۷	۰/۲۸	۰/۶۶	۱۳	علم‌النفس
۰/۳۴	۰/۲۱	۰/۱۵	۰/۲۹	۰/۶۸	۱۲/۶	تفکر و زبان
۰/۷۳	۰/۵۸	۰/۱۱	۰/۲۲	۰/۴۰	۱۱/۸	نروپسیکولوژی
۰/۳۶	۰/۲۲	۰/۱۷	۰/۲۳	۰/۶۲	۱۱/۵	اعتیاد

بالاترین میانگین مربوط به آزمون درس احساس و ادراک و پایین‌ترین میانگین مربوط به آزمون درس اعتیاد است.

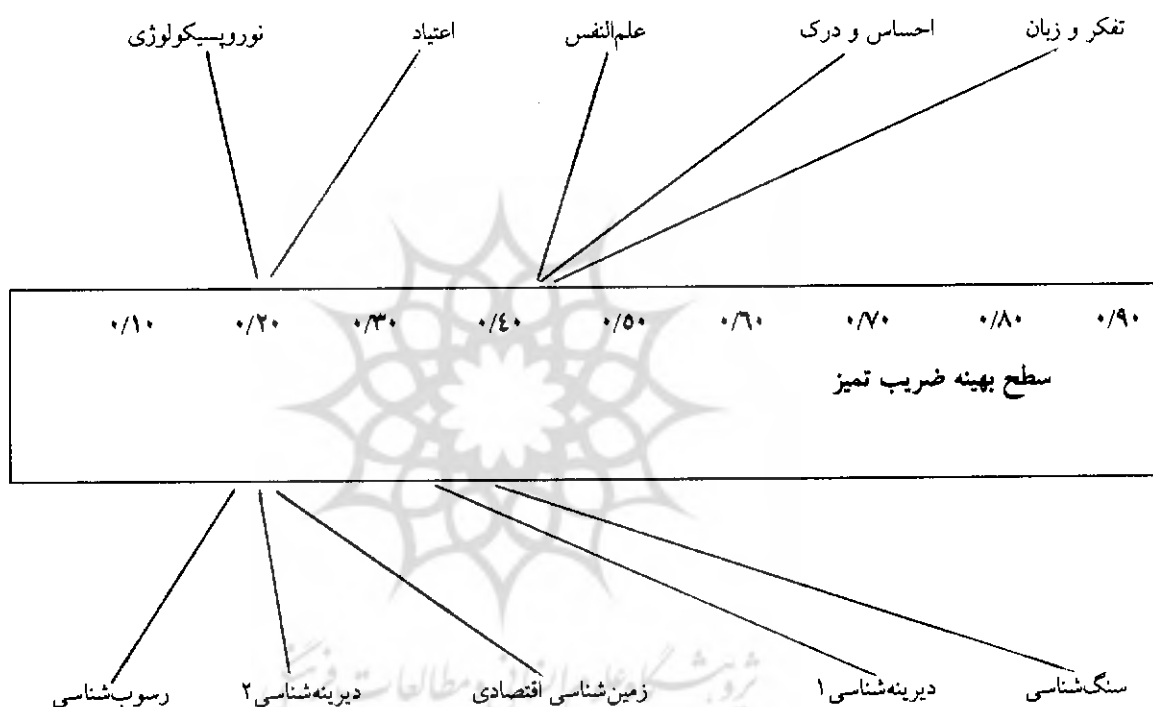
نمودار ۱. جایگاه آزمونهای گروه روان‌شناسی و گروه زمین‌شناسی روی شاخص ضریب دشواری



آزمون نوروپسیکولوژی در داخل سطح بهینه دشواری است و چهار آزمون دیگر به ترتیب از سطح بهینه دور می‌شوند و در سمت سهولت سؤالهای آزمون قرار می‌گیرند.

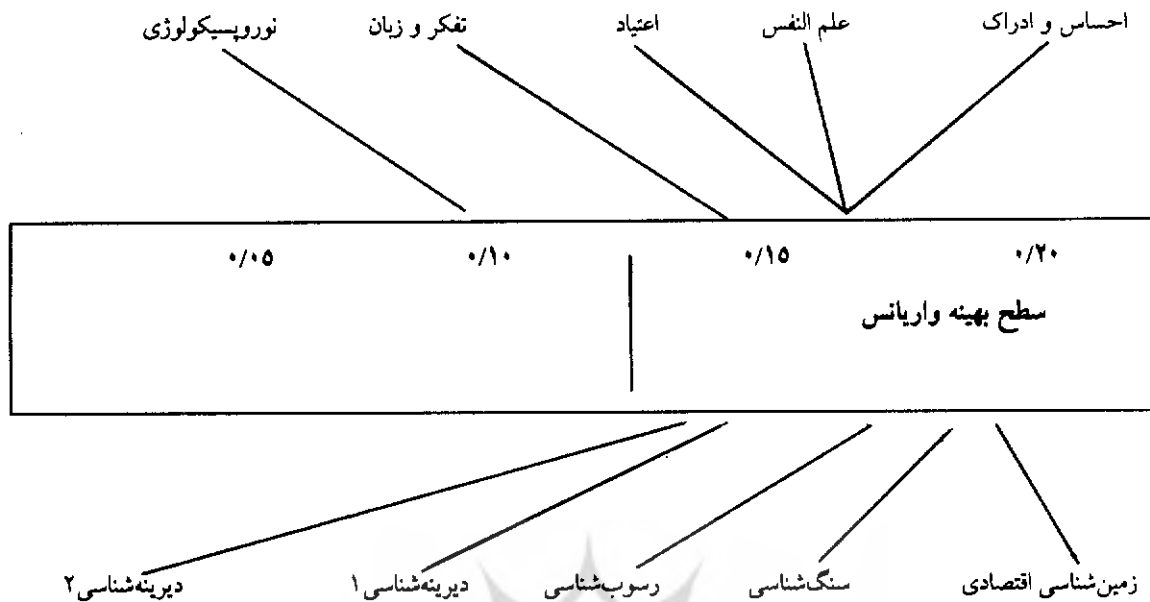
در کل، آزمونهای هر دو گروه خارج از سطح بهینه هستند. روان‌شناسی ۰/۶۱ و زمین‌شناسی ۰/۶۵. آزمونهای گروه زمین‌شناسی ساده‌تر از آزمونهای گروه روان‌شناسی هستند و روایی آزمونهای گروه روان‌شناسی بیشتر است.

نمودار ۲. جایگاه آزمونهای گروه روان‌شناسی و گروه زمین‌شناسی روی شاخص ضریب تمیز



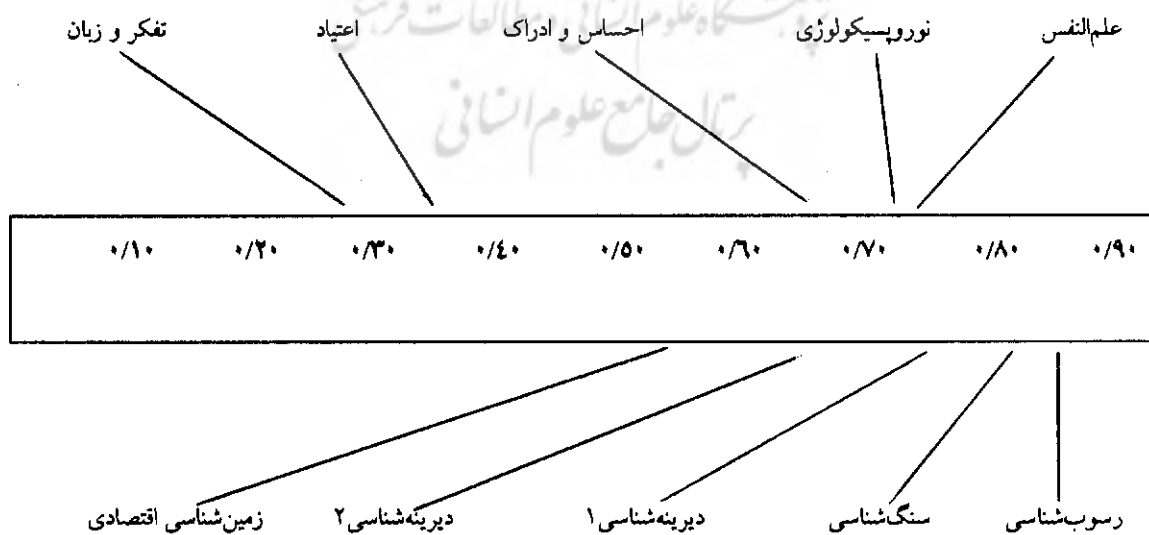
آزمون نوروپسیکولوژی و اعتیاد از گروه روان‌شناسی و آزمونهای رسوب‌شناسی و دیرینه‌شناسی ۲ از قدرت تمیز پایینی برخوردارند و آزمونهای دیگر قدرت تمیز نسبتاً بهتری نشان می‌دهند. قدرت تمیز آزمونهای هر دو گروه ۰/۲۶ است که تقریباً در سطح حداقل قابل قبول قرار دارد.

نمودار ۳. جایگاه آزمونهای گروه روان‌شناسی و گروه زمین‌شناسی روی شاخص واریانس



آزمون نوروسایکولوژی از گروه روان‌شناسی خارج از سطح بهینه واریانس قرار دارند. واریانس کل آزمونهای سطح بهینه و اندک است. آزمونهای دیگر هر دو گروه در زمین‌شناسی ۰/۱۶ و واریانس کل آزمونهای روان‌شناسی ۰/۱۵ است.

نمودار ۴. جایگاه آزمونهای گروه روان‌شناسی و گروه زمین‌شناسی روی شاخص ضریب پایایی



آزمون تفکر و زبان، اعتیاد از گروه روان‌شناسی از پایایی کمی برخوردارند. آزمونهای دیگر هر دو گروه، سطح پایایی خوبی نشان می‌دهند. پایایی آزمونهای گروه زمین‌شناسی بیشتر از گروه روان‌شناسی است.



## بحث

### میانگین ضرایب دشواری آزمونها و میانگین ضرایب

تمیز آزمونهای هر دو گروه خارج از سطح بهینه هستند که فرضیه‌های پژوهش را تأیید می‌کنند. فرضیه دیگر پژوهش در مورد بیرون بودن میانگین واریانس از سطح بهینه رد می‌شود. بالا بودن پایایی آزمونهای گروه زمین‌شناسی از گروه روان‌شناسی نیز تأیید می‌شود.

با توجه به حدود ۴۰ درصد نامناسب بودن سؤالات آزمونهای هر دو گروه، چنانچه در سؤالهای نامناسب و قابل اصلاح تجدیدنظر شود مجموعه سؤالات مفیدی برای نگهداری در بانک سؤال فراهم می‌آید.

تحلیل سؤالهای اول و آخر هر آزمون و رسم منحنی ویژگی سؤال برای آنها نشان می‌دهد که آخرین سؤال آزمونهای دیرینه‌شناسی ۱، سنگ‌شناسی رسوبی، اعتیاد، علم‌النفس، احساس و ادراک، دیرینه‌شناسی ۲ و اولین و آخرین سؤال آزمونهای رسوب‌شناسی، زمین‌شناسی اقتصادی، نوروپسیکولوژی و تفکر و زبان تا حدودی حالت پلکانی دارند و افراد با توجه به ملاک نمره کل خود در آزمون، با نسبت بالاتری به آنها پاسخ داده‌اند.

چون توانایی اعضای علمی دانشگاه در طرح سؤال مطلوب، متفاوت است، پیشنهاد می‌شود از افراد قوی استفاده بیشتری بشود و یا طرح سؤال بهینه و مطلوب به کمیته‌های ویژه واگذار شود. برای آموزش افراد علاقه‌مند، دایر کردن کارگاههای تهیه، تدوین، اجرا و تحلیل آزمون ضروری است.

توصیه می‌شود پژوهش مشابهی در سطح مراکز متعدد یک منطقه قوی و یک منطقه ضعیف اجرا شود. اجرای پژوهش روی سؤالات چند آزمون در سطح کلیه مراکز با استفاده از نرم‌افزارهای کامپیوتری و استفاده از نظریه سؤال - پاسخ در تحلیل سؤالات و آزمونها، اطلاعات علمی ارزشمندی در اختیار دانشگاه قرار می‌دهد.

استفاده از سؤالات و آزمونهای مطلوب که از روایی و پایایی برخوردار باشند، برای اندازه‌گیری پیشرفت تحصیلی دانشجویان و ارزشیابی وضعیت علمی آنان یک ضرورت است.

برای تهیه و تدوین سؤالات چهارگزینه‌ای به آموزش و مهارت نیاز هست. در طرح هر سؤال باید قواعد آزمون‌سازی رعایت گردد و طراحی با توجه به هدفهای آموزشی و طبقات چندگانه شناختی، سؤال خود را تنظیم کند و دقت لازم را در تهیه گزینه‌های انحرافی به کاربرد. آزمونهای معلم ساخته، پس از اجرا نیازمند تحلیل و بازنگری هستند تا طرح از میزان روایی و پایایی و شاخصهای آماری آزمون خود مطلع گردد.

برابر نتایج پژوهش، برخی آزمونها دارای درصد بالاتری از سؤالات مناسب هستند و برخی از آنها درصد بالاتری سؤال نامناسب را در خود جای داده‌اند. در گروه زمین‌شناسی، آزمون سنگ‌شناسی رسوبی و در گروه روان‌شناسی، آزمون نوروپسیکولوژی، بالاترین درصد سؤالات مناسب را نشان می‌دهند.

میانگین برخی آزمونها نظیر دیرینه‌شناسی ۱ در گروه زمین‌شناسی و احساس و ادراک از گروه روان‌شناسی بالاتر و میانگین بقیه آزمونها کمتر است.

بعضی آزمونها نظیر نوروپسیکولوژی، زمین‌شناسی اقتصادی و رسوب‌شناسی در سطح بهینه ضریب دشواری قرار دارند و بقیه دارای ضریب دشواری بالاتر هستند و در بخش سهولت آزمون جای گرفته‌اند. بدیهی است اگر سؤالات خیلی ساده باشد یا خیلی دشوار باشد روایی آزمون آسیب می‌بیند. روایی آزمونهای زمین‌شناسی کمتر از روایی آزمونهای روان‌شناسی به دست آمد.

برخی از آزمونها از حداقل سطح بهینه قدرت تمیز فاصله زیادی دارند و همه ۱۰ آزمون، خارج از سطح بهینه ضریب تمیز هستند و، از این رو، به خوبی نمی‌توانند دو گروه قوی و ضعیف آزمون‌شوندگان را از هم تفکیک کنند.

واریانس آزمونها غیر از آزمون نوروپسیکولوژی که خارج از سطح بهینه است، مناسب می‌باشند. از لحاظ پایایی آزمون، دو آزمون تفکر و زبان و اعتیاد از گروه روان‌شناسی، پایایی کمی نشان می‌دهند و پایایی بقیه آزمونها خوب است. روایی ۱۰ آزمون در سطح ۰/۶۲ می‌باشد.

- منابع
۱. سیف، علی‌اکبر، (۱۳۸۲)، روشهای اندازه‌گیری و ارزشیابی آموزش، تهران، دوران.
  ۲. گنجی، حمزه، (۱۳۷۱)، روان‌شناسی عمومی، تهران، پیام‌نور.
  ۳. ———، (۱۳۷۰)، روانسنجی، تهران، پیام‌نور.
  4. Burney. D.H. (1998). Research methods. Brooksycole publishing company, pacific Grove.
  5. Cohen. R.J. Swedik. M. E. Phillips. S.M.(1996). Psycholoical testing and assessment, may fiet publishing company, California.
  ۸. هومن، حیدرعلی، (۱۳۶۸)، اندازه‌گیریهای روانی و تربیتی، تهران، سلسله.
  ۹. گلاورز، جی. ای، پروینگ، ار. ایچ، (۱۳۷۸)، روان‌شناسی تربیتی، ترجمه علی‌نقی خرازی، تهران، نشر دانشگاهی.
  ۱۰. سلیمی‌زاده، محمدکاظم، (۱۳۷۷) آشنایی با روشهای تجزیه و تحلیل سوالات و آزمونها، تهران، سازمان سنجش کل کشور. ■

