

تاریخ وصول: ۹۰/۸/۳

تاریخ پذیرش: ۹۱/۲/۲۳

ارزش‌یابی آزمون نظری آزمون گواهینامه رانندگی بر اساس نظریه سؤال - پاسخ و مقایسه آن با نظریه کلاسیک آزمون

اسماعیل مام شریفی^۱

علی دلاور^۲

آزاده بلوکی^۳

سمیه شعبانی^۴

چکیده

زمینه: این پژوهش با هدف بررسی ویژگی‌های روان‌سنجی سؤالات آزمون نظری آزمون گواهینامه رانندگی انجام و اجرا شده است. نمونه مورد مطالعه را ۳۵۰ نفر از پاسخ‌دهندگان آزمون تشکیل دادند که به روش نمونه‌گیری خوشه‌ای چند مرحله‌ای انتخاب شده بودند و تعداد ۳۰ سؤال از سؤالات آزمون نظری آزمون گواهینامه رانندگی، که به روش تصادفی ساده استخراج شده بودند، به عنوان نمونه سؤالات مورد بررسی قرار گرفتند. برای بررسی ویژگی‌های روان‌سنجی سؤالات آزمون، از دو نظریه کلاسیک آزمون و سؤال - پاسخ استفاده گردیده و نتایج به دست آمده از آنها با هم مقایسه و بررسی شده است. روش پژوهش توصیفی بوده و در مرحله اول صحت نمونه‌گیری بررسی گردید. جهت تعیین تک بعدی بودن آزمون از روش آلفای کرانباخ و تحلیل عاملی استفاده شد. سپس تجزیه و تحلیل سؤالات آزمون با نظریه کلاسیک و نظریه سؤال - پاسخ انجام گرفت و پارامترهای سؤال (دشواری، تشخیص و حدس) و توانایی به دلیل نامعلوم بودن به روش برآورد همزمان مورد اندازه‌گیری قرار گرفت.

یافته‌ها: نتایج، بیان‌کننده تک بعدی بودن آزمون و استقلال موضعی آن است. پس از تعیین مفروضه‌های اصلی نظریه IRT، برازش مدل - داده‌ها مورد بررسی قرار گرفت که نتایج نشان داد مدل دو پارامتری برازش بهتری با مجموعه داده‌ها دارد. در مرحله بعد معناداری پارامترهای سؤال و توانایی توسط آزمون T وابسته مورد بررسی قرار گرفت. نتایج نشان داد که رابطه معناداری بین دقت برآورد پارامترهای دشواری، شیب و توانایی در دو نظریه کلاسیک و سؤال - پاسخ وجود ندارد. به منظور بررسی پایایی و ثبات نتایج آزمون در اجرای اول، بر روی نمونه‌ای به حجم ۳۰ نفر، بازآزمایی بعمل آمد. از آنجائیکه آزمون مورد نظر از نوع ملاک مرجع می‌باشد، برای بررسی توافق بین تصمیم‌ها از ضریب پایایی کاپا استفاده شد. نتایج نشان داد که رابطه معناداری بین اجرای اول و اجرای دوم وجود دارد و آزمون مورد نظر از پایایی و ثبات کافی در اجرای مختلف برخوردار است.

بحث و نتیجه‌گیری: تحلیل پارامترهای سؤال و آزمودنی‌ها، نشان‌دهنده سادگی آزمون و قدرت جداسازی خوب توانایی آزمودنی‌ها بوده و بر این اساس می‌توان نتیجه گرفت که سؤالات آزمون برای آزمودنی‌هایی با توانایی پایین از مناسبت و دقت بیشتری برخوردار است. میزان توانایی برآورد شده در نظریه سؤال - پاسخ در مقایسه با نمره آزمودنی در نظریه کلاسیک به میزان واقعی نزدیک‌تر است. با توجه به توانایی‌های برآورد شده می‌توان سؤالات متناسب با توانایی آزمودنی‌ها را انتخاب نمود و این می‌تواند به ایجاد بانک سؤال منجر شود.

واژگان کلیدی: ارزش‌یابی، آزمون، آزمون گواهینامه رانندگی، نظریه سؤال - پاسخ، نظریه کلاسیک آزمون.

1. Esmail.sharifi@gmail.com

۱. دانشجوی کارشناسی ارشد (نویسنده مسئول)

۲. استاد دانشگاه علامه طباطبائی

۳. دانشجوی کارشناسی ارشد

۴. دانشجوی کارشناسی ارشد

مقدمه

امروزه در بسیاری از جوامع از سه عامل انسان، راه و وسیله نقلیه به عنوان تعیین کننده های اصلی تصادفات نام برده می‌شود که در این میان عامل انسانی نقش اصلی را به خود اختصاص داده است. یعقوبی (۱۳۷۹) نیز در علت شناسی تصادف‌های رانندگی از ۴ عامل انسانی، جاده، وسیله نقلیه و محیط نام می‌برد که در ۹۰ تا ۹۵ درصد تصادفات رانندگی در ایران، عامل انسانی را دارای نقش اصلی و اول در حوادث رانندگی دانسته است.

سالیانه بیش از ۱/۲۶ میلیون نفر در دنیا به علت تصادفات رانندگی جان خود را از دست می‌دهند (پدین، اسکار فیلد و سلیت^۱، ۲۰۰۴)، و اغلب مرگ و میرهای ناشی از حوادث رانندگی، در کشورهای با درآمد کم و متوسط و برای جوانان اتفاق می‌افتد (کوپیتز و کروپر^۲، ۲۰۰۵). سازمان بهداشت جهانی پیش بینی کرده است که در سال ۲۰۲۰، مرگ و میر ناشی از حوادث رانندگی به ۲/۳۴ میلیون نفر خواهد رسید (رومانا^۳، ۲۰۰۴).

بر اساس آمارها سالانه حدود ۲۵ هزار نفر نیز در جاده های ایران کشته می‌شوند و سوانح رانندگی پس از بیماری‌های قلبی - عروقی، دومین عامل مرگ و میر در ایران محسوب می‌شود. ۱۴۱ هزار کشته، ۱۲/۵ میلیون مصدوم و معلول، از پیامد های ۷ سال تصادفات در کشور است و خسارات جانی و مالی ۷ سال تصادفات رانندگی در ایران با پیامد های یک جنگ یا چند زلزله مهیب برابری می‌کند. موثق‌ترین آمار تصادفات رانندگی در کشور از کشته شدن ۱۴۰ هزار و ۹۳۳ نفر طی ۷ سال گذشته (ابتدای ۷۷ تا انتهای ۸۳) حکایت دارد و این در حالی است که وقوع این حوادث بیش از ۱۲ میلیون مصدوم و ۵۰۰ هزار معلول مادام‌العمر بر جای گذارده است. مرگ ناشی از تصادفات یک ساله در کشور با تلفات انسانی ناشی از زلزله بم برابر بوده، اما چون این حوادث ناگهانی نیست چندان مشهود به نظر نمی‌رسد. با وجود اینکه ایران کمتر از یک صدم جمعیت جهان

1. Peden, Scuffield, & Sleet
2. Kopits & Cropper
3. Romana

را دارد بیش از یک چهارم (۳ درصد) از حوادث ترافیکی را به خود اختصاص داده است (خبرگزاری فارس، ۱۳۸۵).

بر اساس آمارهای سازمان یونسف ۲/۵ درصد از کل تصادفات جاده ای در ایران اتفاق می افتد و این به آن معنا است که آمار تصادفات جاده ای در ایران ۲۰ برابر دیگر کشورهای جهان است. در واقع، میزان کشته شدگان تصادفات جاده ای در دنیا، به ازای هر ده هزار خودرو ۳ نفر بوده، در حالی که در ایران به ازای هر ده هزار خودرو حدود ۳۳ نفر بوده است (سوری و همکاران، ۱۳۸۸). آلمان کمتر از یک چهارم ایران مساحت دارد و تعداد اتومبیل هایش بیش از ۴ برابر ایران است. با این همه شمار قربانیان حوادث رانندگی در ایران بیش از ۱۰ برابر آلمان است (بیمه ایران، ۲۰۰۸). همچنین انگلستان نیز با وجود ۶ برابر بودن تعداد وسایل نقلیه اش، ۳۲ بار کمتر از ایران تلفات جاده ای دارد (سلمانی و همکاران، ۱۳۸۷). در پیامدهای ناشی از حوادث، پدیده کوه یخ، یک فرایند واقعی است. قله کوه یخ که معمولاً مورد توجه قرار می گیرد و به ظاهر کوچک است شامل موارد مرگ ناشی از حوادث است، ولی بدنه عظیم این کوه از بستری شدگان، آسیب دیدگان نیازمند خدمات سرپایی و درمان های خانگی توأم با ناتوانی های گذرا یا ماندگار تشکیل شده است و عواقب روانی، اجتماعی و اقتصادی حوادث نیز بر این بزرگی می افزاید.

پایین بودن سطح فرهنگ صحیح ترافیک و عدم توجه کافی به قوانین و مقررات و حقوق سایرین و رعایت نشدن آیین نامه رانندگی، ناشی از ناکافی بودن آموزش نظری و عملی رانندگی، قبل و بعد از صدور گواهینامه، موجب شده است آمار تصادفات رانندگی در ایران شمار کشته ها و مجروحین ناشی از آن افزایش پیدا کرده و بالاترین رتبه را در بین ملل جهان کسب کند و این یعنی مواجه شدن با پیامدهای جبران ناپذیر از دست رفتن سرمایه انسانی.

اگرچه در یک برنامه پیشگیرانه جامع برای کاهش تصادفات و تلفات، باید به هر ۴ عامل انسانی، جاده، وسیله نقلیه و محیط پرداخت (آیتی، ۱۳۷۱، نقل از یعقوبی، ۱۳۷۹)، اما به نظر می رسد که مداخله در عامل انسانی (با توجه به نقش زیاد آن در تصادفات رانندگی در ایران)

اثربخش تر باشد. چرا که نتایج مطالعات انجام شده نیز نشان می‌دهد که می‌توان از ۹۸ درصد حوادث پیشگیری نمود (عراقی و واحدیان، ۱۳۸۴). مراد از عامل انسانی عابرین، سرنشین‌ها، رانندگان و عوامل روانی، رفتاری، فرهنگی و اجتماعی مؤثر بر آن‌ها می‌باشد.

امروزه نقش آموزش در افزایش کارآیی، بهره‌وری و توسعه اقتصادی کشورها بیش از پیش اهمیت یافته است، چرا که در دنیای کنونی داشتن کارخانه‌ها، زمین، ابزار و ماشین-آلات سرمایه محسوب نمی‌شود، بلکه ثروت در منابع انسانی نهفته است (خالدی، ۱۳۸۷). آموزش این توانایی را دارد تا برای بهبود و پیشرفت کشورهای کمتر توسعه یافته زمینه لازم را فراهم سازد و سیاست‌های لازم برای مقابله با این بحران جهانی را اتخاذ نماید (لوئیز، ۱۹۹۰، نقل از سهرابی، ۱۳۸۴). بسیاری از کشورهای در حال توسعه درباره مرگ و میر در کشور خود اطلاعات محدودی دارند (خی و وانگ^۱، ۲۰۰۴). با مطالعات اپیدمیولوژیک در علوم پزشکی، علوم اجتماعی و علوم رفتاری می‌توان در جهت کاهش حوادث و عواقب ناشی از آن‌ها گام برداشت، چرا که آسیب‌های ناشی از حوادث رانندگی یک مشکل بزرگ بهداشتی در دنیا به شمار می‌آید (پدن و توریان^۲، ۲۰۰۵).

قاعدتاً زمانی یک جامعه به کمال رسیده است که آموزش به گونه‌ای صحیح در آن انتقال یافته باشد و این نیز نمی‌تواند محقق شود مگر با یک برنامه ریزی فراگیر و مستمر به منظور تدوین و آموزش اصول صحیح رانندگی برای سنین مختلف و با روش‌های متنوع و مورد قبول (احمدی، ۱۳۸۵). به هنگام توجه به روند تکاملی کشورهای توسعه یافته، همواره شاهد ارتباط نزدیک بین آموزش، پژوهش، ارزش‌یابی و رشد فناوری از یک سو و پیشرفت اقتصادی و اجتماعی از سوی دیگر می‌باشیم.

یکی از این فعالیت‌ها در زمینه کاستن از رخدادهای مرتبط با رانندگی، توجه ویژه و مجدد به آزمون نظری گواهینامه رانندگی از زوایای مختلف و مقایسه آن با دیگر کشورها است که می‌تواند به عنوان اقدامی پیشگیرانه در جهت کاستن از رخداد حوادث رانندگی به

1. Chi & Wang
2. Peden & Toroyan

شمار آید. آزمون نظری، آزمون استاندارد شده ای است که برای تصمیم گیری در مورد رسیدن فرد به دانش نظری کافی و ساختن راننده ای ایمن و بی خطر از فرد طراحی شده است (ویبرگ^۱، ۲۰۰۴).

در حال حاضر آزمون نظری گواهینامه رانندگی در ایران شامل مجموعه‌ای از پرسشنامه‌های ۳۰ سؤالی است که تعدادی از سؤالات آن به صورت تصویری و رنگی می‌باشد. هر سؤال دارای ۴ گزینه است که یکی از گزینه‌ها صحیح و بقیه انحرافی می‌باشند. زمان پاسخگویی به آزمون محدود، و ۲۰ دقیقه می‌باشد. با توجه به اینکه آزمون از نوع ملاک مرجع و دو ارزشی است، ملاک قبولی در آن پاسخگویی صحیح به ۲۶ سؤال از ۳۰ سؤال (نمره برش ۸۶/۶۷) می‌باشد.

امروزه پس از گذشت چندین سال از برگزاری این آزمون و فعالیت مراکز آموزشی مربوطه در کشور، باید پرسید که:

- این آزمون تا چه اندازه توانسته است به نیازهای فرد و جامعه پاسخ دهند؟
 - آیا این آزمون به اهداف طرح شده اولیه خود رسیده است؟
 - مواد (سؤالات) این آزمون تا چه اندازه از معیارهای روان سنجی و استاندارد های مربوطه برخوردارند؟
- پاسخ به این سؤال‌ها و سؤال‌های مشابه مستلزم استفاده از سازوکار ارزشیابی آزمون از بعد روان سنجی است.

متولیان امر باید به موازات گسترش آموزش‌های ترافیکی، به نتایج آنچه تاکنون در این زمینه هزینه شده، بپردازند. چرا که با هزینه‌های بسیاری که برای این آموزش‌ها صرف می‌شود ایجاد اصلاحات در آن ضروری به نظر می‌رسد. اصلاحاتی که می‌توان از آن به عنوان «نوگرایی^۲» یاد کرد.

یکی از مسائل کلیدی، در دورنمای آزمون نظری گواهینامه رانندگی، بازسازی مداوم آن

از طریق فرایند ارزش‌یابی است. یک نظام آموزشی ایستا پاسخگوی ماهیت پویای توسعه نیست، چرا که نیازها و اولویت‌های یک دهه، مسلماً با نیازهای دهه بعد متفاوت است، گرچه دانش و مهارت‌های بنیادین معینی ممکن است همچنان با اهمیت باقی بمانند، با این حال مسائلی که برای نسل گذشته مهم بوده، امروزه کاملاً بی‌اهمیت است (حسینی، ۱۳۸۰). بر این اساس می‌توان با پاتون^۱ (۱۹۹۷) هم عقیده بود و این گفته را تأیید کرد که اگر نظام آموزشی به سازوکار ارزشیابی مجهز باشد، نتایج مطلوب حاصل خواهد شد (نقل از بازرگان، ۱۳۸۶). بنابراین، ارزشیابی آزمون نظری گواهینامه رانندگی از اهمیت حیاتی برخوردار است.

در این میان علم روان‌سنجی به عنوان شاخه‌ای از روان‌شناسی کاربردی، در پی یافتن شیوه‌هایی است که به کمک آن بتواند صفات روانی انسان را به صورت کمی و مقایسه پذیر ارائه دهد و در این راه اغلب از محاسبات آماری که در خیلی از موارد بسیار پیچیده و پیشرفته نیز می‌شود، سود می‌جوید (ثرندایک^۲، ۱۹۸۲، ترجمه هومن، ۱۳۷۵). به نظر می‌رسد که فعلاً در ایران، جز در موارد استثنایی، ساختن آزمون محدود به تهیه سؤال است و کمتر کسی به خصوصیات علمی آزمون‌ها و سؤالات توجه دارد. این عدم توجه به مبانی علمی آزمون‌سازی در سازمان‌های خارج از دانشگاه‌ها بیشتر نمایان است تا جایی که بعضی از این سازمان‌ها در ظرف چند روز آزمون به خصوصی را آماده اجرا می‌کنند که برای ساختن آن در کشوری مثل آمریکا، شاید یکی دو سال وقت لازم باشد (مگنوسون، ۱۹۶۶، ترجمه براهنی، ۱۳۷۰).

استفاده کنندگان آزمون در انتخاب، اجرا، نمره‌گذاری یا تفسیر یک آزمون باید هدف‌ها، راه‌های وصول به آن‌ها و نتایج احتمالی را بدانند. داشتن هدف‌های مطلوب کافی نیست؛ استفاده کنندگان باید روش‌های ضروری به حداکثر رساندن کارایی و به حداقل رساندن عوامل نامطلوب را بدانند. شایستگی در استفاده از آزمون، ترکیبی است از شناخت اصول روان‌سنجی، اطلاع از موقعیت مشکل‌زا که در آن آزمون انجام می‌شود و مهارت

1. Patton
2. Thorndike

فنی. اگر آزمون‌ها درست به کار برده شوند، ابزار مفیدی هستند؛ در غیر این صورت به ابزار خطرناکی تبدیل می‌شوند (آلن و ین^۱، ۱۹۷۹، ترجمه دلاور، ۱۳۸۴).

نظریه‌های آزمون در علم روان‌سنجی به دو مقوله تقسیم می‌شوند: اول: نظریه کلاسیک آزمون (CTT)، که قدمت آن به مفهوم اسپیرمن^۲ از نمره مشاهده شده، که شامل مؤلفه‌های حقیقی و خطا است می‌رسد. دوم: نظریه سؤال - پاسخ (IRT)، یا نظریه صفت مکنون که در حیطه و قلمرو آزمون نقش مهمی را ایفا می‌نماید. مدل‌های مشتق از IRT در جهت توسعه و بسط آزمون‌ها، همتراز سازی نمرات آزمون‌های غیر موازی، بررسی تورش سؤال و گزارش نمرات مورد استفاده قرار می‌گیرند (همبلتون و واندرلیندن^۳، ۱۹۸۲).

در واقع، یک آزمون می‌تواند از زوایای مختلف مورد مطالعه واقع شده و سؤالات آن نیز بر طبق نظریات متفاوت ارزش گذاری شود. مقایسه مدل‌های IRT از لحاظ دقت برآورد پارامترهای مدل، مورد توجه محققان بسیار طی سال‌های اخیر بوده است. یکی از تحقیقات قابل توجه در زمینه بررسی آزمون نظری آزمون گواهینامه رانندگی، پژوهشی است که توسط ماری ویرگ در سال ۲۰۰۴ و در دانشگاه یومی سوئد انجام گرفته است. آزمون نظری گواهینامه رانندگی در سوئد شامل ۶۵ سؤال چند گزینه ای (۲ تا ۶ گزینه) و از نوع ملاک مرجع با نمره برش ۵۲ (۸۰ درصد) است. نمونه مورد بررسی در این پژوهش شامل ۵۴۰۴ نفر آزمون شونده بوده که به ۶۵ سؤال مذکور پاسخ داده‌اند. برآورد پارامترهای سؤال و توانایی از طریق روش بیشینه درست نمایی حاشیه ای صورت گرفته است. هدف این پژوهش ارزش‌یابی آزمون نظری گواهینامه رانندگی با استفاده از نظریه سؤال - پاسخ و همچنین مقایسه مدل‌های سه گانه این نظریه با همدیگر در برآورد پارامترهای آزمون مورد بررسی، بوده است. نتایج پژوهش وی نشان می‌دهد که نظریه سؤال - پاسخ، ابعاد متفاوت و اطلاعات ارزشمندی به دست می‌دهد. با مقایسه مدل‌های

1. Allen & Yen
2. Spearman
3. Hambleton & Vanderlinden

یک، دو و سه پارامتری در برآورد پارامترها مشخص شده است که مدل سه پارامتری به برآورد دقیق‌تری در مقایسه با مدل‌های یک و دو پارامتری منجر شده است.

همبلتون و کوک^۱ (۱۹۸۳)، با مقایسه این مدل‌ها از طریق شبیه‌سازی کامپیوتری نشان دادند که، مدل منطقی سه پارامتری نسبت به مدل‌های یک و دو پارامتری با آزمون‌های ۲۰ سؤالی دارای قدرت بیشتری در برآورد جایگاه افراد در صفت مکنون و رتبه بندی آزمودنی‌ها براساس صفت مورد سنجش بوده است. رایید^۲ (۱۹۹۳) با هدف مقایسه مدل‌های IRT و با استفاده از تحلیل خرده آزمون‌های محاسبات عددی، تجسم فضایی سه بعدی، خزانه لغات و استدلال ریاضی، نتیجه گرفت که هر ۴ خرده آزمون از لحاظ سطح دشواری سؤالات و قدرت تشخیص بین افراد در سطوح مختلف توانایی تفاوت داشته‌اند. بنابراین مدل‌های یک و دو پارامتری از لحاظ برآورد پارامترها تفاوت معنی دار نشان داده‌اند. لیکن به دلیل تأثیر نداشتن عامل حدس در پاسخگویی به سؤالات، مدل‌های دو و سه پارامتری تفاوت معنی داری در برآورد پارامترها نداشته‌اند.

نتایج مقایسه مدل‌های مختلف از نظر برآورد پارامترهای سؤال و توانایی در پژوهش فراهانی (۱۳۷۵) نشان داد که، مدل کلاسیک و مدل‌های IRT برآوردهای متفاوتی برای پارامترهای سؤال بدست می‌دهند و برآورد پارامترهای سؤال بر پایه مدل‌های IRT به ویژه مدل سه پارامتری، دقیق‌تر از برآورد پارامترهای سؤال بر اساس مدل کلاسیک و مدل‌های ساده‌تر IRT است. همچنین مقایسه مدل‌ها از نقطه نظر برآورد توانایی آزمودنی‌ها نشان داد که مدل سه پارامتری برآورد متفاوت و دقیق‌تری از مدل کلاسیک و حتی مدل‌های یک و دو پارامتری IRT ارائه می‌دهد. در پژوهش محمدزاده (۱۳۷۵)، نیز مشخص شد که مشخصه‌های آماری سؤالات در مدل کلاسیک وابسته به نمونه بوده و برآورد توانایی آزمودنی هم، وابسته به سؤالات آزمون است، اما در مدل‌های جدید اندازه‌گیری این برآوردها مستقل از هم می‌باشند. مقایسه مدل‌ها از لحاظ میزان دخالت پارامتر حدس، اعتبار

1. Cook
2. Reid

اندازه‌گیری و آگاهی دهندگی، برتری مدل‌های جدید اندازه‌گیری را نشان داده است. صالحی (۱۳۷۶) با مقایسه دو نظریه CTT و IRT نشان داد که نظریه سؤال - پاسخ در مقایسه با نظریه کلاسیک توانمندتر و کارآمدتر بوده و با خطای استاندارد اندازه‌گیری کمتر، آگاهی دهندگی و دقت بیشتری پارامترهای سؤال و توانایی را برآورد نموده است. همچنین این نظریه در سطوح بالای توانایی دارای بیشترین میزان آگاهی بوده است. یونسی نیز (۱۳۸۵) با بررسی ویژگی‌های روان‌سنجی سؤالات آزمون‌های فراگیر رشته روان‌شناسی بر اساس نظریه سؤال - پاسخ، به این نتیجه رسید که مدل دو پارامتری نسبت به مدل سه پارامتری برآزش بهتری با مجموعه داده‌ها داشته است و این نشان‌دهنده کم‌اثر بودن عامل حدس در پاسخگویی به سؤالات، تلقی شده است.

امروزه با مدرن شدن آموزش، ارزش‌یابی به عنوان یکی از مؤلفه‌های آموزش مورد استفاده قرار گرفته و نوآوری‌هایی در زمینه آن پدید آمده است. یکی از این نوآوری‌ها مربوط به بکارگیری کامپیوتر در فرایند سنجش و ارزیابی و آزمون متقاضیان گواهینامه رانندگی در کشورهای اروپایی می‌باشد به طوریکه بسیاری از مراکز آموزشی به برگزاری آزمون‌های کامپیوتری اقدام می‌کنند. یکی از مهم‌ترین کاربردهای تجزیه و تحلیل روان‌سنجی سؤالات، ایجاد یک مجموعه (بانک) سؤال استاندارد از سؤالات موجود می‌باشد که این خود مقدمه‌ای است برای ایجاد و گسترش آزمون‌های انطباقی کامپیوتری.

در ایران، آزمون گواهینامه رانندگی شامل یک آزمون نظری و یک آزمون عملی است، که در این پژوهش بر آزمون نظری تأکید شده است. در آزمون نظری گواهینامه رانندگی از نسخه‌های ثابت و متعددی جهت آزمون استفاده می‌شود. اما در ایران، تاکنون تحقیقات منسجمی در ارتباط با ارزش‌یابی سؤالات این آزمون صورت نگرفته و به همین دلیل و نیز اهمیت این آزمون‌ها در تأثیرگذاری بر پیشگیری از رخداد رانندگی از طریق آموزش‌های قبل از صدور گواهینامه رانندگی، ضروری به نظر می‌رسد تحقیقاتی انجام گیرد تا میزان دستیابی به اهداف از پیش تعیین شده در زمینه مؤلفه‌های روان‌سنجی یک

آزمون استاندارد، مشخص گردد. بنابراین هدف از پژوهش حاضر ارزش‌یابی و بررسی ویژگی‌های سؤالات آزمون نظری گواهینامه رانندگی طبق نظریه سؤال - پاسخ (IRT) و مقایسه آن با نظریه کلاسیک آزمون (CTT)، در کنار بررسی برآزش مدل‌های IRT با مجموعه داده‌های آزمون بوده است.

روش

در پژوهش حاضر جامعه آماری مورد مطالعه، شامل کلیه سؤالات (۱۹۴ سؤال) آزمون نظری گواهینامه رانندگی در شهر تهران و در سال ۱۳۸۸ بوده است.

به منظور کسب اطلاعات دقیق در مورد جامعه، بررسی یکایک عناصر جامعه مطلوب به نظر می‌رسد. اما معمولاً این کار غیر ممکن یا غیر عملی است. با توجه به اینکه، نظریه‌های روان‌سنجی، نظریه‌های نمونه‌های بزرگ هستند. قانون کلی در این مورد بزرگ‌ترین اندازه ممکن را تصویب می‌کند. رایت و استون برای مدل یک پارامتری حداقل حجم نمونه آزمودنی‌ها را ۲۰۰ نفر ذکر کرده‌اند. بعضی از محققان هم برای مدل دو پارامتری ۵۰۰ نفر و برای مدل سه پارامتری ۱۰۰۰ نفر را لازم دانسته‌اند تا به برآورد‌های نامتغیر و ناریب از پارامترهای مدل منجر نشود (همبلتون، جونز و راجرز، ۱۹۹۳).

بنابراین با توجه به نتایج تحقیقات صورت گرفته، حجم نمونه سؤالات مورد بررسی، تعداد ۳۰ سؤال بوده که به روش کاملاً تصادفی انتخاب گردید. همچنین در این پژوهش به منظور انتخاب حجم نمونه آزمودنی‌ها، با استفاده از روش نمونه‌گیری خوشه‌ای چند مرحله‌ای، پاسخ‌های یک نمونه ۳۵۰ نفری به آزمون محقق ساخته انتخاب شده است. انتخاب حجم ۳۰ سؤالی به دلیل رعایت شکل صوری آزمون‌های برگزار شده توسط راهور ناجا بوده است.

روش پژوهش نیز با توجه به اهداف پژوهش و بر حسب نحوه گردآوری داده‌ها، توصیفی بوده است.

جهت تجزیه و تحلیل داده های جمع آوری شده، با استفاده از نرم افزارهای SPSS و BILOG - MG، از بررسی های مرتبط با سنجش مؤلفه های روان سنجی آزمون، از قبیل؛ ضریب دشواری، ضریب تمیز، واریانس سؤالات، ضریب همبستگی دو رشته ای نقطه ای استفاده گردید. تک بعدی بودن آزمون با محاسبه آلفای کرونباخ و انجام تحلیل عاملی بررسی و سپس استقلال موضعی آزمون بدست آمد. با استفاده از آزمون خی دو (χ^2) در مورد برازش مدل با داده ها قضاوت شد و پس از اینکه معلوم شد کدام مدل بهترین برازش را با داده های آزمون دارد، پارامترهای برآورد شده سؤال (دشواری، تشخیص و حدس) و توانایی آزمودنی ها با استفاده از روش های برآورد همزمان پارامترهای سؤال و توانایی استخراج گردید. سپس منحنی ویژگی تمامی سؤالات برای رسم نمودارها، تابع آگاهی سؤالات برای مقایسه تفاوت مدل ها، تابع آگاهی آزمون، منحنی خطای استاندارد آزمون و منحنی توانایی های آزمودنی ها ترسیم گردید. در ادامه، بیشینه آگاهی به همراه سطح توانایی سؤالات، بیشینه آگاهی آزمون و همچنین نقطه ای که بیشینه آگاهی در آنها به وجود می آید، محاسبه و نقطه ای که بیشترین میزان آگاهی دهندگی آزمون ها در آن دامنه قرار دارد و نیز توزیع بیشینه آگاهی و خطای استاندارد توانایی آزمون نیز آورده شد.

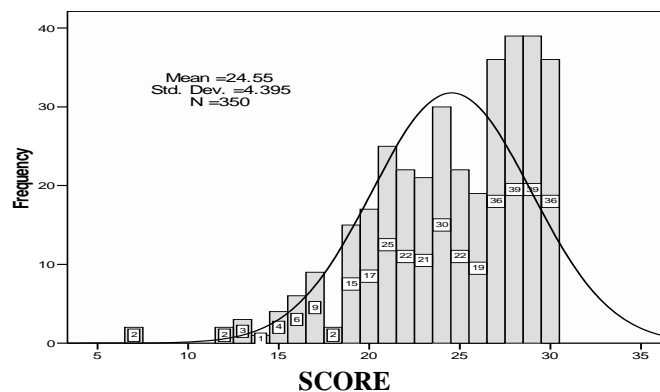
در مرحله بعد همبستگی بین پارامترهای دشواری و تمیز در مدل CTT و مدل ۲PL با استفاده از ضریب همبستگی پیرسون محاسبه و در ادامه به منظور بررسی معناداری پارامترهای سؤال و توانایی در نظریه کلاسیک و سؤال - پاسخ از آزمون معناداری T وابسته استفاده گردید.

همچنین به منظور بررسی پایایی و ثبات نتایج آزمون نظری گواهینامه رانندگی در اجرای اول، مجدداً این آزمون با همان شرایط اولیه، پس از یک هفته سپری شدن از اجرای اول، بر روی ۳۰ نفر از همان نمونه اولیه اجرا گردید. از آنجائیکه آزمون مورد نظر از نوع ملاک مرجع می باشد، جهت بررسی توافق بین تصمیم ها از ضریب پایایی به روش کاپا استفاده شد. داده های اولیه و اصلی جهت استفاده در تجزیه و تحلیل سؤالات آزمون نظری

گواهینامه رانندگی، پاسخ‌های اولیه کلیه شرکت کنندگان به آزمون مذکور در مراکز آموزش و آزمون گیری می‌باشد که در مراحل مختلف جمع آوری شده است. جهت انتخاب سؤالات پرسشنامه، از کتاب "آموزش جامع قوانین و مقررات راهنمایی و رانندگی" (از انتشارات معاونت راهور ناجا و منبع اصلی آزمون نظری). استفاده گردید. آن شامل ۱۰ بخش محتوایی با سؤالات اختصاصی در هر بخش بوده و در مجموع ۱۹۴ سؤال را در بر می‌گیرد. در مرحله اول از میان این سؤالات، ۳۰ سؤال به شیوه کاملاً تصادفی و با در نظر داشتن میزان نسبت در هر بخش محتوایی، انتخاب شد. از میان سؤالات انتخابی تعدادی سؤال تصویری و رنگی نیز در نظر گرفته شد. در انتها گزینه‌های انحرافی و گزینه صحیح سؤالات به شیوه تصادفی برای سؤالات آزمون در نظر گرفته شد.

یافته‌ها

در میان حجم نمونه (۳۵۰ نفر) مورد بررسی ۶۵/۷ درصد افراد مرد و ۳۴/۳ درصد بقیه زن بوده‌اند، که میانگین و انحراف استاندارد نمرات افراد در آزمون نظری برگزار شده، ۲۴/۵۵ و ۴/۴ بوده و با توجه به ملاک قبولی در آزمون (کسب نمره ۲۶ از ۳۰) تعداد ۱۶۹ نفر قبول (۴۸/۲۸ درصد) و ۱۸۱ نفر مردود شده‌اند. این نشان می‌دهد بیش از نصف افراد گروه، به ملاک مورد نظر، جهت قبولی در آزمون نرسیده‌اند (نمودار ۱).



نمودار ۱. منحنی توزیع فراوانی نمرات آزمون گواهینامه رانندگی

در مرحله اول، جهت تجزیه و تحلیل سؤالات آزمون بر طبق نظریه کلاسیک آزمون، درجه دشواری، ضریب تمیز و انحراف استاندارد سؤالات محاسبه گردید که نتایج آن در جدول ۱ ارائه گردیده است.

جدول ۱- پارامترهای سؤالات آزمون گواهینامه رانندگی بر اساس نظریه کلاسیک (CTT)

سؤال	ضریب دشواری (P)	ضریب تمیز (I _{pbis})	انحراف استاندارد	سؤال	ضریب دشواری (P)	ضریب تمیز (I _{pbis})	انحراف استاندارد
۱	.۰۹۷	.۰۲۳	.۰۱۷	۱۶	.۰۵۵	.۰۲۴	.۰۵۰
۲	.۰۷۱	.۰۳۹	.۰۴۵	۱۷	.۰۸۲	.۰۳۵	.۰۳۸
۳	.۰۸۹	.۰۲۹	.۰۳۱	۱۸	.۰۹۳	.۰۲۸	.۰۲۴
۴	.۰۸۶	.۰۳۱	.۰۳۵	۱۹	.۰۸۸	.۰۳۵	.۰۳۲
۵	.۰۹۴	.۰۱۷	.۰۲۳	۲۰	.۰۷۷	.۰۳۹	.۰۴۲
۶	.۰۶۶	.۰۳۳	.۰۴۷	۲۱	.۰۷۴	.۰۳۸	.۰۴۳
۷	.۰۹۴	.۰۳۷	.۰۲۲	۲۲	.۰۷۲	.۰۵۱	.۰۴۵
۸	.۰۹۴	.۰۳۹	.۰۲۴	۲۳	.۰۵۸	.۰۳۲	.۰۴۹
۹	.۰۸۲	.۰۴۴	.۰۳۸	۲۴	.۰۹۲	.۰۲۱	.۰۲۷
۱۰	.۰۹۵	.۰۳۳	.۰۲۰	۲۵	.۰۶۶	.۰۴۲	.۰۴۷
۱۱	.۰۷۸	.۰۰۴	.۰۴۱	۲۶	.۰۸۵	.۰۲۹	.۰۳۶
۱۲	.۰۹۳	.۰۰۴	.۰۲۵	۲۷	.۰۸۲	.۰۳۷	.۰۳۸
۱۳	.۰۹۲	.۰۳۲	.۰۲۷	۲۸	.۰۵۲	.۰۳۰	.۰۵۰
۱۴	.۰۷۹	.۰۳۷	.۰۴۱	۲۹	.۰۷۲	.۰۲۸	.۰۴۵
۱۵	.۰۹۱	.۰۳۲	.۰۲۸	۳۰	.۰۹۵	.۰۱۷	.۰۲۲

نتایج نشان می‌دهد که دامنه ضریب دشواری (P) سؤالات آزمون بین ۰/۵۲ (سؤال ۲۸) و ۰/۹۷ (سؤال ۱) در نوسان بوده است. بنابراین بر اساس این نتایج، سؤالات ۶، ۱۶، ۲۳، ۲۵ و ۲۸ سؤالات مناسبی محسوب شده و بقیه سؤالات نیز به دلیل ضرایب دشواری بالا، سؤالات آسانی محسوب می‌شوند. میانگین و انحراف استاندارد پارامتر دشواری سؤالات ۰/۸۱ و ۰/۱۲ است که در مجموع نشان‌دهنده آسان بودن کل آزمون می‌باشد. دامنه ضریب تمیز (I_{pbis}) سؤالات آزمون بین ۰/۱۷ (سؤالات ۵ و ۳۰) و ۰/۵۱ (سؤال ۲۲) در نوسان بوده است. به استثنای سؤالات ۱، ۳، ۵، ۱۶، ۱۸، ۲۴، ۲۶، ۲۹ و ۳۰ که دارای ضرایب تمیز پایینی می‌باشند، بقیه سؤالات از ضرایب تمیز نسبتاً خوبی (بالاتر از ۰/۳)

برخورد دارند. میانگین و انحراف استاندارد پارامتر تمییز سؤالات ۰/۳۳ و ۰/۰۸ است که نشاندهنده قدرت تمییز نسبتاً مناسب کل آزمون در جداسازی افراد قوی و ضعیف می‌باشد. بر طبق نتایج جدول بالا، ارتباط خاصی میان سؤالات با ضرایب دشواری بالا و پایین مشاهده نمی‌شود.

در مرحله بعد جهت انجام تحلیل سؤالات با استفاده از نظریه سؤال - پاسخ (IRT)، ابتدا مفروضه‌های اساسی و اولیه نظریه IRT، یعنی مفروضه‌های تک بعدی بودن و استقلال موضعی بررسی گردید. مفروضه اول با استفاده از ضریب پایایی و تحلیل عاملی مورد بررسی قرار گرفت.

برای بررسی پایایی سؤالات آزمون به روش لوپ، ابتدا ضریب پایایی کلیه سؤالات و میزان پایایی تک تک سؤالات تعیین گردید. سپس با استفاده از فرمول آلفای کرانباخ ضریب پایایی کل آزمون $\alpha = 0.82$ بدست آمد، که نشاندهنده همسانی درونی بالای سؤالات آزمون می‌باشد.

قبل از انجام تحلیل عاملی باید از اعتبار داده‌ها اطمینان داشته باشیم، برای این منظور از آزمون کایزر- میجر- الکین (KMO) که روشی برای بررسی صحت نمونه‌گیری می‌باشد، استفاده شده است (جدول ۲).

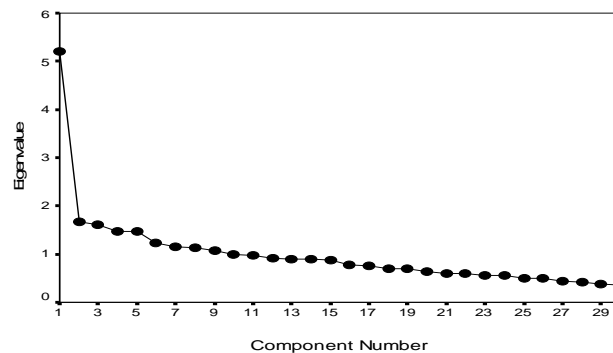
جدول ۲. اندازه KMO و آزمون کرویت بارتلت

۰/۷۶۷	اندازه کفایت نمونه‌گیری کایزر- میجر- الکین (KMO)	
۱۵۱۰/۸۲۴	مجذور کای تقریبی	
۴۳۵	درجات آزادی	آزمون کرویت بارتلت
۰/۰۰۱	سطح معناداری	

با توجه به جدول ۲، مقدار KMO برابر ۰/۷۸۹ بوده که نشان دهنده کفایت نسبتاً بالای نمونه‌گیری و تناسب داده‌ها برای تحلیل عاملی است، و از آنجایی که داده‌های بالاتر از ۰/۷ برای تحلیل مناسب هستند، لذا مقدار KMO بدست آمده رضایت بخش است.

همچنین آزمون کرویت بارتلت در سطح $0/001$ ، با درجات آزادی ۴۳۵ و با مقدار مجذور کای $1510/824$ معنادار بوده است و می توان گفت که داده ها برای اجرای تحلیل عاملی از تناسب قابل قبولی برخوردار بوده و نتایج تحلیل عاملی قابل اعتماد می باشد.

با توجه به نتایج بدست آمده، داده ها برای آزمون تحلیل عاملی از تناسب برخوردار بوده، و ماتریس همبستگی حاصل از تحلیل عاملی نشان می دهد که، عامل اول با ارزش ویژه (P-Value) $5/21$ حدود $17/36$ درصد واریانس عامل ها را تبیین کرده است. (سه برابر عامل دوم با ارزش ویژه $1/67$ و $5/57$ درصد واریانس). این عامل غالب توسط نمودار اسکری (نمودار ۲) بدست آمده نیز تأیید شده است.



نمودار ۲. منحنی اسکری آزمون نظری گواهینامه رانندگی

بنابراین، می توان نتیجه گرفت که آزمون گواهینامه رانندگی تک بعدی بوده و یک عامل غالب، عملکرد افراد را در آزمون تبیین می کند. چنانچه فرض تک بعدی بودن برقرار باشد، مفروضه استقلال موضعی نیز برقرار است (همبلتون، ۱۹۸۹).

پس از اثبات تک بعدی بودن و استقلال موضعی آزمون (نمودار ۲)، که از پیش فرض های مهم تجزیه و تحلیل سؤالات آزمون بر طبق نظریه سؤال - پاسخ (IRT)، است (همبلتون و سوامیناتان، ۱۹۸۵؛ نقل از ویبرگ، ۲۰۰۴)، ضرایب دشواری، شیب، حدس، نوعی آزمون مجذور خی به همراه سطح احتمال مربوطه، برای مدل های یک، دو و سه پارامتری محاسبه گردید که نتایج آن در جداول ۳، ۴ و ۵ آورده شده است.

جدول ۳. مقادیر پارامترهای تعدادی از سؤالات آزمون در مدل یک پارامتری (۱PL) بر اساس

نظریه IRT

سؤال	دشواری (b)	خی دو χ^2	معناداری (p)
۱	-۳/۴۹	۱	۰/۶۱
۱۶	-۰/۲۴	۱۸/۵	*۰/۰۰۱
۲۲	-۱/۰۳	۱۶/۹	*۰/۰۰۱
میانگین	-۱/۸۲		
انحراف استاندارد	۰/۹۳		
-2 Log Likelihood = ۷۹۸۲/۸۲۸۳			

* عدم برازش سؤال با مدل در سطح ۰/۰۱

جدول ۴. مقادیر پارامترهای تعدادی از سؤالات آزمون در مدل سه پارامتری (۳PL) بر اساس

نظریه IRT

سؤال	دشواری (b)	شیب (a)	حدس (c)	خی دو χ^2	معناداری (p)
۴	-۰/۶۵	۱/۴۱	۰/۵	۶	۰/۳
۵	-۲/۳۵	۰/۵۹	۰/۵	۵	۰/۴۱
۲۵	۰/۰۲	۱/۳۸	۰/۳۵	۵/۵	۰/۴۹
میانگین	-۰/۷۶	۱/۳۵	۰/۴۵		
انحراف استاندارد	۰/۹۱	۰/۷۸	۰/۰۵		
-2 Log Likelihood = ۷۸۶۵/۲۹۷۹					

جدول ۵. مقادیر پارامترهای آزمون در مدل دو پارامتری (۲PL) بر اساس نظریه IRT

سؤال	دشواری (b)	شیب (a)	خی دو χ^2	معناداری (p)	سؤال	دشواری (b)	شیب (a)	خی دو χ^2	معناداری (p)
۱	-۲/۷	۱	۱/۶	۰/۴۶	۱۶	-۰/۳۶	۰/۴	۷/۶	۰/۴۹
۲	-۰/۹۶	۰/۷۵	۴/۵	۰/۷۱	۱۷	-۱/۵۳	۰/۷۹	۵	۰/۵۴
۳	-۱/۹۵	۰/۸۴	۸/۱	۰/۰۸	۱۸	-۲/۲۴	۰/۹۴	۰/۴	۰/۹۸
۴	-۱/۷۶	۰/۷۸	۵/۱	۰/۲۷	۱۹	-۱/۸۲	۰/۸۷	۱/۵	۰/۹۱
۵	-۳/۲۶	۰/۵۷	۵/۹	۰/۲۱	۲۰	-۱/۲۲	۰/۸	۷	۰/۴۳
۶	-۰/۷۹	۰/۶۱	۲/۵	۰/۹۶	۲۱	-۱/۱۲	۰/۷۶	۵/۳	۰/۶۲
۷	-۱/۹۷	۱/۴	۴/۶	۰/۲	۲۲	-۰/۸۲	۱/۱۷	۱/۸	۰/۹۳

۱. به دلیل برازش یافتن مدل ۲ پارامتری با مجموعه داده های آزمون، نتایج مقادیر پارامترهای آزمون در این مدل به صورت کامل آورده شده است.

سؤال	دشواری (b)	شیب (a)	خی دو χ^2	معناداری (p)	سؤال	دشواری (b)	شیب (a)	خی دو χ^2	معناداری (p)
۸	-۱/۷۹	۱/۵۹	.۳	.۱۹۵	۲۳	-۰/۴۹	.۵۲	۷/۲	.۵۱
۹	-۱/۲۸	۱/۰۷	.۷	.۱۹۸	۲۴	-۲/۶۳	.۶۵	۳/۳	.۶۶
۱۰	-۲/۰۳	۱/۴۵	.۳	.۱۹۶	۲۵	-۰/۷۱	.۱۸	۳/۱	.۸۷
۱۱	-۱/۲۵	۰/۸۳	۲/۶	.۱۹۱	۲۶	-۱/۹۳	.۶۴	۵/۶	.۴۶
۱۲	-۱/۷۳	۱/۵۷	۲/۴	.۱۲۹	۲۷	-۱/۴۶	.۸۷	۲/۶	.۷۶
۱۳	-۲/۰۶	.۹۶	۵	.۱۲۸	۲۸	-۰/۱۲	.۵۲	۱۵	.۰۵
۱۴	-۱/۳۴	.۷۷	۳/۴	.۱۷۶	۲۹	-۱/۱۹	.۵۵	۱۳/۳	.۰۶
۱۵	-۱/۹۵	.۹۹	۰/۴	.۱۹۸	۳۰	-۳/۱۹	.۶۳	۰/۵	.۹۷

-2 Log Likelihood = ۷۸۸۹/۲۷۷۷

با توجه به نتایج بدست آمده، سؤالات ۱۶ و ۲۲ با مدل یک پارامتری برازش ندارند، اما بقیه سؤالات، با مدل‌های دو و سه پارامتری دارای برازش هستند. سطح معناداری مقادیر χ^2 سؤالات نیز نشان می‌دهد که تمام سؤالات با مدل‌های دو و سه پارامتری برازش دارند ($P > ۰/۰۵$). بنابراین می‌توان نتیجه گرفت که مدل‌های دو و سه پارامتری برآورد دقیق‌تری از پارامترهای سؤال، نسبت به مدل یک پارامتری بدست می‌دهند.

اما جهت تعیین برازش مدل-داده‌ها^۱ و اینکه کدامیک از مدل‌های یک، دو و سه پارامتری با این مجموعه از داده‌ها برازش بهتری دارند، از مقدار $-2\log likelihood$ در آخرین چرخش و آزمون خی دو برای مقایسه آنها به صورت زیر استفاده گردید:

برای مدل‌های یک و دو پارامتری و تعیین اینکه کدامیک برازش بهتری با این مجموعه از داده‌ها دارد، نتیجه زیر بدست آمد:

$$\chi^2 = (-2\log likelihood_{1\rho L}) - (-2\log likelihood_{2\rho L})$$

$$\chi^2 = (۷۹۸۲/۸۲۸۳ - ۷۸۸۹/۲۷۷۷) = ۹۳/۵۵۰۶$$

با توجه به اینکه درجه آزادی آزمون ۳۰ می‌باشد، χ^2 جدول در سطح ۰/۰۵ و ۰/۰۱ به ترتیب ۵۰/۸۹ و ۴۳/۷۷ می‌باشد و چون قدر مطلق χ^2 محاسبه شده از χ^2 جدول در هر دو

سطح ۹۵ و ۹۹ درصد اطمینان بزرگ‌تر است، بنابراین فرض صفر رد می‌شود و نتیجه گرفته می‌شود که مدل دو پارامتری نسبت به مدل یک پارامتری برازش بهتری با این مجموعه از داده‌ها دارد.

برای مدل‌های دو و سه پارامتری و تعیین اینکه کدامیک برازش بهتری با این مجموعه از داده‌ها دارد، نتیجه زیر بدست آمد:

$$\chi^2 = (-2 \log \text{likelihood}_{2\rho L}) - (-2 \log \text{likelihood}_{3\rho L})$$

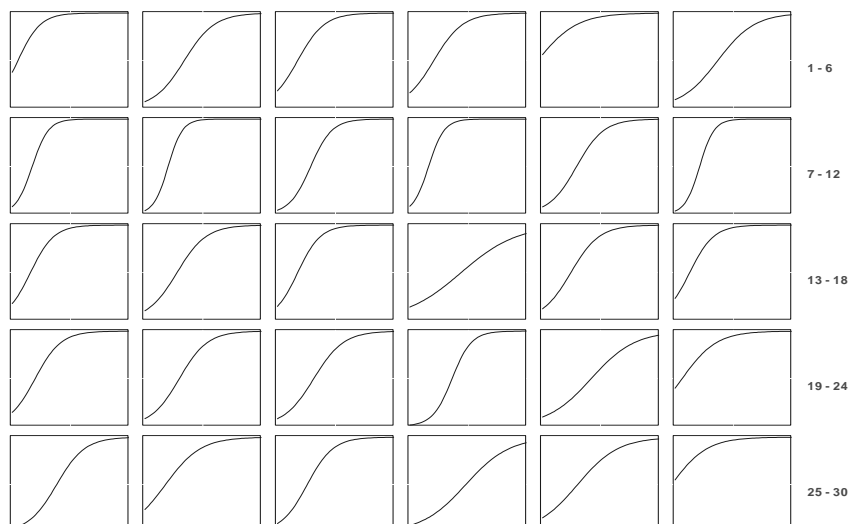
$$\chi^2 = (7889/2777 - 7865/2979) = 23/9798$$

با توجه به اینکه درجه آزادی آزمون ۳۰ می‌باشد، χ^2 جدول در سطح ۰/۰۵ و ۰/۰۱ به ترتیب ۵۰/۸۹ و ۴۳/۷۷ می‌باشد. چون قدر مطلق χ^2 محاسبه شده از χ^2 جدول در هر دو سطح ۹۵ و ۹۹ درصد اطمینان کوچک‌تر است، بنابراین فرض صفر تأیید می‌شود و نتیجه گرفته می‌شود که مدل دو پارامتری نسبت به مدل سه پارامتری برازش بهتری با این مجموعه از داده‌ها دارد.

البته به دلیل کوچک‌تر بودن χ^2 محاسبه شده از χ^2 جدول، به لحاظ آماری بین برازش مدل‌های دو و سه پارامتری تفاوت معناداری وجود ندارد، اما بنا بر اصل امساک (امپرتسون و رایس^۱، ۲۰۰۰؛ نقل از شریفی و همکاران، ۱۳۸۸) مدل دو پارامتری به دلیل دارا بودن پارامترهای کمتر نسبت به مدل سه پارامتری، به عنوان مدلی که بهترین برازش را با مجموعه داده‌ها دارد، معرفی می‌شود.

بنابراین با توجه به نتایج بدست آمده از مقایسه χ^2 ها، نتیجه گرفته می‌شود که مدل دو پارامتری نسبت به مدل‌های یک و سه پارامتری با داده‌های آزمون نظری آزمون گواهینامه رانندگی برازش بهتری دارد.

در نمودارهای زیر (نمودار ۳) منحنی‌های ویژگی‌های سؤالات (ICC) آزمون گواهینامه رانندگی در مدل دو پارامتری آمده است. توجه کنید که سؤالات از سمت چپ به راست شماره گذاری شده‌اند.



نمودار ۳. منحنی ویژگی سؤالات (ICC) آزمون گواهینامه رانندگی در مدل دو پارامتری

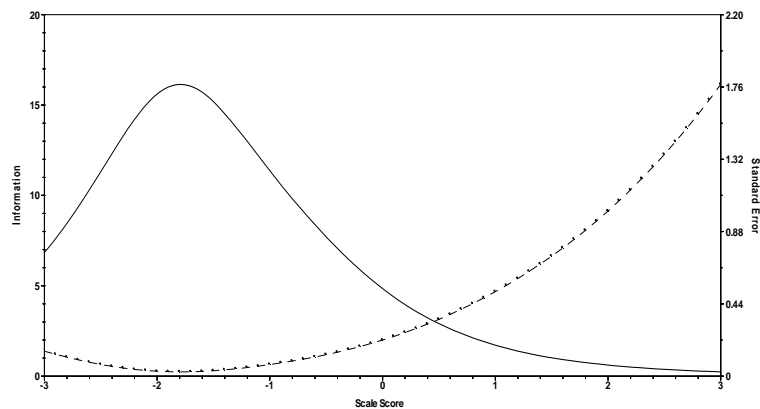
با توجه به مقادیر پارامترهای سؤالات آزمون در مدل دو پارامتری (جدول ۵)، و منحنی ویژگی سؤالات آزمون (نمودار ۳)، برای آزمون گواهینامه رانندگی و در مدل دو پارامتری نتیجه می‌گیریم که به استثنای سؤالات ۲، ۶، ۱۶، ۲۲، ۲۳، ۲۵ و ۲۸ که دارای ضرایب دشواری متوسط می‌باشند، بقیه سؤالات آزمون دارای ضرایب دشواری منفی و پایینی بوده و منحنی ویژگی آنها به سمت چپ متمایل است و بنابراین سؤالات آسانی محسوب می‌شوند. در این میان سؤال ۵ آسان‌ترین سؤال است.

شیب تمام سؤالات نیز قابل قبول بوده و از قدرت تشخیص مناسب برخوردارند، تنها سؤال ۱۶ دارای شیب و قدرت تشخیص نسبتاً پایین و کمتر از ۰/۵ است. در این میان سؤال ۱۲ از بالاترین شیب در بین سؤالات برخوردار است. همچنین بر اساس نتایج بدست آمده، میانگین و انحراف استاندارد پارامتر دشواری سؤالات ۱/۵۸- و ۰/۷۷ است که نشان دهنده آسان بودن کل آزمون می‌باشد. میانگین و انحراف استاندارد پارامتر شیب سؤالات نیز ۰/۸۶ و ۰/۳ است که نشان‌دهنده قدرت تمییز قابل قبول در جداسازی افراد قوی و ضعیف می‌باشد. در ادامه، بیشینه آگاهی سؤالات و سطوح توانایی مربوطه (جدول ۶) و نیز منحنی

آگاهی و خطای استاندارد آزمون گواهینامه رانندگی (نمودار ۴) آورده شده است. باید توجه شود در نقاطی از منحنی که آگاهی به بالاترین حد می‌رسد، خطای استاندارد به صفر میل می‌کند، در واقع با افزایش آگاهی آزمون، خطای استاندارد کاهش می‌یابد و بالعکس.

جدول ۶. بیشینه آگاهی و سطح توانایی تعدادی از سؤالات آزمون در مدل دو پارامتری

سؤال	بیشینه آگاهی	سطح توانایی	سؤال	سطح توانایی	بیشینه آگاهی
۱	.۷۳	-۲/۷	۱۶	.۱۱	-۰/۳۶
۲	.۴	-۰/۹۶	۱۷	.۴۶	-۱/۵۳
۳	.۵۱	-۱/۹۵	۱۸	.۶۴	-۲/۲۴
۴	.۴۴	-۱/۷۶	۱۹	.۵۵	-۱/۸۲
۵	.۲۳	-۳/۲۶	۲۰	.۴۷	-۱/۲۲
۶	.۲۷	-۰/۷۹	۲۱	.۴۲	-۱/۱۲
۷	۱/۴	-۱/۹۷	۲۲	۱	-۰/۸۲
۸	۱/۸۳	-۱/۷۹	۲۳	.۱۹	-۰/۴۹
۹	.۸۲	-۱/۲۸	۲۴	.۳۱	-۲/۶۳
۱۰	۱/۵۳	-۲/۰۳	۲۵	.۴۶	-۰/۷۱
۱۱	.۵	-۱/۲۵	۲۶	.۲۹	-۱/۹۳
۱۲	۱/۷۸	-۱/۷۳	۲۷	.۵۵	-۱/۴۶
۱۳	.۶۷	-۲/۰۶	۲۸	.۱۹	-۰/۱۲
۱۴	.۴۳	-۱/۳۴	۲۹	.۲۲	-۱/۱۹
۱۵	.۷۱	-۱/۹۵	۳۰	.۲۸	-۳/۱۹

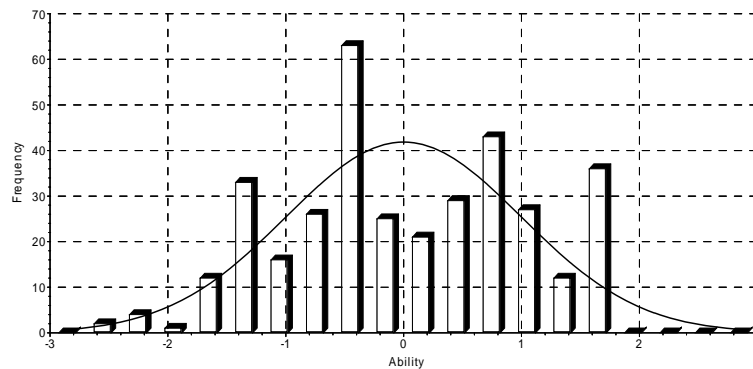


نمودار ۴. منحنی تابع آگاهی (TIF) و خطای استاندارد آزمون گواهینامه رانندگی در مدل دو پارامتری

در منحنی فوق خطوط پیوسته، منحنی آگاهی، و خطوط نقطه چین، منحنی خطای استاندارد آزمون گواهینامه رانندگی را نشان می‌دهد. بیشترین میزان آگاهی دهندگی در دامنه تئای ۲/۲- تا ۱/۴- است و این بدان معنی است که این آزمون برای افرادی که دارای این سطوح از توانایی هستند بیشترین مناسبت و کاربرد را دارد. در واقع میزان آگاهی دهندگی آزمون در سطوح پایین توانایی، بالا می‌باشد. بیشینه مقدار آگاهی آن نیز در سطح توانایی ۱/۹۵- و برابر ۱۵/۹۹ می‌باشد. از طرفی میزان آگاهی دهندگی این آزمون در سطوح توانایی بالا، بسیار اندک است. توزیع بیشینه آگاهی و خطای استاندارد نمودار ۴، و نیز توزیع فراوانی و سطوح مختلف توانایی آزمودنی‌ها، متعلق به نمودار ۵، در جدول ۷ آمده است.

جدول ۷. توزیع بیشینه آگاهی و خطای استاندارد توانایی آزمون گواهینامه رانندگی در مدل دو پارامتری

خطای استاندارد	بیشینه آگاهی	فراوانی	توانایی (تا)
۱/۰۴	۸/۴۴	۰	-۲/۸۵
۰/۶۴	۱۱/۳	۲	-۲/۵۵
۰/۳۶	۱۴/۲	۴	-۲/۲۵
۰/۲۳	۱۵/۹۹	۱	-۱/۹۵
۰/۲۵	۱۵/۷	۱۲	-۱/۶۵
۰/۴	۱۳/۷۷	۳۳	-۱/۳۵
۰/۶۳	۱۱/۳۶	۱۶	-۱/۰۵
۰/۹۴	۹/۰۸	۲۶	-۰/۷۵
۱/۳۲	۷/۰۵	۶۳	-۰/۴۵
۱/۸	۵/۳۴	۲۵	-۰/۱۵
۲/۱	۴/۳۸	۲۱	۰/۱۵
۲/۸۸	۳/۲۱	۲۹	۰/۴۵
۳/۷	۲/۳۴	۴۳	۰/۷۵
۴/۶۶	۱/۷	۲۷	۱/۰۵
۵/۷۸	۱/۲۴	۱۲	۱/۳۵
۷/۰۸	۰/۹۱	۳۶	۱/۶۵
۸/۵۸	۰/۶۷	۰	۱/۹۵
۱۰/۳	۰/۴۹	۰	۲/۲۵
۱۲/۲۶	۰/۳۶	۰	۲/۵۵
۱۴/۴۹	۰/۲۷	۰	۲/۸۵



نمودار ۵. منحنی توزیع توانایی آزمودنی‌ها در آزمون گواهینامه رانندگی در مدل دو پارامتری

برای تعیین معناداری پارامترهای سؤال و توانایی، در مرحله اول میانگین و انحراف استاندارد این پارامترها در مدل کلاسیک و مدل دو پارامتری بر اساس داده‌های بدست آمده اولیه (داده‌های خام) محاسبه گردید (جدول ۸).

جدول ۸. مقایسه پارامترهای سؤال و توانایی در آزمون گواهینامه رانندگی بر اساس نظریه IRT و CTT

شاخص آماری	IRT دشواری (b)	CTT دشواری (P)	IRT تمیز (a)	CTT تمیز (Ipbis)	IRT توانایی (θ)	CTT توانایی
میانگین	-۱/۵۸	/۸۲	/۸۶	/۳۳	- /۰۰۵	/۰۰۰۷
انحراف استاندارد	/۷۷	/۱۳	/۳	/۰۷	۱/۰۱	/۹۹
همبستگی پیرسون	- /۹		/۴۹		---	

سپس همبستگی بین پارامترهای دشواری و تمیز در مدل CTT و مدل ۲PL با استفاده از ضریب همبستگی پیرسون محاسبه گردید. ضرایب همبستگی بدست آمده به لحاظ آماری در سطح ۰/۰۱ معنادار هستند (جدول ۸).

همچنین جهت سنجش تفاوت معناداری بین پارامترهای سؤال در دو مدل، داده‌های محاسبه شده (داده‌های خام) پارامترهای سؤال، به نمرات Z تبدیل و از آنها آزمون معناداری T وابسته به عمل آمد. با توجه به نتایج بدست آمده (جدول ۹) از مقایسه میانگین

پارامترهای دشواری، شیب و توانایی در دو مدل، نتیجه گرفته می‌شود که، مقادیر t بدست آمده در سطح $0/05$ معنادار نبوده و بنابراین تفاوت معناداری بین دقت برآورد پارامترهای مذکور در دو نظریه کلاسیک و سؤال - پاسخ وجود ندارد.

جدول ۹. مقایسه پارامترهای سؤال و توانایی در آزمون گواهینامه رانندگی بر اساس نظریه IRT و

CTT

پارامترها	شاخص مدل	میانگین	انحراف استاندارد	T	df	Sig
دشواری	کلاسیک	$0/01$	$1/05$	$0/061$	۲۹	$0/952$
	دو پارامتری	$-0/01$	۱			
تمییز	کلاسیک	$0/01$	$0/97$	$-0/129$	۲۹	$0/898$
	دو پارامتری	$0/03$	$1/03$			
توانایی	کلاسیک	$0/0007$	$0/99$	$0/047$	۳۴۹	$0/963$
	دو پارامتری	$-0/005$	$1/01$			

در انتها، به منظور بررسی پایایی و ثبات نتایج آزمون نظری گواهینامه رانندگی، مجدداً آزمون مذکور با همان شرایط اولیه، پس از یک هفته سپری شدن از اجرای اول، بر روی ۳۰ نفر از همان نمونه اولیه اجرا گردید. از آنجائی که آزمون مورد نظر از نوع ملاک مرجع می‌باشد، برای بررسی توافق بین تصمیم‌ها از ضریب پایایی به روش کاپا استفاده شد. در این روش مقدار ضریب پایایی، نشان دهنده میزان توافق تصمیم‌های نمونه مورد نظر در ارتباط با چگونگی پاسخگویی به سؤالات آزمون در دو بار اجرای آزمون است. نمره حد تسلط در این آزمون کسب نمره ۲۶ در ۳۰ سؤال مورد نظر است. در واقع می‌توان گفت که ملاک قبولی در آزمون و یا نقطه برش آزمون، $86/67$ می‌باشد.

جدول ۱۰. طرح تعیین پایایی بر اساس توافق بین تصمیم‌ها

آزمون اول

رسیده به حد تسلط	رسیده به حد تسلط	آزمون دوم
۲	۱۰	
۱۷	۱	نرسیده به حد تسلط

$$N = a + b + c + d$$

بر اساس داده‌های اخذ شده از اجرای دوم، مقدار ضریب کاپا برابر ۰/۷۹ بوده و با توجه به اینکه مقدار $P = (۰/۰۰۱)$ در سطح ۰/۹۹ کمتر از ۰/۰۱ می‌باشد، می‌توان نتیجه گرفت که رابطه معناداری بین اجرای اول و اجرای دوم در نمونه مورد نظر وجود دارد. این نتیجه نشان می‌دهد که آزمون مورد نظر از پایایی و ثبات کافی در اجراهای مختلف برخوردار بوده است.

بحث و نتیجه‌گیری

معمولاً در علت‌شناسی تصادف‌های رانندگی از ۴ عامل انسانی، جاده، وسیله نقلیه و محیط نام برده می‌شود. اما در ۹۰ تا ۹۵ درصد تصادفات رانندگی در ایران، عامل انسانی نقش اصلی و اول را در حوادث رانندگی به عهده دارد (یعقوبی، ۱۳۷۹). با اینکه سالیانه بیش از ۱/۲۶ میلیون نفر در دنیا به علت تصادفات رانندگی جان خود را از دست می‌دهند (پدین، اسکار فیلد و سلیت، ۲۰۰۴)، اما بسیاری از کشورهای در حال توسعه درباره مرگ و میر در کشور خود اطلاعات محدودی دارند (خی و وانگ، ۲۰۰۴). با مطالعات اپیدمیولوژیک می‌توان در جهت کاهش حوادث و عواقب ناشی از آنها گام برداشت، چرا که آسیب‌های ناشی از حوادث رانندگی یک مشکل بزرگ بهداشتی در دنیا است (پدن و توریان، ۲۰۰۵).

در واقع آموزش این توانایی را دارد تا برای بهبود و پیشرفت کشورهای کمتر توسعه یافته زمینه لازم را فراهم سازد و سیاست‌های لازم برای مقابله با این بحران جهانی را اتخاذ نماید (لوئیز، ۱۹۹۰، نقل از سهرابی، ۱۳۸۴). نتایج مطالعات انجام شده نشان نیز می‌دهد که می‌توان با اقدامات پیشگیرانه از ۹۸ درصد حوادث جلوگیری نمود (عراقی و واحدیان، ۱۳۸۴).

یکی از فعالیت‌ها در این زمینه، توجه به آزمون‌های نظری گواهینامه رانندگی از زوایای مختلف و مقایسه آن با دیگر کشورها است که می‌تواند به عنوان اقدامی پیشگیرانه در کاهش رخداد حوادث رانندگی به شمار آید. آزمون گواهینامه رانندگی در ایران شامل یک آزمون نظری و یک آزمون عملی (شهری یا جاده) است. آزمون نظری، آزمون استاندارد شده‌ای است که برای تصمیم‌گیری در مورد رسیدن فرد به دانش نظری کافی و ساختن راننده‌ای ایمن و بی‌خطر طراحی شده، و می‌تواند از زوایای متفاوت مطالعه شده و سؤالات آن نیز بر طبق نظریات مختلف ارزش‌گذاری شود (ویبرگ، ۲۰۰۴).

هدف این پژوهش، بررسی ویژگی‌های روان‌سنجی سؤالات آزمون نظری گواهینامه رانندگی بر طبق نظریه کلاسیک آزمون (CTT) و نظریه سؤال - پاسخ (IRT) در ایران بوده است.

تحلیل ویژگی روان‌سنجی سؤالات آزمون نظری گواهینامه رانندگی طبق نظریه کلاسیک نشان داد که، بیشتر سؤالات این آزمون، سؤالات آسانی محسوب شده و افرادی با کمترین توانایی، قادر به پاسخگویی صحیح به تعدادی از سؤالات هستند. همچنین سؤالات آسان دارای انحراف استاندارد و واریانس پایینی بوده و به واریانس کل آزمون کمک زیادی نمی‌کنند و در واقع اطلاعات زیادی را درباره تفاوت‌های آزمودنی‌ها به ما نمی‌دهند. بیشتر سؤالات آزمون از ضرایب تمیز نسبتاً قابل قبولی برخوردارند. این بدان معنی است که بیشتر سؤالات آزمون قادر به جداسازی مناسب آزمودنی‌های قوی و ضعیف می‌باشند.

ضریب پایایی کل آزمون نیز نشان‌دهنده تجانس درونی نسبتاً بالای سؤالات آزمون می‌باشد.

تحلیل ویژگی‌های روان‌سنجی سؤالات آزمون نظری آزمون گواهینامه رانندگی طبق نظریه سؤال - پاسخ نشان داد که پس از برآورد پارامترها و با بررسی میزان برازندگی تک تک سؤالات با مدل‌های یک، دو و سه پارامتری، هیچ سؤالی وجود ندارد که به صورت همزمان با هر سه مدل برازش نداشته باشد. بنابراین هیچکدام از سؤالات حذف نمی‌شوند. بررسی منحنی ویژگی سؤالات نشان می‌دهد که بیشتر سؤالات آزمون دارای ضرایب دشواری منفی و پایینی بوده و منحنی ویژگی آنها به سمت چپ متمایل است و بنابراین سؤالات آسانی محسوب می‌شوند. میانگین و انحراف استاندارد پارامتر دشواری سؤالات ۱/۵۸- و ۰/۷۷. است که نشان دهنده آسان بودن کل آزمون می‌باشد. شیب بیشتر سؤالات آزمون قابل قبول بوده و از قدرت تشخیص مناسب در جداسازی آزمودنی‌های قوی و ضعیف برخوردارند. تنها سؤال ۱۶ دارای شیب و قدرت تشخیص پایین و کمتر از ۰/۵. است. میانگین و انحراف استاندارد پارامتر شیب سؤالات نیز ۰/۸۶. و ۰/۳. است که نشان دهنده قدرت تمیز قابل قبول در جداسازی افراد قوی و ضعیف می‌باشد.

در مجموع می‌توان نتیجه گرفت که بیشتر سؤالات آزمون دارای ضرایب دشواری پایین و ضرایب تمیز قابل قبولی می‌باشند.

بررسی منحنی آگاهی دهندگی نشان می‌دهد که بیشترین میزان آگاهی دهندگی این آزمون در دامنه توانایی ۲/۲- تا ۱/۴- بوده و بیشینه مقدار آگاهی آن نیز در سطح توانایی ۱/۹۵- و برابر ۱۵/۹۹ می‌باشد. این بدان معنی است که این آزمون در سطوح پایین‌تر (توانایی) از آگاهی دهندگی بالایی برخوردار بوده و برای آزمودنی‌هایی با توانایی پایین مناسب می‌باشد. در این میان سؤال ۸ با بیشینه آگاهی ۱/۸۳، در سطح توانایی ۱/۷۹- بیشترین میزان آگاهی دهندگی را در این مجموعه از سؤالات به خود اختصاص داده است. همچنین منحنی توزیع توانایی آزمودنی‌ها نشان می‌دهد که بیشترین فراوانی آزمون شوندگان با ۶۳ نفر متعلق به سطح توانایی ۰/۴۵- می‌باشد.

تحلیل ویژگی‌های روان‌سنجی سؤالات آزمون نظری گواهینامه رانندگی طبق نظریه

سؤال - پاسخ نشان داد که این آزمون تک بعدی می باشد. این نتیجه با تحلیل عاملی سؤالات آزمون بدست آمد. اثبات تک بعدی بودن آزمون به این معنا است که مفروضه استقلال موضعی نیز برقرار است. در مرحله بعد پارامترهای سؤال و توانایی در مدل های یک، دو و سه پارامتری بدست آمد. به منظور مقایسه مدل ها در برازش با داده های آزمون، از آزمون خی دو استفاده گردید. نتایج نشان داد که مدل دو پارامتری نسبت به مدل های یک و سه پارامتری برازش بهتری با داده های آزمون دارد. این بدان معنی است که عامل حدس در پاسخگویی آزمودنی ها به سؤالات نقش زیادی نداشته است. همچنین با توجه به اطلاعات بدست آمده، سؤالات ۱۶ و ۲۲ با مدل یک پارامتری برازش ندارند. اما تمام سؤالات با مدل های دو و سه پارامتری دارای برازش هستند. بنابراین می توان با توجه به برازش تمام سؤالات آزمون، با مدل های دو و سه پارامتری، نتیجه گرفت که مدل های دو و سه پارامتری برآورد دقیق تری از پارامترهای سؤال نسبت به مدل یک پارامتری بدست می دهد.

نتایج این پژوهش از لحاظ تفاوت در دقت برآورد پارامترهای سؤال و توانایی به روش سؤال - پاسخ نسبت به روش کلاسیک، نشان می دهد که، برآورد پارامترهای سؤال و توانایی به روش سؤال - پاسخ از واریانس و دامنه وسیع تری نسبت به روش کلاسیک برخوردار بوده و برخلاف مدل کلاسیک، قادر به برآورد پارامترهای سؤال و توانایی مستقل از همدیگر می باشد. به بیان دیگر آنچه به عنوان نقطه ضعف در نظریه CTT تلقی می شود، وابستگی پارامترهای آن به نمونه مورد بررسی است (همبلتون و سوامیناتان^۱، ۱۹۸۵، نقل از استیج^۲، ۲۰۰۰). در واقع ثبات و استقلال پارامترها در نظریه IRT، ویژگی و مزیت مهم این نظریه به شمار می رود (لرد^۳، ۱۹۸۰، نقل از استیج، ۲۰۰۰). در این رابطه شارکنس و دی آنجلو^۴ (۲۰۱۰)، نیز نشان دادند که اگرچه هر دو نظریه CTT و IRT می توانند در به دست آوردن اطلاعات یکسان در مورد مقدار سؤالات آزمون در رابطه با

1 . Hambleton & Swaminathan
 2 . Stage
 3 . Lord
 4 . Sharkness & DeAngelo

صفت مکنون اندازه‌گیری شده، مفید باشند، اما این دو نظریه اندازه‌های متفاوتی از دقت آزمون بدست می‌دهند.

به منظور مقایسه و تعیین معناداری پارامترهای سؤال و توانایی در آزمون گواهینامه رانندگی، در مرحله اول میانگین و انحراف استاندارد پارامترهای دشواری، شیب و توانایی در دو مدل کلاسیک و دو پارامتری بر اساس داده‌های بدست آمده اولیه (داده‌های خام) محاسبه گردید. سپس همبستگی بین پارامترهای دشواری و تمیز در مدل CTT و مدل ۲PL با استفاده از ضریب همبستگی پیرسون محاسبه گردید، که ضرایب همبستگی بدست آمده به لحاظ آماری در سطح ۰/۰۱ معنادار بودند. در مرحله بعد، داده‌های محاسبه شده (داده‌های خام) برای پارامترهای سؤال و توانایی در دو مدل، به نمرات Z تبدیل و از آن‌ها آزمون معناداری T وابسته به عمل آمد. با توجه به T محاسبه شده از مقایسه دو میانگین و T مندرج در جدول در سطح ۰/۰۵، نتیجه گرفته می‌شود که تفاوت معناداری بین دقت برآورد پارامترهای دشواری، شیب و توانایی در دو نظریه کلاسیک و سؤال - پاسخ وجود ندارد. با مقایسه نتایج متفاوت معناداری‌های بدست آمده توسط آزمون‌های رابطه (پیرسون) و تفاوت (T وابسته) می‌توان نتیجه گرفت که نظریه‌های کلاسیک و سؤال - پاسخ دارای رابطه معناداری بوده و هر دو نظریه از میزان دقت تقریباً مشابه و یکسانی در برآورد پارامترهای سؤال و توانایی برخوردارند و در واقع تفاوت معناداری بین دو نظریه در برآورد پارامترها وجود ندارد.

امروزه، با وجود مزیت‌های نظری شناخته شده نظریه IRT بر CTT، پژوهش‌های انجام یافته از طریق بررسی ویژگی‌های تجربی نیز به نتایج مشابه، متغیر و بعضاً متناقضی منتهی شده است.

بچگار^۱ و همکاران (۲۰۰۳) در مطالعه خود با استفاده از نظریه‌های IRT و CTT به این نتیجه رسیدند که؛ نظریه سؤال - پاسخ حالت گسترش یافته نظریه کلاسیک است، و

مفاهیم هر دو نظریه به هم وابسته هستند. با این حال ثابت شده است که نظریه سؤال - پاسخ در جایی که نظریه کلاسیک ناتوان به نظر می‌رسد می‌تواند مفید واقع شود. بررسی پارامترهای سؤال و توانایی با استفاده از تکنیک مونت کارلو و از طریق داده‌های شبیه‌سازی شده در مطالعه مکدونالد و پانونن^۱ (۲۰۰۲) نشان داد که، برآورد پارامترهای سؤال و توانایی در نظریه IRT و CTT قابل مقایسه، مشابه و دقیق اند. همچنین نتایج حاکی از آن بود که برآورد پارامترهای تمییز مثبتی بر IRT در اغلب شرایط تجربی دقیق بوده، اما برآورد پارامترهای تمییز مثبتی بر CTT در برخی از شرایط خاص از دقت پایین‌تری برخوردار بوده است. پیامدهای نتایج این مطالعه برای تجزیه و تحلیل روان‌سنجی سوالات و انتخاب سؤال مورد بحث قرار گرفته است. نتیجه پژوهش استیج (۲۰۰۰) نیز نشان داد که؛ نظریه‌های جدید و کلاسیک اندازه‌گیری به یک اندازه قادرند داده‌های آزمون را به صورت منظم پیش‌بینی کنند. اما به دلیل تفاوت قابل ملاحظه‌ای موجود بین دو نظریه در تئوری و اجرا و در نتیجه برآورد دقیق‌تر پارامترهای سؤال و توانایی توسط نظریه IRT نسبت به نظریه CTT، که در برخی تحقیقات به اثبات رسیده است، نظریه IRT برتری قابل ملاحظه‌ای بر نظریه CTT دارد.

در مطالعه سنکتهای و سنکتهای^۲ (۲۰۱۰) مشخص گردید که، مدل پاسخ مدرج (IRT)، برای برآورد پارامترهای آماری در یک نمونه بزرگ و نیز تحلیل مدل معادلات ساختاری دقیق‌تر از مدل کلاسیک آزمون است. اما مدل کلاسیک آزمون نیز همچنان ابزار آماری پایایی در تحلیل رگرسیون و در نمونه‌هایی با اندازه کوچک محسوب می‌شود. همچنین هنگامی که پارامترهای فرد، ناشناخته فرض می‌شود و پارامترهای سؤال نیز به صورت شناخته شده و یا شناخته نشده باشد، توان بدست آمده با استفاده از نظریه‌های IRT و CTT مشابه و یکسان، و همیشه پایین‌تر از توان مورد انتظار در نقطه انتهایی توزیع طبیعی

1 . Macdonald & Paunonen

2 . Siengthai & Siengthai

است. تعداد سؤالات نیز تأثیر قابل توجهی بر توان در هر دو روش دارد (سبیل^۱، ۲۰۱۰). تجزیه و تحلیل نهایی پارامترهای سؤال و آزمودنی‌ها در آزمون نظری گواهینامه رانندگی که توسط راهور ناجا برگزار می‌شود، نشان داد که سؤالات این آزمون بر اساس نظریه کلاسیک و نظریه سؤال - پاسخ، برای آزمودنی‌هایی با توانایی پایین از مناسبت و دقت بیشتری برخوردار است. این بدان دلیل است که با توجه به ضرایب دشواری بدست آمده، بیشتر سؤالات آزمون ساده و آسان می‌باشند. اما از آنجائی که آزمون مورد نظر از نوع ملاک مرجع است، سؤالی که از لحاظ محتوایی با ارزش است، لزوماً به دلیل آسان بودن از آزمون کنار گذاشته نمی‌شود (ویبرگ، ۲۰۰۴). پارامتر تمیز آزمون نیز در هر دو نظریه از ضرایب قابل قبولی برخوردار بوده و نشان می‌دهد که این آزمون تا حد زیادی از قدرت جداسازی افراد قوی و ضعیف در سطوح پایین توانایی برخوردار می‌باشد. البته با توجه به اینکه آزمون نظری گواهینامه رانندگی از نوع آزمون‌های ملاکی می‌باشد، حتی سؤال‌هایی با ضرایب تمیز پایین نیز کنار گذاشته نمی‌شوند، چرا که چنین سؤالاتی توانایی اندازه‌گیری هدف را نشان می‌دهند (کیامنش، ۱۳۸۷).

به منظور بررسی پایایی و ثبات نتایج آزمون نظری گواهینامه رانندگی، مجدداً آزمون مذکور با همان شرایط اولیه، پس از یک هفته سپری شدن از اجرای اول، بر روی ۳۰ نفر از همان نمونه اولیه اجرا گردید. بر اساس ضریب کاپای بدست آمده می‌توان نتیجه گرفت که رابطه معناداری بین اجرای اول و اجرای دوم در نمونه مورد نظر وجود دارد. این نتیجه نشان داد که آزمون مورد نظر از پایایی و ثبات کافی در اجراهای مختلف برخوردار است. در واقع تغییر چندانی در تعداد و درصد افرادی که در اجرای اول آزمون موفق و یا ناموفق بوده‌اند، نسبت به اجرای دوم آزمون وجود نداشته و می‌توان نتیجه گرفت که، ثبات تصمیم‌گیری در دو بار اجرای آزمون وجود داشته است.

پایین بودن میانگین نمرات آزمودنی‌ها در این پژوهش، نسبت به نمره ملاک، با در نظر

گرفتن سادگی آزمون، نشاندهنده توانایی پایین آزمون شوندگان می‌باشد. البته میزان توانایی برآورد شده در نظریه سؤال - پاسخ در مقایسه با نمره آزمودنی در نظریه کلاسیک، به میزان واقعی نزدیک‌تر است و با توجه به توانایی‌های برآورد شده، می‌توان سؤالات متناسب با توانایی آزمودنی‌ها را انتخاب نمود و این می‌تواند به ایجاد بانک سؤال و توسعه آزمون‌های انطباقی کامپیوتری (CAT)، در آزمون نظری گواهینامه رانندگی منجر شود. مزیتی که در بررسی هاروی^۱ (۲۰۰۲) نیز بدان اشاره شده است. وی با بررسی سؤالات آزمون تجدید نظر شده مایرز - بریگز (MBTI)، مزیت‌های نظریه IRT بر روش‌های مبتنی بر نظریه CTT را بدین گونه برشمرد:

- (الف) دادن شرح مفصل از عملکرد افراد در سؤالات آزمون
- (ب) ثبات و تغییر ناپذیری شاخص‌های دقت سؤال و آزمون در سراسر طیف نمرات
- (ج) امکان ارزیابی سوگیری سؤال و آزمون در رابطه با زیر گروه‌های جمعیتی
- (د) اندازه‌گیری کیفی و با ثبات از ویژگی‌های هر پاسخ دهنده
- (ه) امکان تهیه آزمون‌های انطباقی کامپیوتری با هدف کاهش زمان آزمایش، بدون فدا کردن دقت اندازه‌گیری.

یافته‌های دیگر (شارکنس و دی آنجلو، ۲۰۱۰) نیز نشان داد که در مجموع نظریه IRT اطلاعات بسیار جامع‌تری در مورد دقت اندازه‌گیری و نیز نقشه راه واضح‌تری برای بهبود مقیاس فراهم می‌کند. همچنین این یافته‌ها، از نظریه IRT برای ساخت مقیاس و توسعه زمینه‌یابی در آموزش عالی پشتیبانی می‌کنند.

منابع

- آلن، مری. جی؛ و وندی، ام. یین (۱۳۸۴). مقدمه ای بر نظریه های اندازه‌گیری (روان‌سنجی). ترجمه علی دلاور. تهران: سمت. صص ۱۳-۱۲. (تاریخ انتشار به زبان اصلی، ۱۹۸۲).
- احمدی، فهیمه (۱۳۸۵). علل وقوع تصادفات در معابر شهری.
- امبرتسون، سوزان ای؛ و رایس، استیون پی (۱۳۸۸). نظریه های جدید روان‌سنجی (IRT) برای روان‌شناسان. ترجمه حسن پاشاشریفی، ولی الله فرزاد، مجتبی حبیبی عسگرآباد و بلال ایزانلو. تهران: رشد. ص ۴۹۰. (تاریخ انتشار به زبان اصلی، ۲۰۰۰).
- بازرگان، عباس (۱۳۸۶). ارزشیابی آموزشی: مفاهیم، الگوها و فرآیند عملیاتی. تهران: سمت.
- بیمه ایران (۲۰۰۸). تفاوت‌های حوادث رانندگی در آلمان و ایران. [Http://www.bih.ir](http://www.bih.ir).
- ثرن‌دایک، آر، ال (۱۳۷۵). روان‌سنجی کاربردی. ترجمه حیدرعلی هومن. تهران: دانشگاه تهران. (تاریخ انتشار به زبان اصلی، ۱۹۸۲).
- حسینی، میرزا حسن (۱۳۸۰). لزوم توجه به نکته های مثبت آموزش از راه دور در دنیا.
- خالدی، محمد (۱۳۸۷). ارزیابی اثربخشی آموزش‌های فنی و حرفه ای، هنرستان‌ها و دوره های علمی و کاربردی به تفکیک خصوصی و دولتی در اشتغال، مؤسسه کار و تأمین اجتماعی. خبرگزاری فارس (۱۳۸۵). گروه اجتماعی، حوزه قضایی و انتظامی. شماره ۱۸۴/۱۰۱۸۵۰۸؛ ۱۳۸۵/۰۸/۰۱.
- سلمانی، محمد؛ رمضان زاده لسبویی، مهدی؛ دریکوند، مسلم و ثابتی، فرخ (۱۳۸۷). بررسی عوامل مؤثر بر تصادفات جاده ای و ارائه راهکارهایی برای کاهش آن، مورد مطالعه: منظومه روستایی جنوب خور و بیابانک. پژوهش‌های جغرافیای انسانی، شماره ۶۵، پاییز ۱۳۸۷، صص ۱۰۴-۸۷.
- سهرابی، منیره (۱۳۸۴). بررسی تطبیقی سیر تحولات آموزش از راه دور در توسعه کمی و کیفی آموزش عالی به ویژه دانشگاه های مجازی در کشورهای انگلستان، آمریکا و ایران، تهران: پایان نامه کارشناسی ارشد، دانشگاه علامه طباطبائی.
- سوری، حمید؛ عینی، الهه؛ موحدی نژاد، عباسعلی؛ محفوظ پور، سعاد؛ موحدی، محمد؛ رضا زاده آذری، منصور؛ وفايي، رضا؛ حاتم‌آبادی، حمیدرضا و مسعودی‌نژاد، محمدرضا (۱۳۸۸). ارایه الگوی عملی نقشه سیاست گذاری در سوانح ترافیکی کشور در سال ۱۳۸۷. مجله پژوهشی حکیم، پاییز ۸۸ دوره دوازدهم، شماره سوم.

صالحی، احمد (۱۳۷۶). بررسی مشخصه های روان سنجی آزمون پره بورده رشته زنان و زایمان کشور ایران بر اساس نظریه سؤال - پاسخ IRT، پایان نامه کارشناسی ارشد، تهران: دانشگاه علامه طباطبائی.

عراقی، عزت؛ و واحدیان، محمد (۱۳۸۴). بررسی عوامل مستعد کننده و آسیب های ناشی از تصادفات با موتورسیکلت در شهرستان مشهد سال ۱۳۸۴. مجله افق دانش (دوره ۱۳، شماره ۱). دانشکده علوم پزشکی و خدمات بهداشتی، درمانی گناباد.

فراهانی، مهدی (۱۳۷۵). مقایسه مدل های اندازه گیری (کلاسیک و سؤال - پاسخ) از لحاظ برآورد پارامترهای سؤال و توانایی. پایان نامه کارشناسی ارشد، تهران: دانشگاه علامه طباطبائی.

کیامتش، علیرضا (۱۳۸۷). روش های ارزشیابی آموزشی. تهران: دانشگاه پیام نور. صص ۱۱۹-۱۱۸.

محمد زاده رومیانی، مهری (۱۳۷۵)، روش های گزینش سؤال در مدل کلاسیک اندازه گیری و مدل های جدید IRT، پایان نامه کارشناسی ارشد، تهران، دانشگاه علامه طباطبائی.

مگنوسون، داوید (۱۳۷۰). مبانی نظری آزمون های روانی. ترجمه محمد نقی براهنی. تهران: نشر دانشگاه تهران. (تاریخ انتشار به زبان اصلی، ۱۹۶۶).

یعقوبی، حمید (۱۳۷۹). بررسی نقش عوامل انسانی در بروز تصادفات رانندگی در ایران. نشریه اندیشه و رفتار، سال ششم، شماره یک.

یونسسی، جلیل (۱۳۸۵)، بررسی ویژگی های روان سنجی سؤالات آزمون های فراگیر رشته روانشناسی دانشگاه پیام نور در سال ۱۳۸۵، پایان نامه کارشناسی ارشد، تهران، دانشگاه علامه طباطبائی.

Bechger, T. M., Maris, Gunter., Verstralen, H. H. F. M. & Béguin, A, A. (2003). Using Classical Test Theory in Combination with Item Response Theory. Applied Psychological Measurement Vol. 27, No 5, pp 319-334.

Chi, GB., & Wang, sy. (2004). Pattern of road traffic injuries in china. Zhonghua lio xing bing zue za zhi. (7):598-601.

Hambleton, R. K., & Vanderlinden, W. J. (1982). Advance in item response theory and application: An introudction applied psychological measurement, 6,4, 372-378.

Hambleton, R. K., & Cook, L. L. (1983). The robustness of item rrsponse models effects of test length and sample aize on the precision of ability estimates. In D jweis(Ed.) New horizons in testing (pp.31- 49) New York: Academic press.

Hambleton, R. K. (1989). Principles and selected applications of item- response theory. In R. Linn (Ed.) Educational measurement, (3rd Ed). New York: Macmillan. 147-200.

- Hambleton, R. K., Jones, R. W. & Rogers, H. J. (1993). Independence of item parameter estimation errors in test development. *Journal of Educational Measurement*, 30, 143 – 155.
- Harvey, R. J., & Hammer, A. L. (2002). *Item Response Theory*. Virginia Polytechnic Institute & State University & Consulting Psychologists Press, Inc.
- Kopits, E., & Cropper, M. (2005). Traffic fatalities and economic growth, *Accidentalprev*; 37(1):169-78.
- Macdonald, P., & Paunonen, S. V. (2002). A Monte Carlo Comparison of Item and Person Statistics Based on Item Response Theory versus Classical Test Theory. *Educational and Psychological Measurement* Vol. 62 No. 6, pp 921-943. University of Western Ontario.
- Peden, M., & Toroyan, T. (2005). Counting road traffic deaths and injuries: poor data should not detract from doing some thing. *Annals of emergency medicine*. 46(2):158-60.
- Peden, M., Scuffield, R., & Sleet, D. (2004). *World report on road traffic injury prevention*. Geneva: world health organization.
- Reid, C. A. (1993). Latent trait modeling of the general aptitude test battery used with a rehabilitation client population: An investigation of Model – Data Fit. Source: DAI – B 54/12 , P. 6497, JUN 1994.
- Romana, E. (2004). *World health day: road safety is no accident*. Paris, France.
- Sébille, V., Hardouin, J.B., Le Néel, T., Kubis, G., Boyer, F., Guillemain, F., & Falissard, B. (2010). Methodological issues regarding power of classical test theory (CTT) and item response theory (IRT) -based approaches for the comparison of patient-reported outcomes in two groups of patients- a simulation study. *BMC Medical Research Methodology*.
- Sharkness, J., & DeAngelo, L. (2010). Measuring Student Involvement: A Comparison of Classical Test Theory and Item Response Theory in the Construction of Scales from Student Surveys. *Res High Educ* 52, pp 480–507.
- Siengthai, Sukirno & Sununta. (2010). The comparison of graded response model and classical test theory in human resource research: a model fitness test. *Research and Practice in Human Resource Management*. 18 (2), pp77-90. Singapore, Human Resources Institute & Curtin University of Technology.
- Stage, C. (2000). *A Comparison Between Item Analysis Based on Item Response Theory and Classical Test Theory. A Study of the SweSAT Subtest ERC*.
- Wiberg, M. (2004). Classical test theory vs. item response theory: An evaluation of the theory test the Swedish Driving-License test. 1-27.