

# بررسی ویژگی‌های روان‌سنجی سؤالات تخصصی آزمون‌های فراگیر رشته روان‌شناسی دانشگاه پیام نور در سال ۱۳۸۵<sup>۱</sup>

جلیل یونسی<sup>۲</sup>

دکتر علی دلاور<sup>۳</sup>

دکتر محمدرضا فلسفی نژاد<sup>۴</sup>

تاریخ پذیرش: ۸۹/۹/۲۰

تاریخ وصول: ۸۹/۴/۲۶

## چکیده

هدف پژوهش حاضر بررسی ویژگی‌های روان‌سنجی سؤالات تخصصی آزمون فراگیر رشته روان‌شناسی دانشگاه پیام نور در سال ۱۳۸۵ بوده است. بدین منظور از بین کلیه شرکت‌کنندگان در آزمون فراگیر رشته روان‌شناسی، یک نمونه ۲۰۰۰ نفری از آزمودنی‌ها به کمک نرم افزار SPSS به صورت تصادفی انتخاب گردید. سپس این نمونه ۲۰۰۰ نفری به صورت تصادفی به دو گروه ۱۰۰۰ تایی تقسیم شد. یکی از این نمونه‌ها جهت برآورد پارامترها و نمونه دیگر جهت برازش مدل با داده‌ها مورد استفاده قرار گرفت. جهت تجزیه و تحلیل آزمون و سؤالات آن طبق نظریه کلاسیک اندازه‌گیری علاوه بر محاسبه توزیع فراوانی گزینه‌های انحرافی و واریانس سؤالات، ضریب دشواری، ضریب تمیز و ضریب پایایی آزمون محاسبه شد. جهت تجزیه و تحلیل آزمون و سؤالات آن براساس نظریه سؤال- پاسخ (IRT)، ابتدا مفروض‌های تک بعدی بودن و استقلال موضعی بررسی گردید. در این تحقیق جهت تعیین تک بعدی بودن، کلیه آزمون‌های تخصصی رشته روان‌شناسی با برنامه TESTFACT مورد تحلیل عاملی قرار گرفتند. نتایج تجزیه و تحلیل آزمون‌های

۱- این پژوهش با حمایت مالی پژوهشکده آموزش باز و از راه دور انجام گرفته است

۲- دانشجوی دکتری سنجش و اندازه‌گیری دانشگاه علامه طباطبائی

۳- عضو هیئت علمی دانشگاه علامه طباطبائی

۴- عضو هیئت علمی دانشگاه علامه طباطبائی

مذکور نشان داد که کلیه آزمون‌ها تک بعدی هستند و در نتیجه مفروضه استقلال موضعی نیز برقرار است. برازش مدل- داده‌ها از طریق نرم افزار BILOG-MG مورد بررسی قرار گرفت و در نهایت پارامترهای برآورد شده سؤال (دشواری، تشخیص و حدس) و پارامتر توانایی آزمودنی‌ها استخراج گردید.

تحلیل گزینه‌های انحرافی نشان می‌دهد که گزینه‌های انحرافی تمام سؤالات، هم احتمال نیستند و عملکرد معیوبی داشته‌اند. همچنین نتایج نشان داد که در آزمون‌های روان‌شناسی و جامعه‌شناسی مدل دو پارامتری و در آزمون فلسفه مدل سه پارامتری برازش بهتری با مجموعه داده‌های آزمون دارد. در عین حال علیرغم این که میانگین تمامی آزمون‌ها پایین تر از نمره ملاک است، اما به نظر می‌رسد که مشکلات در درس فلسفه بیشتر بوده و این آزمون در شکست تحصیلی آزمودنی‌ها تأثیر بیشتری داشته است.

**واژگان کلیدی:** CTT، JRT، TESTFACT، BILOG-MG، آزمون فراگیر دانشگاه پیام نور.

#### مقدمه

در قرن حاضر با تعارضات و بحران‌های اجتماعی و اقتصادی شدیدی روبرو هستیم. بخصوص در کشورهای جهان سوم و توسعه نیافته این بحران‌ها جدی‌تر است. آموزش این توانایی را دارد تا برای بهبود و پیشرفت کشورهای توسعه نیافته زمینه لازم را فراهم سازد و سیاست‌های لازم برای مقابله با این بحران جهانی و بحران‌های منطقه‌ای و ملی را اتخاذ نماید (لوئیز<sup>۱</sup>، ۱۹۹۰، به نقل از سهرابی، ۱۳۸۴).

هنگامی که به روند تکاملی کشورهای توسعه یافته توجه می‌کنیم همواره شاهد ارتباط نزدیک بین آموزش، پژوهش علمی و رشد فناوری از یک سو و پیشرفت سیاسی، اقتصادی و اجتماعی از سوی دیگر می‌باشیم. آموزش عالی نقش ارزنده‌ای در تربیت نیروی انسانی مورد نیاز و افزایش کیفیت نیروی انسانی موجود دارد و توسعه فرهنگی، اقتصادی،

---

1. Luiz

اجتماعی و سیاسی به صورت مداوم نیازمند گسترش و نوسازی آموزش عالی است. از سوی دیگر محدودیت‌های نظام آموزش عالی کشورمان و افزایش تقاضای آموزش عالی و اقبال عمومی به آن باعث شده است که سالانه حدود ۱/۵ میلیون نفر داوطلب آموزش عالی باشند، در صورتی که نظام آموزش عالی کشورمان توانایی جذب بیش از یک چهارم آن را ندارد و به دلیل ساختار جمعیتی کشورمان چنانچه راه‌های جدیدی برای افزایش دسترسی به آموزش عالی فراروی متقاضیان قرار نگیرد، سایر افراد باید از حقوق اولیه خود یعنی آموزش چشم‌پوشند. در چنین شرایطی با توجه به افزایش تقاضای آموزش عالی و پذیرش این مسأله که آموزش علاوه بر اینکه حقوق اولیه اجتماعی هر فرد است، پیش‌نیاز توسعه پایدار هم می‌باشد، بنابراین بایستی به دنبال راهکارهایی به منظور گسترش و تنوع بخشیدن به نظام آموزش عالی کشورمان باشیم. یکی از این راهکارها توسعه نظام آموزش باز و از راه دور می‌باشد.

### مسئله پژوهش

نظام آموزش باز و از راه دور دانشگاه پیام نور به دلیل ویژگی‌های ساختاری خود، توان توسعه سریع کمی و کیفی را در زمان‌ها و مکان‌های مختلف و همچنین سطوح مختلف آموزش را دارد. از این رو گسترش آموزش و همگانی کردن آن در همه سطوح به ویژه آموزش عالی با رویکرد به نظام آموزش باز و از راه دور از طریق دوره‌های فراگیر مورد توجه سیاست‌گذاران و برنامه‌ریزان آموزشی در کشورهای مختلف می‌باشد. در کشورهای پیشرفته این نظام آموزشی بیشتر در خدمت کسانی است که به دلایلی نمی‌خواهند یا نمی‌توانند از آموزش سنتی تمام وقت برخوردار شوند ولی در کشورهای جهان سوم، این نظام بخشی از پیکره آموزش کشور است و در حقیقت همان کارکرد نظام سنتی را دارد. یکی از ویژگی‌های دانشگاه‌های باز<sup>۱</sup> تکیه بر آمادگی و توانمندی‌های داوطلبان ورود به آموزش عالی برای برنامه‌های تحصیلی ویژه رشته و گرایش تخصصی است که داوطلب علاقه‌مند به تحصیل در آن است. در این راستا شرکت در دوره‌های فراگیر برای هر

---

1. Open University

رشته‌ای که داوطلب علاقه تحصیلی آن را داشته باشد، بدون توجه به زمینه تحصیلی قبلی وی امکان‌پذیر است و برای سنجش میزان توانمندی‌ها و آمادگی تحصیلی داوطلبان به آنان اجازه داده شده که در رشته مورد علاقه به عنوان دانش‌پذیر ثبت نام نموده و تعدادی از واحدهای درسی رشته انتخابی خویش را طبق برنامه دانشگاه مطالعه و با کسب آمادگی‌های علمی لازم در آزمون پایان نیمسال دانش‌پذیری که توسط سازمان سنجش آموزش کشور به صورت سراسری برگزار می‌شود، شرکت نمایند.

نکته قابل توجه آن است که داوطلب فراگیر از فیلتری به نام آزمون سراسری عبور نکرده است و آزمون پایان ترم اول دانش‌پذیری، جایگزین آزمون سراسری شده است و آزمون دانش‌پذیری صرفاً از دروس اختصاصی رشته تحصیلی فراگیر خواهد بود. به عبارتی سؤالات آزمون‌های فراگیر نسبت به سؤالات آزمون‌های سراسری فنی‌تر و تخصصی‌تر هستند و کسب حد نصاب‌های علمی لازم در این آزمون منجر به تبدیل وضعیت فراگیر از دانش‌پذیری به دانشجویی خواهد شد. به همین جهت توجه به اصول و قواعد آزمون‌سازی و تحلیل ویژگی‌های روان‌سنجی سؤالات آزمون‌های فراگیر از اهمیت فوق‌العاده‌ای برخوردار خواهد بود. چرا که عدم توجه به اصول مذکور موجب می‌شود که آزمون‌ها در عمل، چیزی جز خطا را اندازه‌گیری نکنند و در نتیجه براساس اطلاعات غلط، تصمیمات غلط اتخاذ خواهد شد. بنابراین پژوهش حاضر درصدد پاسخگویی به این پرسش اساسی است که ویژگی‌های روان‌سنجی سؤالات تخصصی آزمون‌های فراگیر (رشته روان‌شناسی) چگونه است؟

### هدف پژوهش

پژوهش حاضر با هدف بررسی ویژگی‌های روان‌سنجی سؤالات آزمون‌های فراگیر رشته روان‌شناسی دانشگاه پیام نور در سال ۱۳۸۵ صورت گرفته است. در عین حال محقق ویژگی‌های سؤالات اختصاصی آزمون‌های مذکور را طبق نظریه کلاسیک و نظریه سؤال-پاسخ مورد بررسی قرار داده و به برآزش مدل‌های IRT با مجموعه داده‌های آزمون پرداخته است.

### سؤالات پژوهش

۱. ویژگی‌های سؤالات اختصاصی آزمون‌های فراگیر رشته روان‌شناسی بر طبق نظریه کلاسیک چگونه است؟
۲. ویژگی‌های سؤالات اختصاصی آزمون‌های فراگیر رشته روان‌شناسی بر طبق نظریه IRT چگونه است؟
۳. کدامیک از مدل‌های IRT بهترین برازش را با این مجموعه از داده‌ها دارد؟

### مبانی نظری و پیشینه پژوهش

در حال حاضر دو نظریه غالب یعنی نظریه نمونه‌گیری تصادفی<sup>۱</sup> و نظریه سؤال-پاسخ<sup>۲</sup> در روان‌سنجی وجود دارد. در نظریه نمونه‌گیری تصادفی به رابطه عملکرد مشاهده شده و عملکرد واقعی همانند تعمیم از یک نمونه به یک جامعه نگریسته می‌شود، در حالی که بر اساس نظریه سؤال-پاسخ اگر چگونگی ارتباط هر سؤال آزمون با فرد مورد مطالعه درک شود، می‌توان نمره واقعی فرد را به صورت مستقیم برآورد کرد. نظریه نمونه‌گیری تصادفی شامل دو نظریه نسبتاً متفاوت یعنی نظریه کلاسیک<sup>۳</sup> یا پایایی کلاسیک و نظریه تعمیم‌پذیری<sup>۴</sup> است (کیامنش، ۱۳۷۸).

نظریه کلاسیک علیرغم ضعف‌ها و کاستی‌ها، هنوز مورد استفاده قرار گرفته و به عبارتی با توجه به اهمیتی که این نظریه کلاسیک در تجزیه و تحلیل آزمون‌های مختلف روانی تربیتی دارد، در بررسی آزمون‌ها به طور وسیع مورد استفاده قرار می‌گیرد. نکته قابل توجه آن است که نظریه کلاسیک در مقابل نظریه جدید سؤال-پاسخ نبوده بلکه به عنوان مکمل آن می‌باشد (لرد<sup>۵</sup>، ۱۹۸۰؛ نانالی<sup>۶</sup> و برنشتاین<sup>۷</sup>، ۱۹۹۴).

- 
1. Random Sampling Theory
  2. Item Response Theory (IRT)
  3. Classical Test Theory (CTT)
  4. Generalizability Theory (GT)
  5. Lord
  6. Nunnally
  7. Bernstein

### مروری بر پژوهش‌های انجام شده در ارتباط با موضوع پژوهش

در زمینه تحلیل روان‌سنجی سؤالات، تحقیقات ولی زاده (۱۳۷۸) و ستاری (۱۳۷۷) را می‌توان نام برد. ولی زاده در تحقیقی به تحلیل مشخصات روان‌سنجی آزمون ریاضی، فیزیک و شیمی ورودی دانشگاه پیام نور دوره نظام جدید (۱۳۷۹-۱۳۷۸) براساس نظریه سؤال- پاسخ پرداخته است. در واقع این تحقیق به منظور تعیین چگونگی اندازه‌گیری سطوح چیرگی آزمون‌های مذکور براساس نظریه سؤال- پاسخ و همچنین تعیین چگونگی تابع آگاهی در سطح چیرگی انجام شده است. نتایج تحقیق وی نشان داد که تابع آگاهی آزمون‌های ریاضی، فیزیک و شیمی در سطح چیرگی از کشیدگی قابل توجهی برخوردار نیست و سطح چیرگی این آزمون‌ها به دقت اندازه‌گیری نمی‌شود ولی در آزمون شیمی در سطح چیرگی تعیین شده، تابع آگاهی تا حدی مناسب بوده و سطح چیرگی تا حدی به طور مناسب اندازه‌گیری می‌شود. در مورد تعداد سؤالات نیز مشخص گردید که سؤالات هر سه آزمون در حدی بودند که می‌توانستند سطح چیرگی را با دقت اندازه‌گیری کنند.

ستاری (۱۳۷۷) کاربرد مدل سه پارامتری نظریه سؤال- پاسخ در تجزیه و تحلیل و ساخت آزمون‌های توانایی را با استفاده از نرم افزار ASCAL در خصوص سؤالات آزمون ریاضی (۳۰ سؤالی) گروه علوم تجربی کنکور سراسری سال ۱۳۷۵ و بر روی نمونه ۲۰۰۰ نفری بررسی کرده است. وی در این تحقیق علاوه بر برآورد پارامتر توانایی افراد و پارامترهای سؤال (پارامترهای دشواری، تمیز و حدس) نسبت به ترسیم منحنی ویژگی سؤالات، منحنی ویژگی آزمون و ترسیم منحنی آگاهی آزمون اقدام کرده و در نهایت داده‌ها و منحنی‌های بدست آمده از اجرای مدل سه پارامتری را تفسیر کرده است.

در خارج از کشور به دلیل نیاز به سؤالات کارآمد و مناسب، تمامی شرکت‌های آزمون‌سازی مثل شرکت خدمات سنجش آموزشی (ETS)<sup>۱</sup> هر ساله سؤالات آزمون‌های خود (نظیر TESOL و TOEFL) را مورد تجزیه و تحلیل روان‌سنجی قرار داده و سؤالات خود را به طور مداوم پالایش می‌کنند.

### روش اجرای پژوهش

پژوهش حاضر از نوع توصیفی است. لازم به ذکر است که تحقیق توصیفی شامل مجموعه روش‌هایی است که هدف آنها توصیف کردن شرایط یا پدیده‌های مورد بررسی است و اجرای این گونه تحقیق‌ها می‌تواند صرفاً برای شناخت بیشتر شرایط موجود یا یاری دادن به فرایند تصمیم‌گیری باشد.

### جامعه آماری و گروه نمونه مورد مطالعه

جامعه آماری مورد مطالعه، کلیه سؤالات اختصاصی آزمون فراگیر رشته روان‌شناسی دانشگاه پیام نور در سال ۱۳۸۵ می‌باشد. لازم به ذکر است که از آزمودنی‌ها اقدام به نمونه‌گیری به عمل آمده و به کمک نرم‌افزار SPSS یک نمونه ۲۰۰۰ نفری به صورت تصادفی از بین کلیه آزمودنی‌های رشته فراگیر روان‌شناسی سال ۱۳۸۵ انتخاب گردید. سپس این نمونه ۲۰۰۰ نفری به صورت تصادفی به دو گروه ۱۰۰۰ نفری تقسیم شد. یکی از این نمونه‌ها جهت برآورد پارامترها (نمونه ضریب‌یاب<sup>۱</sup>) استفاده شد و نمونه دیگر جهت برآزش مدل با داده‌ها<sup>۲</sup> (نمونه ارزشیاب<sup>۳</sup>) مورد استفاده قرار گرفت (ر. ک. لرد و ناویک<sup>۴</sup>، ۱۹۶۸، ص ۲۸۵).

در نمونه‌گیری و تعیین حجم نمونه آزمودنی‌ها موارد زیر مد نظر قرار گرفت: اولاً با توجه به این که نظریه‌های روان‌سنجی، نظریه‌های نمونه‌های بزرگ هستند، جهت برآورد با ثبات پارامترها، نمونه‌ای با حجم بالا انتخاب گردید. ثانیاً آزمودنی‌هایی که در آزمون‌ها نمره کامل یا نمره صفر گرفته بودند از جامعه اصلی حذف شدند و سپس عمل نمونه‌گیری از بقیه آزمودنی‌ها در هر آزمون صورت پذیرفت. جدول ۱ توزیع فراوانی آزمودنی‌ها در جامعه و نمونه به تفکیک جنسیت را نشان می‌دهد:

- 
1. Calibration Sample
  2. Model – Data Fit
  3. Validation Sample
  4. Novick

جدول ۱. توزیع فراوانی جنسیت در جامعه و نمونه

گروه‌ها	توزیع فراوانی جامعه					
	توزیع فراوانی جامعه		نمونه ضریب یاب		نمونه ارزشیاب	
جنسیت	فراوانی	در صد	فراوانی	در صد	فراوانی	در صد
مرد	۲۰۳۵	۲۱/۰	۲۰۲	۲۰/۲	۱۹۸	۱۹/۸
زن	۷۶۷۶	۷۹/۰	۷۹۸	۷۹/۸	۸۰۲	۸۰/۲
کل	۹۷۱۱	۱۰۰/۰	۱۰۰۰	۱۰۰/۰	۱۰۰۰	۱۰۰/۰

## روش جمع‌آوری اطلاعات

داده‌های اولیه و اصلی جهت استفاده در تجزیه و تحلیل سؤالات اختصاصی آزمون فراگیر رشته روان‌شناسی، پاسخ‌های کلیه شرکت‌کنندگان در آزمون پایان ترم دوره دانش‌پذیری دانشگاه پیام نور بوده است. جدول ۲ مشخصات مربوط به مواد درسی آزمون‌های مذکور را نشان می‌دهد.

جدول ۲. دروس و منابع درسی آزمون فراگیر رشته روان‌شناسی

نام درس	نام کتاب	مؤلف	تعداد سؤال
روان‌شناسی	روان‌شناسی هیلگارد	مترجم رفیعی	۲۰
جامعه‌شناسی	درآمدی بر جامعه‌شناسی	بروس کوئن، مترجم ثلاثی	۲۰
فلسفه	کلیات فلسفه	دادبه	۲۰

## ابزارها و روش تحلیل داده‌ها

جهت تجزیه و تحلیل اطلاعات جمع‌آوری شده از نرم افزارهای کامپیوتری SPSS، EXCEL، TESTFACT و BILOG-MG استفاده گردید.

لازم به ذکر است که به منظور تجزیه و تحلیل آزمون و سؤالات آن طبق نظریه کلاسیک اندازه‌گیری علاوه بر محاسبه توزیع فراوانی گزینه‌های انحرافی و واریانس سؤالات، ضریب دشواری، ضریب تمیز و ضریب پایایی آزمون محاسبه شد. جهت تجزیه



و تحلیل آزمون و سؤالات آن براساس طبق نظریه سؤال-پاسخ (IRT)، ابتدا مفروضه‌های تک بعدی بودن<sup>۱</sup> و استقلال موضعی<sup>۲</sup> بررسی گردید.

برای سنجش تک بعدی بودن آزمون روش‌های متعددی پیشنهاد شده است. برای نمونه هتیه<sup>۳</sup> ۳۰ روش را پیشنهاد کرده است. در این تحقیق جهت تعیین تک بعدی بودن آزمون‌ها، روش پیشنهادی همبلتون<sup>۴</sup> مورد استفاده قرار گرفت، بدین صورت که کلیه آزمون‌های تخصصی رشته روان‌شناسی با برنامه TESTFACT مورد تحلیل عاملی قرار گرفتند.

این نرم افزار بدان دلیل مورد استفاده قرار گرفته است که تحلیل عاملی را براساس رویکرد غیرخطی<sup>۵</sup> محاسبه می‌نماید. والر<sup>۶</sup> (۲۰۰۰) اشاره کرده است که وقتی تحلیل عاملی خطی برای داده‌های چند ارزشی به کار می‌رود، نتایج تورش‌داری بدست می‌دهد. مک دونالد<sup>۷</sup> و آهلاوت (۱۹۷۴) نیز نشان داده‌اند که وقتی شکل رگرسیون‌های سؤال-عامل غیرخطی به گونه قابل توجهی در همه سؤالات تفاوت داشته باشد، مدل تحلیل عاملی خطی، عامل‌های کاذب تولید می‌کند (هومن و عسگری، ۱۳۸۵). بنا به دلایل فوق برای بررسی مفروضه تک بعدی بودن آزمون‌ها از برنامه TESTFACT استفاده گردید. چرا که اولاً نسبت به برنامه‌های دیگر تحلیل عاملی ارجح‌تر است و اطلاعات کامل‌تری را در اختیار ما قرار می‌دهد و ثانیاً در تشکیل ماتریس همبستگی از ضریب همبستگی تراکوریک<sup>۸</sup> به جای ضریب همبستگی پیرسون استفاده می‌کند (دو تویت<sup>۹</sup>، ۲۰۰۳).

جهت تعیین تعداد ابعاد آزمون‌ها می‌توان از نسبت حاصل ارزش ویژه<sup>۱۰</sup> عامل اول به ارزش ویژه عامل دوم استفاده کرد. چنانچه ارزش ویژه عامل اول، دو تا سه برابر ارزش

- 
1. Unidimensionality
  2. Local Independence
  3. Hattie
  4. Hambleton
  5. Nonlinear
  6. Waller
  7. Mc Donald8. Tetrachoric Correlation
  8. Tetrachoric Correlation
  9. du Toit, Mathilda
  10. Eigenvalue

ویژه عامل دوم باشد، می‌توان نتیجه گرفت که آزمون مورد نظر تک بعدی است، یا به عبارتی دیگر یک عامل غالب، عملکرد افراد در آزمون را تبیین می‌کند. یکی دیگر از روش‌های بررسی تک بعدی بودن، نمودار اسکری<sup>۱</sup> است.

باید توجه داشت که استقلال موضعی، به طور خودکار از تک بعدی بودن پیروی می‌کند و مفروضه‌ای جداگانه به حساب نمی‌آید (لرد، ۱۹۸۰؛ همبلتون، ۱۹۸۹).

برآزش مدل-داده‌ها از طریق نرم افزار BILOG-MG مورد بررسی قرار گرفت و در نهایت پارامترهای برآورد شده سؤال (دشواری، تشخیص و حدس) و پارامتر توانایی آزمودنی‌ها استخراج گردید.

در پژوهش حاضر به این دلیل که پارامترهای سؤال و توانایی آزمودنی‌ها نامعلوم بوده‌اند، از روش‌های برآورد همزمان پارامترهای سؤال و توانایی استفاده گردیده است. سپس منحنی ویژگی سؤال، تابع آگاهی سؤال، تابع آگاهی آزمون و منحنی خطای استاندارد آزمون از طریق نرم افزار مذکور ترسیم گردید.

به منظور بررسی ثبات برآورد پارامترهای سؤال که به معنای عدم تأثیر پذیری پارامترهای سؤال از زیر گروه‌های (نمونه‌های) آزمودنی جامعه مورد نظر است، نمونه مذکور (۲۰۰۰ نفر) به دو زیر گروه ۱۰۰۰ نفری تقسیم شد. روش کار چنین بوده است که ابتدا مقادیر پارامترهای سؤالات از طریق نمونه ضریب یاب برآورد گردید، سپس مقادیر به دست آمده به عنوان مقادیر اولیه در نمونه ارزشیاب قرار داده شد و مجدداً پارامترهای سؤالات برآورد گردید و نمودارهای آنها با هم مقایسه شد. در ادامه، بیشینه آگاهی سؤالات، بیشینه آگاهی آزمون و همچنین نقطه‌ای که بیشینه آگاهی در آنها به وجود می‌آید آمده است. در نهایت محاسبه گردید که بیشترین میزان آگاهی‌دهندگی آزمون‌ها در چه دامنه‌ای قرار دارد.

یافته‌های پژوهش

الف) درس روان‌شناسی

جدول ۳ نتایج تحلیل سؤالات درس روان‌شناسی را به تفکیک دو نظریه کلاسیک و سؤال- پاسخ نشان می‌دهد.

جدول ۳. تحلیل سؤالات درس روان‌شناسی براساس نظریه کلاسیک و نظریه سؤال- پاسخ

سؤال	الف	ب	ج	د	تحلیل پاسخ‌ها بر اساس نظریه کلاسیک								
					میانگین نمره صحیح	میانگین نمره نادرست	میانگین نمره کل	میانگین نمره صحیح	میانگین نمره نادرست	میانگین نمره کل	میانگین نمره صحیح		
۱	۱۲۰	۶۱	*۴۶۸	۲۱۳	۱۳۸	۰/۵۵	۰/۲۸	۰/۲۵	۰/۶۴	۰/۳۰۰	۰/۲۲۷	۱۲/۴	۰/۱۹
۲	۷۴	۲۱۱	۲۹۷	*۲۱۶	۲۰۲	۰/۲۸	۰/۲۸	۰/۲۰	۰/۶۴	۲/۸۵۲	۰/۲۶۶	۳/۹	۰/۹۲
۳	۱۹۶	*۱۸۷	۱۶۴	۱۲۸	۲۲۵	۰/۳۷	۰/۲۷	۰/۲۳	۰/۶۴	۲/۰۸۸	۰/۲۶۴	۱۴/۹	۰/۰۶
۴	۹۸	*۳۰۱	۲۶۴	۲۰۶	۱۳۱	۰/۳۴	۰/۳۴	۰/۲۲	۰/۶۳	۱/۴۵۳	۰/۳۹۳	۱۸/۶	۰/۰۳
۵	۴۳۶	۳۲	*۳۳۴	۹۷	۱۰۱	۰/۳۸	۰/۳۲	۰/۲۴	۰/۶۳	۱/۳۱۳	۰/۳۲۰	۱۵/۲	۰/۰۸
۶	۲۱۵	۱۰۸	۲۱۱	*۳۷۸	۸۸	۰/۴۱	۰/۲۳	۰/۲۴	۰/۶۴	۱/۷۲۴	۰/۱۷۵	۱۱/۱	۰/۲۷
۷	۴۳۳	۴۸	*۱۶۶	۱۱۷	۲۳۶	۰/۲۰	۰/۴۱	۰/۱۶	۰/۶۲	۱/۸۱۶	۰/۶۶۸	۱۳/۹	۰/۰۵
۸	*۴۴۶	۷۳	۱۹۳	۱۲۹	۱۵۹	۰/۵۴	۰/۴۵	۰/۲۵	۰/۶۱	۰/۲۵۰	۰/۶۷۸	۳۲/۳	*۰/۰۰
۹	۱۷۹	*۴۹۲	۸۸	۱۰۳	۱۳۸	۰/۵۶	۰/۳۱	۰/۲۵	۰/۶۳	۰/۱۰۸	۰/۳۷۱	۳۸/۰	*۰/۰۰
۱۰	۱۱۷	۶۵	۱۵۹	*۴۰۴	۲۲۵	۰/۵۶	۰/۳۱	۰/۲۵	۰/۶۳	۰/۹۲۵	۰/۲۴۷	۱۶/۹	۰/۰۵
۱۱	*۲۱۶	۱۲۵	۴۵۱	۱۱۰	۹۸	۰/۲۴	۰/۲۲	۰/۱۸	۰/۶۴	۳/۵۲۲	۰/۲۱۷	۸/۲	۰/۵۱
۱۲	۲۲۱	۹۹	۱۰۳	*۴۳۷	۱۴۰	۰/۵۲	۰/۴۶	۰/۲۵	۰/۶۱	۰/۲۵۳	۰/۷۴۷	۴۹/۷	*۰/۰۰
۱۳	*۳۹۰	۱۱۵	۱۴۴	۲۴۰	۲۴۰	۰/۵۰	۰/۴۶	۰/۲۵	۰/۶۱	۰/۴۸۶	۰/۷۷۳	۳۶/۸	*۰/۰۰
۱۴	۲۰۰	۲۱۹	*۱۲۵	۲۲۸	۲۲۸	۰/۱۶	۰/۰۹	۰/۱۳	۰/۶۵	۱۰/۱۱۳	۰/۱۰۹	۱۷/۶	۰/۰۴
۱۵	۱۷۶	۵۶۲*	۱۳۹	۶۰	۶۳	۰/۵۸	۰/۳۸	۰/۲۴	۰/۶۲	-/۲۹۸	۰/۴۷۲	۳۵/۵	*۰/۰۰
۱۶	۴۳	۶۶	*۳۳۷	۳۰۷	۲۴۷	۰/۴۳	۰/۴۰	۰/۲۵	۰/۶۲	۱/۰۴۵	۰/۴۷۱	۲۱/۸	*۰/۰۰
۱۷	۳۴۸	*۲۶۵	۹۰	۱۰۲	۱۹۵	۰/۳۱	۰/۴۴	۰/۲۱	۰/۶۱	۱/۱۸۶	۰/۶۵۴	۱۹/۰	۰/۰۱
۱۸	۱۱۰	۲۳۳	۱۶۴	۱۶۳*	۳۳۰	۰/۲۴	۰/۳۴	۰/۱۸	۰/۶۳	۲/۴۲۷	۰/۴۵۰	۱۸/۷	۰/۰۱
۱۹	۲۲۳	۲۵۵	*۲۴۵	۱۰۱	۱۷۶	۰/۳۰	۰/۲۹	۰/۲۱	۰/۶۴	۲/۶۷۱	۰/۲۵۹	۶/۴	۰/۶۰
۲۰	*۱۲۵	۶۵	۲۴۹	۲۱۶	۱۴۵	۰/۳۸	۰/۵۱	۰/۲۴	۰/۶۰	۰/۶۷۲	۰/۹۱۷	۴۰/۷	*۰/۰۰

-2 Log Likelihood	۰/۴۳۴	۱/۷۴۶	--	۰/۳۴	۰/۳۹	میانگین	
=	۴۶۱۰۳/۳۶	۰/۲۳۱	۲/۲۲۲	--	۰/۱۰	۰/۱۳	انحراف معیار
* عدم برازش سؤال با مدل در سطح ۰/۰۱				* گزینه صحیح (کلید سؤال)			

در تحلیل ویژگی‌های روان‌سنجی سؤالات آزمون روان‌شناسی طبق نظریه کلاسیک، سؤالات ۷ و ۱۴ درجه دشواری کمتر از ۰/۳۰ یا بالاتر از ۰/۷۰ دارند و این مطلب بدان معنی است که این سؤالات خیلی آسان و یا خیلی دشوار بوده‌اند و در نتیجه اطلاعات زیادی را درباره تفاوت‌های آزمودنی‌ها ارائه نمی‌کنند. همچنین از نظر ضریب تمیز، به غیر از سؤال ۱۴ بقیه سؤالات دارای وضعیت مناسبی می‌باشند. ضریب پایایی کل آزمون با استفاده از فرمول آلفای کرونباخ ۰/۶۴ تعیین گردید، سپس بعد از حذف هر سؤال میزان پایایی مجدداً مورد بررسی قرار گرفت. نتایج نشان داد که با حذف سؤال ۱۴ میزان ضریب پایایی افزایش پیدا می‌کند. بنابراین، این سؤال تجانس کمتری با بقیه داشت. همچنین نحوه پاسخ‌گویی آزمودنی‌ها به گزینه‌های انحرافی نشان می‌دهد که در نیمی از سؤالات آزمون نحوه توزیع پاسخ‌ها به گزینه‌های انحرافی نامناسب می‌باشد و گزینه‌های انحرافی از این حیث، عملکرد معیوبی داشته‌اند.

تحلیل ویژگی‌های روان‌سنجی سؤالات آزمون روان‌شناسی طبق نظریه IRT نشان داد که این آزمون تک بعدی است؛ لذا مفروضه استقلال موضعی نیز برقرار است. در جدول ۴ درصد واریانس هر یک از دو عامل استخراج شده سه آزمون مورد بررسی به همراه مقادیر ویژه نشان داده شده است.

جدول ۴: نتایج بررسی مفروضه‌های IRT در سه آزمون مورد بررسی

روان‌شناسی		جامعه‌شناسی		فلسفه	
مقادیر ویژه	در صد واریانس	مقادیر ویژه	در صد واریانس	مقادیر ویژه	در صد واریانس
۳/۳۸۸	۱۵/۸۸۱	۴/۶۳۸	۲۲/۷۷۵	۲/۸۳۸	۱۳/۵۹
۰/۵۱۴	۲/۴۱۸	۰/۴۷۴	۲/۲۹۱	۰/۶۲۷	۳/۱۰

آزمون‌های دو نشان داد که مدل دو پارامتری نسبت به مدل‌های یک و سه پارامتری با داده‌های آزمون روان‌شناسی برازش بهتری دارد. پس از برآورد پارامترها میزان برازندگی تک تک سؤالات با مدل‌های یک، دو و سه پارامتری محاسبه شد. نتایج نشان داد در مدل سه پارامتری ۱۲ سؤال، در مدل دو پارامتری ۷ سؤال و در مدل یک پارامتری ۱۳ سؤال با مدل‌های مربوطه برازش ندارند که در بین این سؤالات، ۸، ۹، ۱۲، ۱۳، ۱۵، ۱۶ و ۲۰ با هیچکدام از مدل‌های مذکور برازش ندارند و باید حذف گردند. بررسی منحنی ویژگی سؤالات حاکی از آن است که کل سؤالات آزمون، قدرت تشخیص نسبتاً پایینی دارند و در تفکیک آزمودنی‌های قوی و ضعیف چندان موفق عمل نمی‌کنند. در بررسی منحنی ویژگی سؤالات شیب ملایم منحنی نیز مبین همین مسأله بود. اکثر سؤالات این آزمون دشواری متوسطی داشتند و حدود ۶ سؤال دشواری بالایی داشتند.

جدول شماره ۵ نشان می‌دهد که بر طبق نظریه کلاسیک و IRT کدام سؤالات از نظر دشواری و قدرت تشخیص مناسب و کدام سؤالات نامناسب هستند. ملاک نامناسب بودن سؤال از نظر دشواری در نظریه کلاسیک، درجه دشواری کمتر از ۰/۳۰ یا بالاتر از ۰/۷۰ بوده است، بدین دلیل که در این دامنه واریانس سؤال به حداقل می‌رسد. ملاک نامناسب بودن سؤال از نظر قدرت تشخیص در نظریه کلاسیک، ضریب تشخیص کمتر از ۰/۲۰ و در IRT کمتر از ۰/۵ بوده است.

جدول ۵. توزیع فراوانی و درصد سؤالات طبق CTT و IRT از نظر دشواری و تشخیص در درس روان‌شناسی

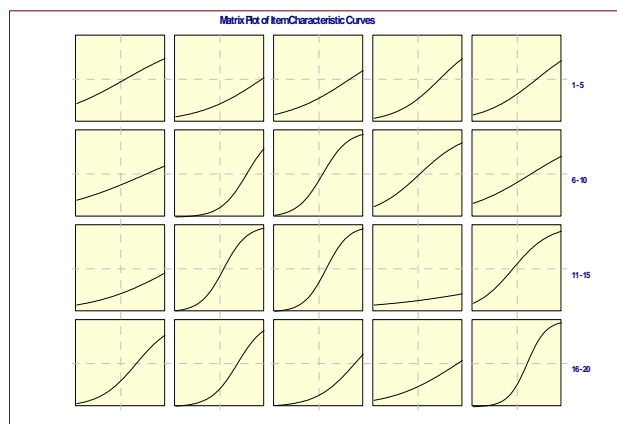
IRT									
ضریب تشخیص				ضریب دشواری				مناسب	نا مناسب
نامناسب		مناسب		عدم برازش با مدل		برازش با مدل			
تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	
۶۵	۱۳	۳۰	۶	۳۵	۷	۵۵	۱۱	۱۱	
۵	۱	۰	۰	۰	۰	۱۰	۲	۲	

جدول ۵ بیانگر آن است که طبق نظریه کلاسیک و IRT، از نظر دشواری ۱۱ سؤال و از نظر قدرت تشخیص ۶ سؤال در هر دو مدل مناسب می‌باشند. این مطلب حاکی از آن است که این دو نظریه در ۵۵ درصد موارد با هم توافق و در ۴۵ درصد موارد در تشخیص

سؤالات مناسب و نامناسب با هم تناقض دارند. ولی این دو نظریه در ضریب تشخیص در ۳۵ درصد موارد با هم توافق و در ۶۵ درصد موارد در تشخیص سؤالات مناسب و نامناسب با هم تناقض دارند.

جهت بررسی برازش مدل- داده‌ها (مقایسه مدل IPL با 2PL و نیز مقایسه مدل 2PL با 3PL) از آزمون خلی دو استفاده شده است.<sup>۱</sup> خلی دو محاسبه شده برای مقایسه مدل IPL با 2PL برابر ۵۳۸/۰۹ بود که با توجه به درجات آزادی ۲۰ در سطح اطمینان ۹۹ درصد ( $p < ۰/۰۱$ ) معنی دار است، یعنی مدل دو پارامتری برازش بهتری با مجموعه داده‌های آزمون دارد. همچنین خلی دو محاسبه شده برای مقایسه مدل 2PL با 3PL برابر ۲۶/۵۳ است که با درجات آزادی ۲۰ در سطح اطمینان ۹۵ درصد ( $p > ۰/۰۵$ ) معنی دار نیست، لذا مدل 2PL نسبت به مدل 3PL برازش بهتری با داده‌ها دارد.

نمودار ۱ منحنی‌های ویژگی تمامی سؤالات آزمون روان‌شناسی در مدل 2PL را نشان می‌دهد. توجه کنید که سؤالات از سمت چپ به راست شماره گذاری شده‌اند.

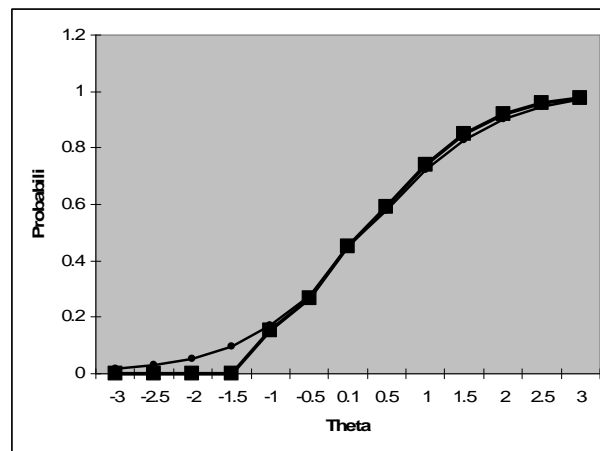


نمودار ۱. منحنی‌های ویژگی تمامی سؤالات آزمون درس روان‌شناسی در مدل دو پارامتری

<sup>۱</sup> - خلی دو محاسبه شده بر اساس این فرمول برآورد می‌گردد:

$$\chi^2 = (-2 \log \text{likelihood}_{1PL}) - (-2 \log \text{likelihood}_{2PL})$$

لازم به ذکر است که مقادیر پارامترهای سؤالات، به منظور بررسی ثبات برآورد پارامترها از طریق نمونه ارزشیاب برآورد گردید. نتایج نشان داد که در مدل دو پارامتری صرفاً ۵ سؤال در هر دو برآورد از طریق نمونه ضریب یاب و نمونه ارزشیاب برآزش نداشته‌اند. نمودار ۲ نمونه‌ای از برآزش پارامترهای سؤال را در دو برآورد مذکور نشان می‌دهد.



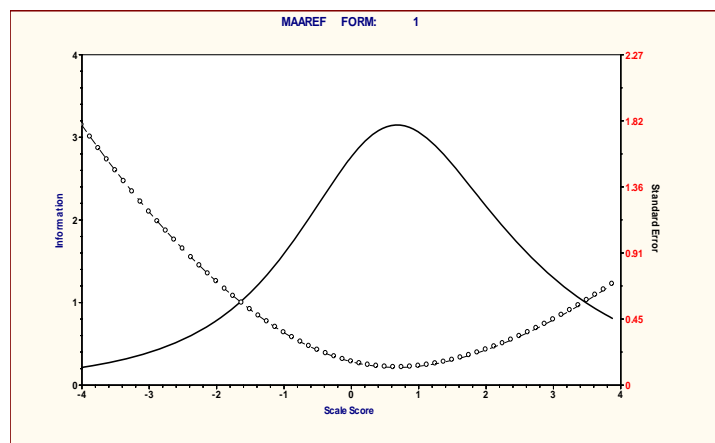
نمودار ۲. مقایسه منحنی ویژگی سؤال ۱۲ آزمون روان‌شناسی براساس دو نمونه ضریب یاب و ارزشیاب

نمودار فوق مبین این مسئله است که در سطح توانایی تتای ۳- تا ۳، منحنی‌های مورد انتظار و مشاهده شده برآزش خوبی با یکدیگر دارند. جدول ۶ بیشینه آگاهی سؤالات و همچنین نقطه‌ای که بیشینه آگاهی در آن به وجود می‌آید را نشان می‌دهد.

جدول ۶. بیشینه آگاهی سؤال و نقاط بیشینه آگاهی در مدل دو پارامتری درس روان‌شناسی

سؤال	بیشینه آگاهی سؤال	نقطه بیشینه آگاهی	سؤال	بیشینه آگاهی سؤال	نقطه بیشینه آگاهی
۱	۰/۰۳۷	۰/۲۹۹	۱۱	۰/۳۳	۳/۵۳۱
۲	۰/۰۵۱	۲/۸۵۱	۱۲	۰/۴۰۳	۰/۲۵۲
۳	۰/۰۵۰	۲/۰۸۸	۱۳	۰/۴۳۲	۰/۴۸۶
۴	۰/۱۱	۱/۴۵۳	۱۴	۰/۰۰۸	۱۰/۱۱۳
۵	۰/۰۷۴	۱/۳۱۲	۱۵	۰/۱۶۱	-۰/۲۹۸
۶	۰/۰۲۲	۱/۷۲۴	۱۶	۰/۱۶۰	۱/۰۴۵
۷	۰/۳۲۲	۱/۸۱۵	۱۷	۰/۳۰۹	۱/۱۸۶
۸	۰/۳۳۲	۰/۲۴۵	۱۸	۰/۳۴۶	۲/۴۲۶
۹	۰/۱۰۰	۰/۱۰۷	۱۹	۰/۰۴۸	۲/۶۷۱
۱۰	۰/۰۴۴	۰/۹۲۴	۲۰	۰/۶۰۷	۰/۶۷۱

نمودار ۳، منحنی آگاهی و منحنی خطای استاندارد آزمون روان‌شناسی را نشان می‌دهد. توجه داشته باشید که خطوط نقطه چین بیانگر منحنی خطای استاندارد آزمون است.



نمودار ۳. منحنی آگاهی و منحنی خطای استاندارد آزمون درس روان‌شناسی

مشاهده و بررسی منحنی تابع آگاهی آزمون روان‌شناسی نشان داد که بیشترین میزان آگاهی‌دهندگی این آزمون در فاصله تئای ۰/۱ تا ۱/۱ قرار گرفته است. به عبارت دیگر این آزمون برای افرادی که دارای این سطح توانایی هستند مناسب و کاربرد بهتری می‌تواند



داشته باشد. با توجه به این مطلب بیشترین میزان آگاهی‌دهندگی این آزمون در دامنه نمره ۶/۹ تا ۱۰/۲ است.

بیشترین میزان آگاهی‌دهندگی این آزمون ۳/۱۴۵ است که در نقطه تتا برابر ۰/۶۲۵ واقع شده است. با توجه به اینکه میانگین و انحراف استاندارد نمرات این آزمون به ترتیب ۶/۵۶ و ۳/۲۸ می باشد، نمره ملاک ۸/۶ معادل نقطه تتا برابر ۰/۶۲۵ خواهد بود. با توجه به نمره ملاک ۸/۶ در سطح ۹۵ درصد اطمینان می‌توان نتیجه‌گیری کرد که قضاوت کردن راجع به نمرات افراد در دامنه ۷/۶ تا ۹/۶ با بیشترین دقت همراه است. با توجه به این شرایط اگر نمره ملاک ۸/۶ نسبت به بقیه نمره‌های ملاک انتخاب شود، خطای کمتری در تصمیم‌گیری راجع به توانایی آزمودنی‌ها جهت ادامه تحصیل به عنوان دانشجو خواهیم داشت. طبق نمره ملاک ۸/۶ در ۷۵/۱ درصد افراد در این آزمون مردودند در صورتی که طبق نمره ملاک ۱۰، ۸۱/۲ درصد آزمودنی‌ها در این آزمون رد شده‌اند. با توجه به اینکه نمره ملاک ۱۰ که به عنوان نمره قبولی توسط دانشگاه اعلام شده است، با نمره ملاک پیشنهادی ۸/۶ فاصله دارد، همچنان این سؤال باقی می‌ماند که آیا انتخاب نمره ملاک ۱۰ عادلانه است هم برای دانشجو و هم از این لحاظ که بتوانیم محتوای کتاب را در حد قابل قبولی مورد سنجش قرار دهیم یا خیر؟

#### ب) درس جامعه‌شناسی

نتایج تحلیل سؤالات درس جامعه‌شناسی به تفکیک دو نظریه کلاسیک و سؤال- پاسخ در جدول ۷ نشان داده شده است.



در تحلیل ویژگی‌های روان‌سنجی سؤالات آزمون جامعه‌شناسی طبق نظریه کلاسیک سؤالات ۲، ۶، ۱۲ و ۱۳ درجه دشواری کمتر از ۰/۳۰ یا بالاتر از ۰/۷۰ دارند و این مطلب بدان معنی است که این سؤالات خیلی آسان و یا خیلی دشوار بوده‌اند. همچنین از نظر ضریب تمیز، سؤالات دارای وضعیت نسبتاً مناسبی بودند. ضریب پایایی کل آزمون با استفاده از فرمول آلفای کرونباخ ۰/۷۵ بود. همچنین نتایج نشان داد که حذف سؤالات موجب افزایش یا کاهش چشمگیر در میزان ضریب پایایی نمی‌شود. یعنی سؤالات از همسانی درونی برخوردار هستند. نحوه پاسخگویی آزمودنی‌ها به گزینه‌های انحرافی نیز نشان می‌دهد که در تعداد کمی از سؤالات آزمون نحوه توزیع پاسخ‌ها به گزینه‌های انحرافی نامناسب می‌باشد و گزینه‌های انحرافی از این حیث، عملکرد معیوبی داشته‌اند.

ویژگی‌های روان‌سنجی سؤالات آزمون جامعه‌شناسی طبق نظریه IRT نیز بررسی شده و نتایج نشان داد که این آزمون علاوه بر برقراری مفروضه‌های تک بعدی بودن و استقلال موضعی، با مدل 2PL برازش بهتری با داده‌ها دارد. بررسی منحنی ویژگی سؤالات حاکی از آن است که نیمی از سؤالات آزمون، قدرت تشخیص نسبتاً پایین و نیمی دیگر قدرت تشخیص متوسطی دارند و در تفکیک آزمودنی‌های قوی و ضعیف چندان موفق عمل نمی‌کنند. اکثر سؤالات این آزمون دشواری متوسطی داشتند، ۲ سؤال دشواری بالا و ۷ سؤال خیلی آسان بودند. جدول ۸ نشان می‌دهد که کدامیک از سؤالات براساس نظریه کلاسیک و IRT از نظر دشواری و تشخیص مناسب و یا نامناسب هستند.

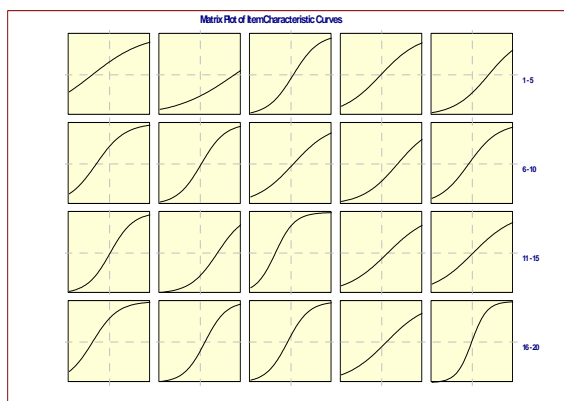
جدول ۸. سؤالات مناسب و نامناسب بر طبق CTT و IRT از نظر دشواری و تشخیص در درس جامعه‌شناسی

		IRT							
		ضریب دشواری				ضریب تشخیص			
		برازش با مدل		عدم برازش با مدل		مناسب		نامناسب	
		تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد
CTT	مناسب	۱۶	۸۰	۰	۰	۱۱	۵۵	۹	۴۵
	نامناسب	۴	۲۰	۰	۰	۰	۰	۰	۰

با توجه به نظریه کلاسیک و IRT، از نظر دشواری ۱۶ سؤال در هر دو مدل مناسب می‌باشند. این مطلب حاکی از این است که این دو نظریه در ۸۰ درصد موارد با هم توافق و در ۲۰ درصد موارد در تشخیص سؤالات مناسب و نامناسب با هم تناقض دارند. در عین حال از نظر قدرت تشخیص نیز ۱۱ سؤال در هر دو مدل مناسب می‌باشند. این مطلب حاکی از این است که این دو نظریه در ۵۵ درصد موارد با هم توافق و در ۴۵ درصد موارد در تشخیص سؤالات مناسب و نامناسب با هم تناقض دارند.

جهت بررسی برازش مدل - داده‌ها (مقایسه مدل 1PL با 2PL و نیز مقایسه مدل 2PL با 3PL) از آزمون خبی دو استفاده شده است. خبی دو محاسبه شده برای مقایسه مدل 1PL با 2PL برابر ۱۷۱/۳۶۰ بود که با توجه به درجات آزادی ۲۰ در سطح اطمینان ۹۹ درصد ( $p < 0/01$ ) معنی‌دار است، یعنی مدل دو پارامتری برازش بهتری با مجموعه داده‌های آزمون دارد. همچنین خبی دو محاسبه شده برای مقایسه مدل 2PL با 3PL برابر ۲۸/۸۱۱ است که با درجات آزادی ۲۰ در سطح اطمینان ۹۵ درصد ( $p > 0/05$ ) معنی‌دار نیست، لذا مدل 2PL نسبت به مدل 3PL برازش بهتری با داده‌ها دارد.

در نمودار ۴ منحنی‌های ویژگی تمامی سؤالات آزمون جامعه‌شناسی در مدل دو پارامتری آمده است.



نمودار ۴. منحنی‌های ویژگی تمامی سؤالات آزمون جامعه‌شناسی در مدل دو پارامتری

مقادیر پارامترهای سؤالات، به منظور بررسی ثبات برآورد پارامترها از طریق نمونه ارزشیاب نیز برآورد گردید. نتایج نشان داد که در مدل دو پارامتری ۲ سؤال وجود داشته که در برآورد از طریق نمونه ضریب یاب با مدل برازش نداشته‌اند، در حالی که در برآورد از طریق نمونه ارزشیاب با مدل مذکور برازش داشته‌اند.

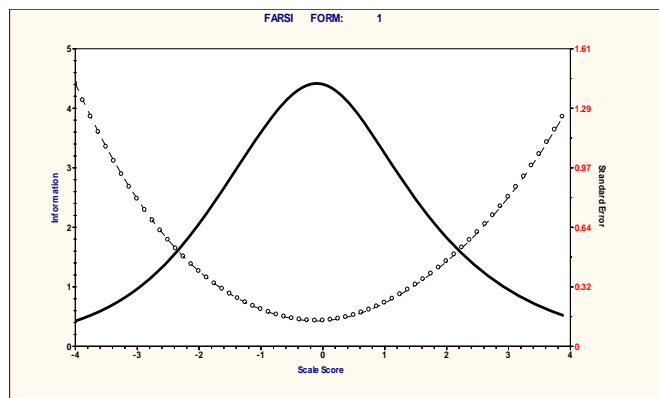
بیشینه آگاهی سؤالات و همچنین نقطه‌ای که بیشینه آگاهی در آن به وجود می‌آید، در جدول ۹ آورده شده است.

جدول ۹. بیشینه آگاهی سؤال و نقاط بیشینه آگاهی در مدل دو پارامتری درس جامعه‌شناسی

سؤال	بیشینه آگاهی سؤال	نقطه بیشینه آگاهی	سؤال	بیشینه آگاهی سؤال	نقطه بیشینه آگاهی
۱	۰/۰۶۷	-۱/۲۹۹	۱۱	۰/۳۳۹	۰/۰۳۵
۲	۰/۰۵۱	۲/۵۱۷	۱۲	۰/۲۴۲	۱/۲۶۷
۳	۰/۲۸۱	۰/۱۶۱	۱۳	۰/۴۸۵	-۱/۱۶۲
۴	۰/۱۲۱	-۰/۰۶۹	۱۴	۰/۱۰۴	۰/۴۳۸
۵	۰/۱۵۳	۱/۱۹۴	۱۵	۰/۱۰۸	۰/۰۴۸
۶	۰/۲۲۱	-۰/۹۹۶	۱۶	۰/۲۸۲	-۱/۲۷۳
۷	۰/۳۱۱	۰/۰۶۹	۱۷	۰/۴۰۴	۰/۳۴۲
۸	۰/۱۲۶	۰/۱۷۰	۱۸	۰/۳۹۳	-۰/۲۵۷
۹	۰/۱۴۶	۰/۱۷۳	۱۹	۰/۱۱۱	۰/۳۴۷
۱۰	۰/۲۰۷	-۰/۲۷۹	۲۰	۰/۷۹۰	-۰/۰۵۲

منحنی آگاهی و منحنی خطای استاندارد آزمون جامعه‌شناسی در نمودار ۵ ترسیم

شده است.



نمودار ۵. منحنی آگاهی و منحنی خطای استاندارد آزمون درس جامعه‌شناسی

مشاهده و بررسی منحنی تابع آگاهی آزمون جامعه‌شناسی نشان داد که بیشترین میزان آگاهی‌دهندگی این آزمون در فاصله تنای ۰/۶- تا ۰/۴ قرار گرفته است. به عبارت دیگر این آزمون برای افرادی که دارای این سطح توانایی هستند مناسب و کاربرد بهتری می‌تواند داشته باشد. با توجه به این مطلب بیشترین میزان آگاهی‌دهندگی این آزمون در دامنه ۷/۳ تا ۹/۳ است.

بیشترین میزان آگاهی‌دهندگی این آزمون ۴/۴۱۳ است که در نقطه تنای برابر ۰/۱۲۵- واقع شده است. با توجه به اینکه میانگین و انحراف استاندارد نمرات این آزمون به ترتیب ۹/۸ و ۴/۱ می‌باشد، نمره ملاک ۹/۳ معادل نقطه تنای برابر ۰/۱۲۵- خواهد بود. با توجه به نمره ملاک ۹/۳، در سطح ۹۵ درصد اطمینان می‌توان نتیجه‌گیری کرد که قضاوت کردن راجع به نمرات افراد در دامنه ۸/۵ تا ۱۰/۱ با بیشترین دقت همراه است. با توجه به این شرایط اگر نمره ملاک ۹/۳ نسبت به بقیه نمره‌های ملاک انتخاب شود، خطای کمتری در تصمیم‌گیری راجع به توانایی آزمودنی‌ها جهت ادامه تحصیل به عنوان دانشجو خواهیم داشت. طبق نمرات ملاک ۱۰ و ۹/۳، ۵۱/۷ درصد افراد در این آزمون مردودند. با توجه به اینکه نمره ملاک ۱۰ که به عنوان نمره قبولی توسط دانشگاه اعلام شده است، نزدیک به نمره ملاک پیشنهادی ۱۰/۸ است، همچنان این سؤال باقی می‌ماند که آیا انتخاب نمره

ملاک ۱۰ هم برای دانشجو و هم از این لحاظ که بتوانیم محتوای کتاب را در حد قابل قبولی مورد سنجش قرار دهیم عادلانه است یا خیر.

ج) درس فلسفه

جدول ۱۰ نتایج تحلیل سؤالات درس فلسفه را به تفکیک دو نظریه کلاسیک و سؤال- پاسخ نشان می‌دهد.

جدول ۱۰. تحلیل سؤالات درس فلسفه بر اساس نظریه کلاسیک و نظریه سؤال- پاسخ

سؤال	تحلیل پاسخ‌ها بر اساس نظریه کلاسیک										تحلیل پاسخ‌ها بر اساس نظریه سؤال- پاسخ									
	الف	ب	ج	د	پاسخ صحیح	پاسخ نادرست	پاسخ صحیح	پاسخ نادرست	پاسخ صحیح	پاسخ نادرست	پاسخ صحیح	پاسخ نادرست	پاسخ صحیح	پاسخ نادرست	پاسخ صحیح	پاسخ نادرست				
۱	۳۵۴	*۴۷۲	۵۷	۶۵	۵۲	۰/۵۲	۰/۴۴	۰/۲۵	۰/۵۷	۰/۶۲۷	۰/۸۵۸	۰/۱۹۷	۴۰/۰	*۰/۰۰						
۲	*۲۲۳	۱۱۰	۳۸۴	۱۸۹	۹۴	۰/۲۵	۰/۲۹	۰/۱۹	۰/۵۹	۲/۳۵۵	۰/۶۴۶	۰/۱۳۶	۷/۸	۰/۴۶						
۳	*۲۵۶	۱۰۷	۱۱۴	۲۳۵	۲۸۸	۰/۳۶	۰/۳۳	۰/۲۳	۰/۵۸	۲/۱۴۶	۰/۵۷۵	۰/۱۳۴	۲۱/۰	*۰/۰۰						
۴	*۳۵۳	۴۱۸	۵۷	۵۰	۱۲۲	۰/۴۰	۰/۴۲	۰/۲۴	۰/۵۷	۰/۵۲۸	۱/۲۸۰	۰/۱۵۳	۳۰/۸	*۰/۰۰						
۵	۹۷	۱۷۶	*۲۰۵	۱۶۶	۳۵۶	۰/۳۲	۰/۲۹	۰/۲۲	۰/۵۸	۲/۲۴۰	۰/۶۷۸	۰/۱۱۲	۱۳/۱	۰/۰۷						
۶	*۳۰۷	۱۹۷	۱۸۶	۱۲۵	۱۸۵	۰/۳۸	۰/۴۱	۰/۲۴	۰/۵۷	۱/۱۵۹	۰/۷۸۶	۰/۰۹۴	۳۰/۸	*۰/۰۰						
۷	۱۶۹	۳۹	۲۰۲	*۳۸۶	۲۰۴	۰/۴۸	۰/۳۲	۰/۲۵	۰/۵۸	۱/۴۱۷	۰/۶۳۴	۰/۲۱۰	۱۶/۵	۰/۰۳						
۸	۱۱۹	*۳۲۶	۸۶	۷۱	۳۹۸	۰/۵۴	۰/۳۶	۰/۲۵	۰/۵۷	۱/۱۹۷	۰/۷۱۱	۰/۱۱۱	۳۴/۲	*۰/۰۰						
۹	*۴۳۶	۸۳	۱۴۸	۱۵۶	۱۷۷	۰/۵۲	۰/۳۳	۰/۲۵	۰/۵۸	۰/۷۱۸	۰/۶۴۹	۰/۱۳۹	۵۴/۵	*۰/۰۰						
۱۰	۴۵	۱۲۷	*۴۵۷	۱۵۱	۲۲۰	۰/۵۹	۰/۲۱	۰/۲۴	۰/۶۰	۱/۵۶۹	۰/۳۵۴	۰/۲۳۱	۵۲/۹	*۰/۰۰						
۱۱	۳۱۵	۱۶۲	۱۷۳	*۱۹۱	۱۵۹	۰/۲۳	۰/۴۲	۰/۱۸	۰/۵۷	۱/۴۹۳	۰/۱۶۱	۰/۰۷۰	۱۰/۵	۰/۱۶						
۱۲	۹۵	۱۵۹	۱۹۰	*۴۶۶	۹۰	۰/۵۰	۰/۳۲	۰/۲۵	۰/۵۹	۱/۲۲۷	۰/۵۸۴	۰/۲۷۱	۱۲/۸	*۰/۰۰						
۱۳	۲۳۳	۱۲۸	*۱۷۰	۵۸	۴۱۱	۰/۳۰	۰/۱۳	۰/۲۱	۰/۶۰	۴/۴۹۷	۰/۴۶۲	۰/۱۴۱	۱۵/۸	۰/۰۴						
۱۴	۴۷۵	*۱۵۵	۱۳۳	۵۸	۱۷۹	۰/۱۷	۰/۱۰	۰/۱۴	۰/۶۰	۵/۱۸۵	۰/۴۱۶	۰/۱۳۲	۱۶/۳	۰/۰۴						
۱۵	۲۵۲	*۱۰۷	۱۴۶	۳۲۸	۱۶۷	۰/۱۴	۰/۳۰	۰/۱۲	۰/۵۹	۲/۵۳۷	۰/۷۱۱	۰/۰۶۴	۸/۹	۰/۲۶						
۱۶	۶۸	۱۵۰	*۲۳۶	۱۵۱	۳۹۵	۰/۳۸	۰/۳۶	۰/۲۴	۰/۵۸	۱/۵۷۸	۰/۷۴۵	۰/۱۳۳	۱۱/۰	۰/۲۰						
۱۷	۱۰۵	*۱۲۴	۱۶۷	۱۱۵	۴۸۹	۰/۲۲	۰/۱۱	۰/۱۷	۰/۶۰	۳/۵۶۷	۰/۸۱۰	۰/۱۱۸	۱۶/۶	۰/۰۳						
۱۸	۱۱۵	۳۲۸	۹۰	*۳۰۰	۲۵۷	۰/۴۰	۰/۴۳	۰/۲۴	۰/۵۷	۱/۲۱۲	۰/۹۵۴	۰/۱۲۲	۳۱/۳	*۰/۰۰						
۱۹	۲۵۵	۱۹۱	۲۱۰	*۷۱	۲۷۳	۰/۱۰	۰/۱۵	۰/۰۹	۰/۶۰	۳/۷۰۵	۰/۷۷۷	۰/۰۶۰	۲/۳	۰/۹۷						

	۰/۰۰*	۴۳/۳	۰/۱۵۰	۱/۷۳۱	۰/۷۴۹	۰/۵۷	۰/۲۵	۰/۳۶	۰/۵۴	۲۰۱	۱۳۱	۴۲۹*	۱۰۷	۱۳۲	۲۰
2 Log Likelihood		۰/۱۳۹	۰/۷۲۶	۲/۰۶۵	--	۰/۳۰	۰/۳۷					میانگین			
=		۰/۰۵۴	۰/۲۲۴	۱/۳۳۳	--	۰/۱۱	۰/۱۵					انحراف معیار			
	۲۱۵۱۳/۴۶														
	* گزینه صحیح (کلید سؤال)											* عدم برازش سؤال با مدل در سطح ۰/۰۱			

در تحلیل ویژگی‌های روان‌سنجی سؤالات آزمون فلسفه طبق نظریه کلاسیک، سؤالات ۱۴، ۱۵، ۱۷ و ۱۹ درجه دشواری کمتر از ۰/۳۰ یا بالاتر از ۰/۷۰ دارند و این مطلب بدان معنی است که این سؤالات خیلی آسان و یا خیلی دشوار بوده‌اند. همچنین سؤالات ۱۳، ۱۴، ۱۷ و ۱۹ دارای ضریب تمیز پایین می‌باشند. ضریب آلفای کرونباخ محاسبه شده نیز برابر ۰/۶۰ بود. همچنین نتایج نشان داد که حذف سؤالات موجب افزایش چشمگیر در میزان ضریب پایایی نمی‌شود، بنابراین سؤالات از این نظر مناسب می‌باشند. نحوه پاسخ‌گویی آزمودنی‌ها به گزینه‌های انحرافی نیز نشان می‌دهد که در نیمی از سؤالات آزمون نحوه توزیع پاسخ‌ها به گزینه‌های انحرافی نامناسب می‌باشد و گزینه‌های انحرافی از این حیث، عملکرد معیوبی داشته‌اند.

تحلیل ویژگی‌های روان‌سنجی سؤالات این آزمون طبق IRT نشان داد که آزمون فلسفه تک بعدی بوده، مفروضه استقلال موضعی نیز در آن برقرار است و در عین حال بررسی برازش مدل نشان داد که در مدل سه پارامتری ۱۰ سؤال، در مدل دو پارامتری ۵ سؤال و در مدل یک پارامتری ۱۰ سؤال با مدل‌های مربوطه برازش ندارند که در بین این سؤالات، سؤالات ۴، ۹، ۱۰، ۱۸ و ۲۰ با هیچکدام از مدل‌های مذکور برازش ندارند و باید حذف گردند. دلیل اینکه نیمی از سؤالات آزمون مذکور با این مدل برازش ندارند، احتمالاً به این دلیل است که به غیر از پارامترهای شیب، دشواری و عامل حدس، عوامل دیگری پاسخ‌های آزمودنی را تحت تأثیر قرار داده است. بررسی منحنی ویژگی سؤالات حاکی از آن است که کل سؤالات آزمون، قدرت تشخیص نسبتاً پایینی دارند و در تفکیک آزمودنی‌های قوی و ضعیف چندان موفق عمل نمی‌کنند. نیمی از سؤالات این آزمون دشواری متوسطی و نیمی دیگر دشواری بالایی داشتند. از نظر پارامتر حدس بجز ۶ سؤال، بقیه سؤالات آزمون، میزان حدس پذیری کمتر از ۰/۱۵ داشتند.



جدول ۱۱ توزیع فراوانی و درصد سؤالات مناسب و نامناسب از نظر دشواری و قدرت تشخیص با توجه به هر دو نظریه کلاسیک و IRT را نشان می‌دهد.

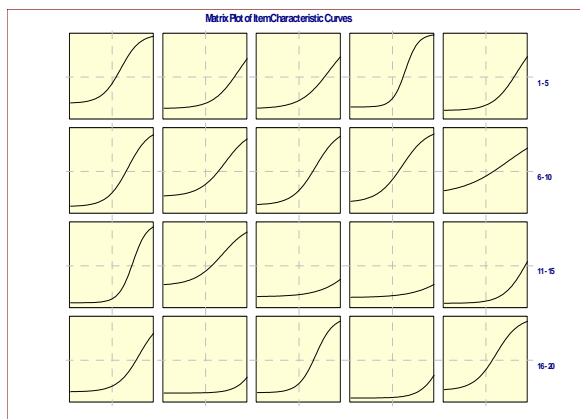
جدول ۱۱. سؤالات مناسب و نامناسب بر طبق CTT و IRT از نظر دشواری و تشخیص در درس فلسفه

IRT								CTT
ضریب دشواری				ضریب تشخیص				
برازش با مدل		عدم برازش با مدل		مناسب		نامناسب		
تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	
۶	۳۰	۱۰	۵۰	۱۴	۷۰	۲	۱۰	مناسب
۴	۲۰	۰	۰	۲	۱۰	۲	۱۰	نامناسب

از نظر دشواری در ۳۰ درصد موارد و از نظر قدرت تشخیص در ۸۰ درصد موارد، بین نظریه کلاسیک و IRT توافق وجود دارد.

جهت بررسی برازش مدل - داده‌ها (مقایسه مدل 1PL با 2PL و نیز مقایسه مدل 2PL با 3PL) از آزمون خی دو استفاده شده است. خی دو محاسبه شده برای مقایسه مدل 1PL با 2PL برابر ۱۵۸/۸۲۶ بود که با توجه به درجات آزادی ۲۰ در سطح اطمینان ۹۹ درصد ( $p < ۰/۰۱$ ) معنی دار است، یعنی مدل دو پارامتری برازش بهتری با مجموعه داده‌های آزمون دارد. همچنین خی دو محاسبه شده برای مقایسه مدل 2PL با 3PL برابر ۴۵/۱۲۹ است که با درجات آزادی ۲۰ در سطح اطمینان ۹۹ درصد ( $p < ۰/۰۱$ ) معنی دار است، لذا مدل 3PL نسبت به مدل 2PL برازش بهتری با داده‌ها دارد.

نمودار ۶ منحنی‌های ویژگی تمامی سؤالات آزمون فلسفه را در مدل سه پارامتری نشان می‌دهد.



نمودار ۶. منحنی‌های ویژگی تمامی سؤالات آزمون درس فلسفه در مدل سه پارامتری

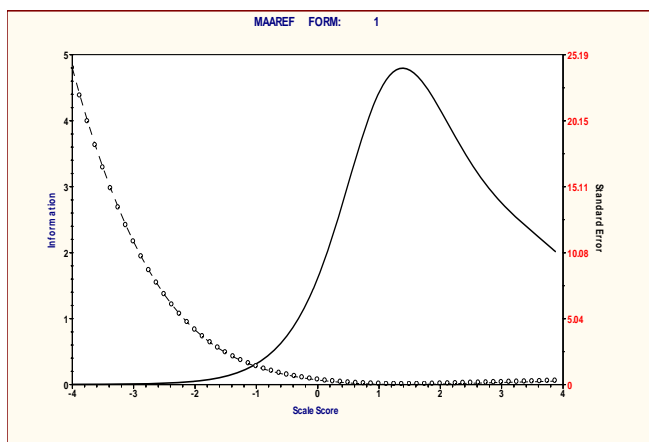
لازم به ذکر است که مقادیر پارامترهای سؤالات، به منظور بررسی ثبات برآورد پارامترها از طریق نمونه ارزشیاب برآورد گردید. نتایج نشان داد که در مدل سه پارامتری صرفاً ۷ سؤال در هر دو برآورد از طریق نمونه ضریب یاب و نمونه ارزشیاب برآزش نداشته‌اند.

بیشینه آگاهی سؤالات و همچنین نقطه‌ای که بیشینه آگاهی در آن به وجود می‌آید نیز محاسبه شده است که در جدول ۱۲ مشاهده می‌شود.

جدول ۱۲. بیشینه آگاهی سؤال و نقاط بیشینه آگاهی در مدل سه پارامتری درس فلسفه

سؤال	بیشینه آگاهی سؤال	نقطه بیشینه آگاهی	سؤال	بیشینه آگاهی سؤال	نقطه بیشینه آگاهی
۱	۰/۳۶۴	۰/۸۰۸	۱۱	۰/۸۴۸	۱/۵۵۲
۲	۰/۲۳۲	۲/۵۳۷	۱۲	۰/۱۴۶	۱/۵۶۰
۳	۰/۱۸۴	۲/۳۴۹	۱۳	۰/۱۱۷	۴/۷۶۰
۴	۰/۸۸۲	۱/۰۲۹	۱۴	۰/۰۹۷	۵/۴۶۱
۵	۰/۲۶۷	۲/۳۹۰	۱۵	۰/۳۲۲	۳/۰۲۷
۶	۰/۳۷۲	۱/۲۷۱	۱۶	۰/۳۱۰	۲/۱۳۴
۷	۰/۱۹۴	۱/۶۷۳	۱۷	۰/۳۷۷	۴/۰۹۷
۸	۰/۲۹۴	۱/۳۳۹	۱۸	۰/۵۱۹	۱/۳۲۵
۹	۰/۲۳۳	۰/۹۰۳	۱۹	۰/۳۸۸	۳/۷۸۲
۱۰	۰/۰۵۸	۲/۰۶۰	۲۰	۰/۲۸۹	۰/۹۲۳

منحنی آگاهی و منحنی خطای استاندارد آزمون فلسفه نیز در نمودار ۷ آمده است.



نمودار ۷. منحنی آگاهی و منحنی خطای استاندارد آزمون درس فلسفه

مشاهده و بررسی منحنی تابع آگاهی آزمون فلسفه نشان داد که بیشترین میزان آگاهی‌دهندگی این آزمون در فاصله تئای ۰/۸ تا ۱/۹ قرار گرفته است. به عبارت دیگر این آزمون برای افرادی که دارای این سطح توانایی هستند مناسب و کاربرد بهتری می‌تواند داشته باشد. با توجه به این مطلب بیشترین میزان آگاهی‌دهندگی این آزمون در دامنه ۸/۱ تا ۱۱/۴ است.

بیشترین میزان آگاهی‌دهندگی این آزمون ۴/۷۹۳ است که در نقطه تئای برابر ۱/۳۷۵ واقع شده است. با توجه به اینکه میانگین و انحراف استاندارد نمرات این آزمون به ترتیب ۵/۶۷ و ۳/۰۲ می‌باشد، نمره ملاک ۹/۸ معادل نقطه تئای برابر ۱/۳۷۵ خواهد بود. با توجه به نمره ملاک ۹/۳، در سطح ۹۵ درصد اطمینان می‌توان نتیجه‌گیری کرد که قضاوت کردن راجع به نمرات افراد در دامنه ۹/۱ تا ۱۰/۵ با بیشترین دقت همراه است. با توجه به این شرایط اگر نمره ملاک ۹/۸ نسبت به بقیه نمره‌های ملاک انتخاب شود، خطای کمتری در تصمیم‌گیری راجع به توانایی آزمودنی‌ها جهت ادامه تحصیل به عنوان دانشجو خواهیم داشت. طبق نمرات ملاک ۱۰ و ۹/۸، ۸۸/۳ درصد افراد در این آزمون مردودند. با توجه به اینکه نمره ملاک ۱۰ که به عنوان نمره قبولی توسط دانشگاه اعلام شده است، نزدیک به

نمره ملاک پیشنهادی ۱۰/۸ است، کماکان این سؤال باقی است که آیا انتخاب نمره ملاک ۱۰ هم برای دانشجو و هم از این لحاظ که بتوانیم محتوای کتاب را در حد قابل قبولی مورد سنجش قرار دهیم، عادلانه است یا خیر.

### نتیجه‌گیری نهایی

بر طبق تحلیل‌های نظریه کلاسیک آزمون، میانگین ضریب پایایی کل آزمون‌ها (روانشناسی، جامعه‌شناسی و فلسفه) با استفاده از فرمول آلفای کرونباخ ۰/۶۶ بدست آمد و این مطلب حاکی از آن است که آزمون‌های مذکور از این نظر در سطح نسبتاً مطلوبی قرار دارند. تحلیل گزینه‌های انحرافی نشان داد که گزینه‌های انحرافی تمامی سؤالات در همه آزمون‌ها هم احتمال نیستند و میزان جذابیت این گزینه‌ها برای آزمودنی‌ها یکسان نیست و گزینه‌های انحرافی از این حیث، عملکرد معیوبی داشته‌اند.

تحلیل ویژگی‌های روان‌سنجی کل سؤالات آزمون‌ها طبق IRT نشان داد که آزمون‌ها تک بعدی بوده و استقلال موضعی در آنها برقرار است. در بررسی پارامتر دشواری سؤالات مشخص گردید که سؤالات دارای پارامتر دشواری منظمی نیستند و از ساده به مشکل مرتب نشده‌اند. اگر قرار باشد که این آزمون‌ها بار دیگر اجرا شوند، بهتر است که به دلیل ایجاد انگیزه سؤال‌های آسان در ابتدای آزمون قرار گیرد (آلن و یسن، ترجمه دلاور، ۱۳۷۴).

با توجه به اینکه در آزمون‌های روان‌شناسی و جامعه‌شناسی مدل دو پارامتری نسبت به مدل سه پارامتری برآزش بهتری با این مجموعه از داده‌ها دارد، می‌توان دریافت که عامل حدس در پاسخ‌گویی آزمودنی‌ها به سؤالات نقش زیادی بازی نمی‌کند و این نکته از مزایای این آزمون‌ها محسوب می‌شود. همچنین در آزمون فلسفه وضعیت حدس‌پذیری سؤالات در حد معقولی می‌باشد.

مشاهده و بررسی منحنی تابع آگاهی آزمون‌ها نشان داد که آزمون‌های روان‌شناسی و جامعه‌شناسی برای افرادی که در محدوده توانایی پایین و آزمون فلسفه برای افرادی که در محدوده توانایی بالا هستند مناسب و کاربرد بهتری دارد.

با توجه به اینکه میانگین تمامی آزمون‌ها در پایین نمره ملاک قرار گرفته است، اما به نظر می‌رسد که مشکلات در درس فلسفه بیشتر بوده، هم به جهت اینکه میانگین نمرات آزمودنی‌ها در این آزمون پایین‌تر بوده و هم از این جهت که میانگین پارامتر دشواری سؤالات در این آزمون از بقیه آزمون‌ها بالاتر می‌باشد. البته پایین بودن نمرات ممکن است به دلیل برنامه نامناسب آموزشی، دشوار بودن سؤالات آزمون، توانایی پایین آزمودنی‌ها، عدم تلاش و یا انگیزه پایین آنها باشد. به هر حال این آزمون در شکست تحصیلی آزمودنی‌ها بیشتر تأثیر داشته است و آزمودنی‌ها در این درس ضعف بیشتری دارند. لذا به دانشگاه پیام نور پیشنهاد می‌شود چنانچه بخواهد آزمون‌هایش را اصلاح کند از آزمون فلسفه شروع کند و به منظور اصلاح، یا کتاب‌هایی را تألیف و معرفی کند که قابلیت خودآموز بودن آنها برای فراگیران بیشتر است یا اینکه برنامه آموزشی مناسب‌تری را برای این درس تعریف و اجرا نماید.

حال اگر بخواهیم نتیجه‌گیری کنیم که آزمون‌های مذکور در رشته روان‌شناسی در کل به چه صورت عمل کرده‌اند، باید به این نکته بسیار مهم توجه داشت که تحصیل و کسب علم در دانشگاه پیام نور که مبتنی بر نظام آموزش باز و از راه دور است، اساساً با تحصیل در دیگر دانشگاه‌ها و مؤسسات آموزش عالی که با رویکردهای مرسوم (سنتی) به ارائه آموزش می‌پردازند، تفاوت‌های عمده‌ای دارد. بر خلاف نظام‌های مرسوم دانشگاهی که بر محوریت استاد و تشکیل کلاس‌های چهره به چهره تأکید دارد، نظام آموزش باز و از راه دور، دانشجو را محور اصلی یادگیری می‌داند. بنابراین، دانشگاه در دوره‌های تحصیلی با رویکرد یادگیری مستقل عمل می‌کند. بنابراین یکی از مهم‌ترین دلایلی که باعث مردودی حدود ۸۵ درصد آزمودنی‌ها در آزمون رشته روان‌شناسی شده است، با توجه به این نکته که آزمون مذکور به صورت رقابتی نبوده است، شاید بدین دلیل بوده است که اکثر فراگیران با همان ذهنیت نظام مرسوم آموزشی و توقع حضور کامل در کلاس، داوطلب ادامه تحصیل در این نظام آموزشی می‌شوند و احتمالاً به دلیل عدم آشنایی با مهارت‌های مطالعه و مهارت‌های فراشناختی با حجم زیاد منابع آزمون روبرو و در نتیجه بی‌انگیزه شده و این امر موجب عدم موفقیت آنان گردیده است.

## منابع

- آلن، مری جی؛ و وندی ام. پین. (۱۹۸۲). *مقدمه‌ای بر نظریه‌های اندازه‌گیری*، ترجمه علی دلاور (۱۳۷۴). تهران: سمت.
- آناستازی ا. (۱۹۷۶). *روان‌آزمایی*. ترجمه محمد نقی براهنی (۱۳۷۱)، انتشارات دانشگاه تهران.
- بیکر، فرانک. (۲۰۰۱). *پایه‌های اساسی تئوری سؤال پاسخ*. ترجمه حیدر علی هومن، (۱۳۸۱). تهران: نشر پارسا.
- دلاور، علی. (۱۳۸۳). *روش تحقیق در روان‌شناسی و علوم تربیتی*، نشر ویرایش ستاری، بهزاد. (۱۳۷۷). کاربرد مدل سه پارامتری نظریه سؤال-پاسخ در تجزیه و تحلیل و ساخت آزمون‌های توانایی، (پایان‌نامه)، تهران: دانشگاه تربیت معلم.
- سهرابی، منیره. (۱۳۸۴). *بررسی تطبیقی سیر تحولات آموزش از راه دور در توسعه کمی و کیفی آموزش عالی به ویژه دانشگاه‌های مجازی در کشورهای انگلستان، آمریکا و ایران*، پایان‌نامه کارشناسی ارشد دانشگاه علامه طباطبائی.
- سیف، علی اکبر. (۱۳۸۱). *روش‌های اندازه‌گیری و ارزشیابی آموزشی (ویرایش دوم)*. تهران: نشر دوران.
- فراهانی، مهدی. (۱۳۷۵). *مقایسه مدل‌های اندازه‌گیری (کلاسیک و سؤال-پاسخ) از لحاظ برآورد پارامترهای سؤال و توانایی*، (پایان‌نامه)، تهران: دانشگاه علامه طباطبائی.
- کیامنش، علیرضا. (۱۳۷۸). *نظریه تعمیم‌پذیری در اندازه‌گیری آموزشی*. مجله علوم تربیتی دانشگاه تهران: دوره ۱۶، شماره ۱ و ۲، ص ۲۵-۴۸.
- مگنوسون، داوید. (۱۹۶۶). *مبانی نظری آزمون‌های روانی*. ترجمه محمد نقی براهنی (۱۳۷۰)، انتشارات دانشگاه تهران.
- مینایی، اصغر. (۱۳۷۷). *بررسی مقاومت مدل‌های منطقی IRT در برابر نقض مفروضات*، (پایان‌نامه)، تهران: دانشگاه علامه طباطبائی.
- هومن، حیدرعلی؛ عسگری، علی. (۱۳۸۴). *تحلیل عاملی: دشواری‌ها و تنگناهای آن*. مجله روان‌شناسی و علوم تربیتی دانشگاه تهران، دوره جدید، سال سی پنجم، شماره ۲، پاییز و زمستان.

- Anastasi, A. (1990). *Psychological Testing* (6<sup>th</sup> Ed.) New York: Mc Millan Publishing.
- Anderson. E.B. (1972). *The numerical solution of a set of conditional estimation equations*. Journal of the Royal Statistical Society, series B, 34, 42- 54.
- Embereston, Susan E.; Reise, Steven P. (2000). *Item Response Theory for psychologists*. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey. London.
- Hambleton, R.K.& Cook, L.L.(1977). *Latent trait models and their use in the analysis of educational test data*. Journal of Educational Measurement, 14, 75-96.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item Response Theory: Principle and application*. Boston, Ma: Kluwer Academic Publishers.
- Hambleton, R.K. (1989). *Principles and selected applications of item- response theory*. In R. Linn (Ed.) Educational measurement, (3<sup>rd</sup> Ed). New York: Macmillan. 147-200.
- Hableton, R. K., Rogers, H. J., & Swaminathan, H. (1991). *Fundamentals of item response theory*. Newbury Park, Cliff: Sage Publications.
- Hulin, C.L. & Drasgow, D.E. (1982). *Recovery of two and three parameter logistic item characteristics curves*. Applied Psychological Measurement, 6. 153- 160.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Lord & Novick (1974). *Statistical Theories of mental test scores*. Addison- Wesley Publishing company.
- McDonald, R.P. (1967). *Non- linear factor analysis* (Psychometric Monograph No. 15). Psychometric Society.
- Nunnally. J.C. & Bernstein (19۹۴). *Psychometric Theory*. Mc Graw- Hill book Co.
- Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B., Mislevy, R., L., S., & Thissen, D. (2000). *Computerized adaptive testing: A primer* (Second ed.). Hillsdale: NJ: Lawrence Erlbaum Associates.