

## روایی آزمون<sup>۱</sup> و عوامل به وجود آورنده<sup>۲</sup> واریانس نامربوط به سازه<sup>۳</sup>: بررسی نقش رایانه، جنسیت، سن، زبان مادری، رشته تحصیلی در نتایج

تافل آی. بی. تی

محمد صالحی\*

استادیار آموزش زبان انگلیسی، مرکز زبانها و زبانشناسی، دانشگاه صنعتی شریف،  
تهران، ایران

آرش نعمت زاده\*\*

کارشناسی ارشد آموزش زبان انگلیسی، مرکز زبانها و زبانشناسی، دانشگاه صنعتی شریف،  
تهران، ایران

(تاریخ دریافت: ۹۶/۰۷/۱۱، تاریخ تصویب: ۹۶/۱۰/۳۰، تاریخ چاپ: اردیبهشت ۱۳۹۷)

### چکیده

عوامل متعددی می‌توانند در روایی یک آزمون تأثیر داشته باشند. این عوامل می‌توانند نوعی تهدید به حساب بیایند. آری، جیکاب، سورنسون و واکر (۲۰۱۴)<sup>۳</sup> از آن‌ها با عنوان عوامل واریانس نامربوط به سازه یاد می‌کنند. رایانه به مثابه یکی از ابزارهای فنی برجسته زندگی روزمره ما را متحول ساخته است. رایانه در محیط‌های آموزشی، بر شیوه یادگیری ما تأثیر گذاشته است. در برخی موارد، رایانه می‌تواند نتایج آزمون‌ها را به حاشیه برد و تبدیل به منبع تهدیدی برای روایی آزمون شود. علاوه بر رایانه، جنسیت، سن، زبان مادری و رشته تحصیلی نیز متغیرهایی هستند که می‌توانند واریانس نامربوط به سازه تولید کنند. در تحقیق حاضر، نمرات آزمون تافل آی. بی. تی به صورت متغیر وابسته مورد تحقیق قرار گرفت؛ بدین منظور، صد نفر در قالب دو گروه علوم انسانی و مهندسی به پرسشنامه‌هایی در مورد آشنایی با رایانه و دیدگاهشان درباره رایانه پاسخ دادند و نیز در یک تافل آی. بی. تی شرکت کردند؛ در مرحله اول، سطح آشنایی با رایانه بررسی شد و در مرحله بعدی، دیدگاه شرکت‌کنندگان در مورد رایانه و تأثیر سه متغیر مشخص شده (سن، جنسیت و زبان مادری) بر نتایج آزمون مورد مطالعه قرار گرفت. نتایج با استفاده از روش‌های متفاوت، از جمله تجزیه و تحلیل عامل یک‌طرفه و سه‌طرفه بررسی شد. نتایج به دست آمده نشان‌دهنده روابط معناداری بین این متغیرها و نمرات آزمون نبود و با ارجاع به این نتایج، این‌طور می‌توان ادعا کرد که رایانه و سایر عوامل ذکر شده، هیچ تأثیر نامطلوبی (واریانس نامربوط به سازه) بر نتایج آزمون نداشتند.

**واژه‌های کلیدی:** آشنایی با رایانه، روایی آزمون، واریانس نامربوط به سازه، آزمون‌های مبتنی بر رایانه.

1. Test Validity

2. Construct Irrelevant Variance Factors

\* E-mail: [m\\_salehi@shrarif.ir](mailto:m_salehi@shrarif.ir) نویسنده مسئول

\*\* E-mail: [m\\_salehi@shrarif.ir](mailto:m_salehi@shrarif.ir)

3. Ary, Jacobs, Sorensen, & Walker

## ۱- مقدمه

در طول چند دهه اخیر، به واسطه پیشرفت‌های حاصل در فناوری به‌خصوص رایانه، شاهد تغییرات متعددی در جنبه‌های مختلف زندگی‌مان بوده‌ایم. آزمون زبان، بخشی از این حوزه گسترده است که دچار تحولات متعددی شده است. یکی از عواقب این پدیده را می‌توان تغییر فرمت یک آزمون رایج مانند تافل (آزمون زبان انگلیسی به‌عنوان زبانی خارجی) دانست که اساساً بر روی کاغذ چاپ می‌شد و در اختیار آزمون‌دهندگان قرار می‌گرفت. با این حال، بعد از معرفی رایانه به این حوزه، آزمون‌هایی مانند تافل رایانه‌ای برگزار شدند که در آن، زبان‌آموزان، به سؤالات مختلف آزمون، از طریق رایانه پاسخ می‌دادند. همان‌طور که آلدرسن<sup>۱</sup> (۲۰۰۰) ادعا کرده است، اولین نسخه CBT در ۱۹۹۸ عرضه شد. از آن زمان تاکنون، آزمون‌های بیشتری از طریق اینترنت، سی‌دی رام یا هردوی آن‌ها عرضه شدند. یکی از مزیت‌های این آزمون‌های رایانه‌ای، در زمینه آموزش از راه دور است؛ برای مثال، بسیاری از نهادها و دانشگاه‌ها مانند دانشگاه کالیفرنیا و لوس آنجلس (UCLA)، آزمون‌های تعیین سطح زبان دوم را عرضه کردند که مبتنی بر وب هستند و در آینده، می‌توان این آزمون‌ها را در اختیار زبان‌آموزانی قرارداد که خواهان ورود به ایندانشگاه‌ها هستند. فولچر<sup>۲</sup> (۱۹۹۹) به آزمون تعیین سطح زبان انگلیسی رایانه‌ای اشاره کرده است که به‌مثابه ابزاری برای بررسی سوگیری بالقوه علیه دانش‌آموزانی استفاده می‌شد که آشنایی کافی با رایانه نداشتند یا دیدگاهی منفی نسبت به فناوری داشتند.

هنگام استفاده از آزمون‌های رایانه‌ای، مسائل بحث‌برانگیز بسیاری قابل طرح هستند. یکی از ویژگی‌هایی که تیلور، کیرش، جمیسون و ایگنور<sup>۳</sup> (۱۹۹۹) به آن اشاره کردند، این است که رایانه‌ها ممکن است واریانس ساختاری غیرمرتبط را وارد محیط آزمون بکنند. رایانه در جایگاه یکی از نمادهای فناوری، می‌تواند باعث تغییر عقاید مرسوم در مورد مسائل گوناگونی شود. این ظرفیت رایانه در حوزه آزمون‌سازی زبان، قابل توجه است. تغییر قالب آزمون‌ها و طراحی آزمون‌های جدید، نشان‌دهنده تلاش‌های طراحان آزمون‌ها در جهت هم‌راستایی با تغییرات فناوری هستند. با توجه به این تغییرات، رایانه تأثیر اجتناب‌ناپذیر خود بر فرایند اجرای آزمون را نشان داده است. آزمون‌های مرسوم کاغذی، پیش‌شرط‌های متفاوتی دارند و این امر، به آن‌ها

1. Alderson

2. Fulcher

3. Taylor, Kirsch, Jamieson, Eignor

ظاهری می دهد که زبان آموزان با آن ها به خوبی آشنا هستند و از طریق فنون متفاوتی، می توانند این پیش شرط ها و الزامات را رعایت کنند. از سوی دیگر، با توجه به ماهیت متمایز رایانه، شیوه انجام آزمون رایانه ای، نسبت به روش های قبلی، کمی متفاوت تر است. این تفاوت منجر به پدیده ای می شود که مؤلفان متفاوتی مانند تیلور و همکاران (۱۹۹۹) و ویر<sup>۱</sup> (۲۰۰۵) آن را آشنایی با رایانه می نامند. آشنایی با رایانه به مثابه یکی از ویژگی های متمایز فناوری در این دوره، معیار مهمی است که بر آزمون دهندگان اعمال می شود.

محققان متفاوتی سعی در بررسی تأثیر رایانه بر عملکرد آزمون و نتایج آزمون دهندگان داشته اند؛ برای مثال، ویر او سالیوان، جین یان و باکس<sup>۲</sup> (۲۰۰۷) به بررسی واکنش آزمودنی ها به قالب رایانه ای در مقابل قالب کاغذی سستی بخش نوشتاری آزمون آیلتس پرداخت. مهم ترین نتیجه این پژوهش، نبود تفاوتی معنادار در نمراتی بود که دو ارزیاب مستقل به عملکرد آزمون دهندگان تحت این دو وضعیت، یعنی آزمون کاغذی و آزمون رایانه ای دادند. تیلور و همکاران (۱۹۹۹) به بررسی رابطه بین آشنایی با رایانه و عملکرد در پاسخ دهی به سؤالات زبانی رایانه ای پرداختند. آن ها هیچ رابطه معناداری بین این دو متغیر پیدا نکردند. تحقیقات متعددی درباره این مفهوم صورت گرفته اند. هیچ نتیجه گیری قطعی در مورد رابطه، بین آشنایی با رایانه و عملکرد آزمون وجود ندارد.

علاوه بر قالب آزمون، عوامل دیگری چون جنسیت، سن، رشته تحصیلی و زبان مادری، می توانند در نتایج آزمون ها واریانس نامربوط به سازه تولید کنند. گرانیپیه و کونان<sup>۳</sup> (۲۰۰۷) نقش سن را در یک آزمون بررسی کردند. صالحی و طیبی (۲۰۱۲) نقش جنسیت را در یک آزمون مهم و حیاتی بررسی کردند؛ همین طور کوی، کیم، و بو<sup>۴</sup> (۲۰۰۳) تحقیق هایی در مورد آزمون سی. بی. تی انجام داده اند. اما کمتر محققانی تأثیر عوامل فوق را بر روی نمرات آزمون تافل آی. بی. تی بررسی کرده است. در نتیجه، هدف از این پژوهش، بررسی رابطه متغیرهای متفاوتی مانند جنسیت، سن، زبان مادری و رشته رشته تحصیلی با نتیجه آزمون تافل iBT هست. برای پاسخ به سؤالات خاص این پژوهش، محققان از دو پرسشنامه و یک تافل اینترنتی بهره بردند. صد شرکت کننده از سه دانشگاه مختلف (دانشگاه شریف، تبریز و پیام نور تبریز)

1. Weir

2. Weir, O'Sullivan, Jin Yan Bax

3. Geranpayeh, Kunnan

4. Choi, Kim, Boo

به این پرسشنامه‌ها و همچنین آزمون تافل اینترنتی پاسخ دادند تحقیق حاضر، به‌طور مشخص در جست‌وجوی پاسخ به سؤال‌های زیر است:

- ۱- آیا آشنایی یا عدم آشنایی با رایانه، باعث ایجاد واریانس نامربوط با سازه در آزمون تافل می‌شود؟
- ۲- آیا رشته تحصیلی تأثیری در نمرات آزمون دارد؟
- ۳- دیدگاه شرکت‌کنندگان در آزمون نسبت به رایانه و فناوری چیست؟
- ۴- آیا سن، زبان مادری و جنسیت در نمرات آزمون آی. بی. تی تأثیری دارند؟

## ۲- مرور پیشینه تحقیقات

### ۱-۲- مقدمه

رایانه‌ها وارد حوزه آزمون و سنجش شده‌اند و شرایط را برای نوآوری‌هایی در زمینه طراحی و اجرای آزمون فراهم آورده‌اند. سهولت استفاده از ابزارهای سمعی و بصری مختلف و نمره‌دهی، آزمون‌های رایانه‌ای را به ابزارهای قدرتمندی در این حوزه تبدیل می‌کنند. متغیرهای متفاوتی وجود دارند که می‌توانند نتایج آزمون را تحت تأثیر قرار دهند. جنسیت، سن، آشنایی با رایانه، تجربه استفاده از رایانه، رشته تحصیلی و غیره می‌توانند باعث تفاوت‌هایی در نتایج حاصل شوند. اندازه‌گیری سطح آشنایی با رایانه به فعالان حوزه آزمون، این امکان را می‌دهد تا بفهمند آیا رابطه‌ای بین تجربه استفاده از رایانه و عملکرد آزمون وجود دارد یا خیر. در نتیجه، آشنایی با رایانه به‌مثابه عاملی بیرونی، ممکن است باعث پراکندگی در نمرات آزمون شود. هنگامی که در مورد نتایج آزمون صحبت می‌کنیم، لاجرم بحث روایی مطرح می‌شود. این ویژگی سنجش، ابزاری مهم در جهت ارزیابی نتایج آزمون و تفسیر مبتنی بر آن نمرات است. پر واضح است که روایی و فرایند آزمون، دو روی یک سکه هستند. نتایج آزمون باید با این مفهوم توجیه شوند.

فناوری، حوزه‌های علمی متعددی را تحت تأثیر قرار داده است که از بین آن‌ها، می‌توان به یادگیری زبان و ارزیابی اشاره کرد. آزمون زبان انگلیسی مانند تافل آی. بی. تی بر روی رایانه اجرا می‌شود و می‌توان نسخه‌های دیگر آزمون‌های رایانه‌ای را هم یافت که در نهادها و مراکز دانشگاهی مختلف استفاده می‌شوند. آلدسون (۲۰۰۰) به بررسی مزایا و معایب آزمون‌های رایانه‌ای می‌پردازد. همان‌طور که آلدسون بیان می‌کند، هر نوآوری دارای نقایص خاص خود است و آزمون‌های رایانه‌ای نیز استثنا نیستند. یکی از نقاط ضعف موجود، وابستگی این

آزمون‌ها به برخی روش‌های آزمون مانند آزمون‌های چندگزینه‌ای، آزمون‌های حاوی متن با جای خالی (کلوز تست) و پر کردن جاهای خالی است؛ زیرا نمره‌دهی به این موارد، از طریق ماشین راحت‌تر است.

سواد رایانه‌ای یا آشنایی با رایانه، مفهوم دیگری است که آزمون‌های رایانه‌ای را از آزمون‌های کاغذی متمایز می‌کند. استفاده از موس و کیبورد و خواندن متن از ماینیتور، نیازمند درجه‌ای از آشنایی با رایانه است. یکی دیگر از اشکالات این نوع شیوه آزمون، عدم توانایی ذاتی رایانه‌ها برای هماهنگی با آزمون‌های تشریحی مانند گفتن و نوشتن است. برخلاف آزمون‌های کاغذی مرسوم که باید در زمان مشخصی تحویل داده شوند و گروهی از ناظران برای انجام رویه آزمون موردنیاز هستند، در آزمون‌های رایانه‌ای فرد می‌تواند در هر زمان و مکانی از آن استفاده کند. رایانه چارچوبی را فراهم می‌کند که معلمان می‌توانند در قالب این چارچوب، آزمون‌های استانداردتری برگزار کنند. با استفاده از این چارچوب، طراحان آزمون و معلمان، از اینترنت برای به اشتراک گذاشتن سؤالات خود و کاربرد منابع آنلاین برای آزمون‌هایی کامل بهره می‌برند. اینترنت ابزار مفیدی است که طراحان آزمون‌ها، از طریق آن، می‌توانند آزمون‌های خود را در محل مورد نظر تحویل دهند. ولی این برای آزمون‌های بسیار مهمی مانند تافل اجراشدنی نیست؛ زیرا هکرها و خطرهایی از این دست در اینترنت حضور دارند.

## ۲-۲- روایی<sup>۱</sup> و فرایند روایی سنجی

یکی از مهم‌ترین ویژگی‌های رویه آزمون، تفسیر و استفاده از نمرات آزمون است. روایی، رابطه نزدیکی با مسائل مرتبط با نمرات آزمون دارد (بکمن<sup>۲</sup>، ۱۹۹۰). آری، جیکوبز، سورنسن و واکر (۲۰۱۴) روایی را این‌گونه تعریف کردند: میزانی که ابزار خاصی به اندازه‌گیری آنچه ادعای اندازه‌گیری‌اش را دارد، می‌پردازد. آزمون زبان، عملی اجتماعی است، در نتیجه باید تأثیرات آن بر جامعه و آموزش را مورد توجه قرار داد. به‌منظور اتخاذ تصمیمی مناسب براساس نتایج به‌دست‌آمده، فرد باید به مفاهیم مهم و خاصی توجه داشته باشد. یکی از این مفاهیم محتوای آزمون است. محتوای یک آزمون زبان، نشان‌دهنده ایده و هدف انجام آن آزمون است و باید به اندازه کافی، به هدف آزمون‌گیرنده و توانایی و نیازهای آزمون‌دهندگان

1. Validity

2. Bachman

نزدیک باشد. معیار مهم دیگری که باید به آن اشاره کرد، عملکرد آزمون‌دهندگان است. فرایند روایی‌سنجی عملکرد، با توجه به سطح توانایی آزمون‌دهندگان صورت می‌گیرد. به عبارت دیگر، یک آزمون برای اینکه معتبر باشد، باید سطح توانایی آزمون‌دهندگان را مدنظر قرار دهد. در نتیجه، این‌طور می‌توان گفت که محتوای آزمون و عملکرد آزمون‌دهندگان، مفاهیم کلیدی فرایند روایی‌سنجی هستند. از طرف دیگر، این امر مشهود است که محتوای آزمون و نمرات آزمون، تنها راه‌های روایی‌سنجی یک آزمون نیستند. این تفسیر و کاربرد آتی نمرات آزمون هستند که می‌توانند روایی یک آزمون را مورد قضاوت قرار دهند. محققان علاقه‌مند به فرایند ارزیابی، ساختارهای متفاوتی مانند اضطراب، استعداد و تفکر انتقادی هستند، اما برای ارزیابی این ساختارها مانند مفاهیم فیزیکی ابزارهای مستقیمی وجود ندارد، در نتیجه برای دستیابی به تفسیر و نتیجه مطلوب، محقق باید ساختار را به صورتی نظری تعریف کند.

### ۳-۲- جنبه‌های روایی

شواهد مورد نیاز برای روایی هر فرایند ارزیابی را می‌توان در سه نوع کلی قرار داد: تناسب محتوایی، ارتباط با معیار و معناداری ساختار. محققان در بخش بعدی، به بحث پیرامون این انواع متفاوت می‌پردازند، اما اکنون اشاره به این امر حائز اهمیت است که معمولاً این مقوله‌ها به مثابه مثال‌های متفاوتی از فرایند روایی‌سنجی در نظر گرفته شده‌اند که هر یک از آن‌ها، پیش‌شرط‌های متفاوت و معانی ضمنی مختلفی دارند. اما کاملاً روشن است که فرد در فرایند اعتبارسنجی، باید به جمع‌آوری شواهد از همه این جوانب اقدام کند و آن‌ها را مفاهیم مکملی بداند که از طریق آن‌ها می‌تواند به مطالعه ویژگی‌های کاربرد آزمون و تفسیر آزمون بپردازد و به صورت معناداری به روایی‌سنجی ابزار سنجشی (مانند آزمون) در محیط آموزشی و همچنین در حوزه‌های روان‌شناختی اقدام کند.

همان‌طور که مسیک<sup>۱</sup> (۱۹۹۳) اشاره کرده است، مقوله‌بندی سنتی روایی، شامل روایی‌های محتوایی، ملاکی و سازه‌ای است. وی اعتقاد دارد که این تعریف و گروه‌بندی مفهوم روایی‌سنجی ناقص است؛ زیرا براساس این طبقه‌بندی، نمی‌توان نمرات آزمون‌ها و عواقب اجتماعی آن‌ها را در نظر گرفت. مسیک در ارتباط با روایی، به مفهوم جدیدی دست یافت که به آن، مفهوم یکپارچه روایی می‌گویند. هدف اصلی از معرفی این اصطلاح، توجه به

1. Messick

ویژگی‌های متفاوت فرایند روایی‌سنجی در قالب یک مفهوم منفرد و جامع یعنی «روایی سازه» بود. این ادغام را که شکل جدیدی به روایی سازه می‌دهد، روایی ارزیابی کلی برای اهداف کاربردی و علمی می‌دانند. براساس ایده مسیک درباره روایی و ادغام مفاهیم مختلف در مورد این جزء مهم و حیاتی آزمون‌سازی، نتیجه می‌گیریم که روایی، صرفاً مربوط به کنش و برداشت‌های تفسیری منتج از امتیازات آزمون نیست، بلکه به تفسیر رفتارهای منسجم یا ویژگی‌هایی می‌پردازد که حوزه بررسی ارزشمندی به‌شمار می‌روند.

مسیک در تحقیقاتش در مورد اساس روایی و فرایند روایی‌سنجی، ما را به معنادار بودن نمرات ارجاع می‌دهد. وی اعتقاد دارد که کاربرد کلی نمره، شامل مفاهیم و مضامین کمی و کیفی است. تفسیرهای مختلفی مانند قضاوت‌های بالینی، رفتاری و قضاوت‌های مختلف عملکرد و نیز گزارش‌های رایانه‌ای نمرات حاصل، در این گستره وسیع جای دارند. نمرات، نه تنها محدود به عملکرد یا ویژگی‌های یک فرد خاص هستند، بلکه در این حوزه جدید، نمره شاخص خوبی از سطح انسجام عملکردی و ویژگی‌های گروه‌ها، شرایط محیطی، موضوعات یا موارد آموزشی در زمینه‌های مختلف هستند. نمره را می‌توان برای ارزیابی هر یک از این ویژگی‌ها، از جمله عملکرد و خصیصه‌های فرد یا گروه به‌کار برد. واقعیت مهمی که باید در هنگام ارزیابی مورد توجه قرار بگیرد، در نظر گرفتن روایی به‌صورت نظام‌مند و مشابه با هر مسئله‌سنجی پایه‌ای مانند قابلیت اعتماد (پایایی)<sup>۱</sup>، قیاس‌پذیری و غیره است. همان‌طور که مسیک اشاره کرده است، روایی و سایر ویژگی‌های سنجشی، ابزارهای ارزیابی نیستند، بلکه ارزش‌های اجتماعی هستند که تأثیرات مشخصی بر جهان خارج از محیط سنجش دارند. در نتیجه، نمی‌توان در مورد روایی ارزیابی بر اساس ضریب همبستگی قضاوت خاصی کرد یا نظرهای تخصصی داد. گفته‌های کرونباخ<sup>۲</sup> (۱۹۷۱) بیان‌کننده این نکته است که روایی‌سنجی رابطه‌ای با خود آزمون یا ابزارهای مشاهده‌ای ندارد، بلکه مستقیماً به نتایج نمرات آزمون مربوط می‌شود.

تا اینجا، هدف اصلی بحث در مورد روایی‌سنجی و مؤلفه‌های حیاتی این فرایند، نمره و تفسیر آن بود. همان‌طور که مسیک اشاره کرده است، فرد باید به برخی سؤالات حیاتی در مورد فرایند روایی‌سنجی و خود مفهوم روایی پاسخ دهد: اولاً، فرد باید در مورد منطقی و

1. Reliability

2. Cronbach

مناسب بودن فرایند تحقیق و شیوه تحلیل و قرار دادن متغیرها در ترتیب مناسب اطمینان پیدا کند؛ مرحله بعدی شامل بازیافت اطلاعات از دست‌رفته است؛ سپس، فرد باید به ارزیابی شیوه موجود پردازد تا مشخص کند آیا آن روش منبع ناروایی می‌تواند باشد یا خیر. این مفهوم «واحد» دارای جنبه متفاوتی از جنبه محتوایی است که رابطه محتوایی، بازنمایی و کیفیت فنی است (امبرتسون<sup>۱</sup>، ۱۹۸۳). جنبه دیگر روایی سازه، مرتبط با جنبه سازه‌ای است که لوینگر<sup>۲</sup> (۱۹۵۷) آن را مرتبط با تناسب ساختار امتیازبندی با حوزه سازه می‌داند. کمپل و فیسک<sup>۳</sup> (۱۹۵۹) در مورد بعد دیگر روایی سازه صحبت کردند که جنبه‌ای بیرونی است و شامل شواهد همگرایی و افتراقی منتج از مقایسه‌های چندخصلتی - چندروشی است. ویژگی دیگری که باید در این زمینه به آن اشاره کنیم، همان جنبه عمومیت‌بخشی است که به بررسی میزانی می‌پردازد که ویژگی‌های نمرات و تفسیر را می‌توان در بین گروه‌ها، محیط‌ها و کارهای مختلف تعمیم داد (کوک و کمپل<sup>۴</sup>، ۱۹۷۹؛ شولمن<sup>۵</sup>، ۱۹۷۰). مسیک (۱۹۸۰) در مورد جنبه نتیجه محوری صحبت کرده است که مرتبط با تفاسیر مبتنی بر اساس نمرات آزمون هستند.

در این بخش، بهتر است به بحث پیرامون انواع روایی‌های مرسوم و محدودیت آن‌ها در فرایند ارزیابی و سنجش پردازیم. همان‌طور که قبلاً اشاره شد، به‌طور سنتی، مفهوم روایی، معمولاً به انواع مجزا و قابل جایگزینی تقسیم می‌شد: روایی محتوایی، ملاکی و سازه‌ای. همچنین بیان شده است که این انواع مجزا و جنبه‌های روایی، هنگام بحث پیرامون تفسیر نمرات برای کاربردهای آتی، مناسب نیستند. در وهله اول، به تعریف جنبه‌هایی می‌پردازیم که انجمن روان‌شناسی آمریکا معرفی کرده است (۱۹۵۴ و ۱۹۶۶).

روایی محتوایی با این امر ارزیابی می‌شود که محتوای آزمون، تا چه اندازه‌ای نشان‌دهنده نمونه طبقه موقعیت یا موضوعی است که باید در آن مورد به نتیجه رسید.

روایی ملاکی با مقایسه نمرات آزمون‌ها، به‌صورت تخمینی یا هم‌زمان، با یک یا چند متغیر بیرونی (با نام قواعد) ارزیابی می‌شود که معیار مستقیمی از آن ویژگی یا رفتار مدنظر را عرضه می‌کند.

1. Embretson
2. Loevinger
3. Campbell, Fiske
4. Cook, T. D., & Campbell
5. Shulman



روایی سازه با ارزیابی ویژگی‌هایی ارزیابی می‌شود که یک آزمون باید به سنجش آن‌ها پردازد؛ یعنی میزان توضیحی که مفاهیم توضیحی معیار یا سازه‌ها از عملکرد یک آزمون می‌دهند.

در طول تاریخچه آزمون و سنجش عملکرد، این مفاهیم همچنان به کار رفته‌اند، ولی دستخوش تغییر و تحولات چشمگیری بوده‌اند. لنون<sup>۱</sup> (۱۹۵۶) یکی از این تغییرات را در قالب مفهوم روایی محتوایی معرفی کرد. وی تمرکز مفهوم روایی محتوایی را از نمونه‌گیری موقعیت‌ها یا موضوع مورد بررسی به سمت نمونه‌گیری از رفتارها و فرایندهای حاکم تغییر داد. مسیک (۱۹۷۵، ۱۹۸۰) همچنین مفهوم روایی سازه را به گونه‌ای بسط داد که دربرگیرنده هر دو مفهوم محتوا و ملاک باشد. در اینجا بهتر است به بحث پیرامون ویژگی‌های روایی محتوایی بپردازیم. این مفهوم، منتج از تعاریف مرسوم است که روایی محتوایی با محتوای آزمون و محتوای حوزه رفتاری و پوشش مؤلفه‌ها یا کارهای خاصی متناسب است. در روایی، فرد نمی‌تواند در مورد فرایند پاسخ‌گویی، ساختار آزمون، تفاوت در عملکرد و حتی نتایج اجتماعی به‌کارگیری فرایند آزمون اطلاعات کافی کسب کند. در نتیجه، این ادعا مطرح می‌شود که جنبه روایی - روایی محتوایی - مربوط به نتایج آزمون، تفسیر نتایج یا عملکرد آزمون‌دهندگان نیست، بلکه حوزه تحقیقاتی مهم در این نوع روایی، مطالعه تناسب محتوایی است. جنبه دیگر روایی که باید به آن اشاره کنیم، روایی ملاکی است.

همان‌طور که در بالا اشاره شد، در قالب روایی محتوایی، دغدغه اصلی معطوف به محتوای آزمون و تناسب آن با حوزه مدنظر است؛ از طرف دیگر، در روایی ملاکی، دغدغه اصلی معطوف به نمرات آزمون و ضرایب همبستگی و یا رگرسیون‌های آن‌ها با نمرات ملاکی است. در اینجا، ویژگی مهم صرفاً الگوی روابط بین نمرات نیست،

بلکه روابط گزینشی در آزمون‌های عملی خاص است که مهم شمرده می‌شوند. براساس تفسیر امتیازات آزمون، به کمک روایی ملاکی، محققانی مانند اشمیت، هانتز، پیرلمن، هیرش، ساکت اشمیت و زیدیک<sup>۲</sup> (۱۹۸۵) در مورد تعمیم‌پذیری نمره آزمون، به نسبت تفسیر نمره آزمون مطالبی نگاشته‌اند. روایی ملاکی مرتبط با کاربرد ضرایب همبستگی یا رگرسیون آزمون، به‌مثابه ابزاری برای پیش‌بینی است. باین‌حال، مسئله مهمی در زمان استفاده از این نوع روایی وجود دارد. در روایی ملاکی، پارادوکسی وجود دارد که به آن پارادوکس ذاتی یا درونی

1. Lennon

2. Schmidt, Hunter, Pearlman, Hirsh, Sackett, Schmitt, Zedeck

می‌گویند. این به آن معناست که نمرات به‌مثابه برون‌داد این روایی، باید به شیوه‌ای مشابه با یک آزمون ارزیابی بشوند ولی نباید از آن‌ها به‌صورت استانداردهایی در جهت ارزیابی این امتیازات استفاده کرد. با این حال، مؤلفان و کارشناسان حوزه ارزیابی و آزمون به نتیجه‌گیری خاصی، برای رفع این مشکل دست یافته‌اند. آن‌ها به ارزیابی هر دو نظریه سازه‌ای معیار و آزمون‌محور، در قالب حوزه ملاکی خاصی پرداختند. در اینجا، روایی سازه‌ای مورد نظر ماست. روایی سازه‌ای به‌مثابه جنبه مجزایی از فرایند روایی‌سنجی، یک روایی جامع است که تفسیر امتیاز آزمون و معنای آن را در نظر می‌گیرد. هم‌راستا با روایی سازه، کرونباخ و میهل<sup>۱</sup> (۱۹۵۵) اشاره کردند که در این نوع روایی، نمره را معادل سازه نمی‌گیرند و در نتیجه، نمی‌توان به تعریف مفهوم سازه پرداخت. نمره و معیار به‌صورت مجموعه‌ای از شاخص‌های مرتبط با سازه در روایی سازه‌ای در نظر گرفته می‌شوند.

#### ۴-۲- روش‌های مختلف سنجش اعتبار سازه

در ادبیات روایی، چندین روش وجود دارد که از آن‌ها به مختصر یاد می‌کنیم.

##### ۱- واگرایی<sup>۲</sup> و همگرایی<sup>۳</sup>

در این شیوه ماتریکس چندشیوه‌ای - چندخصیصه‌ای (multi-trait multi-method) بسیار اهمیت دارد که کمپل و فسبیک در سال ۱۹۵۹ مطرح کرده‌اند. درحوزه آزمون‌سازی زبان، از این شیوه کمتر استفاده شده است. ولی در ایران از تحقیق رضایی و صالحی (۲۰۰۸) می‌توان نام برد که در مورد یک آزمون بسندگی<sup>۴</sup> زبان دانشگاه تهران انجام گرفت. در این شیوه، از ضرایب همبستگی استفاده می‌شود و دو یا چند خصیصه را می‌توان با دو یا چند شیوه مورد مقایسه قرارداد؛ برای مثال دانش واژه و دانش دستور را می‌توان به‌صورت دو خصیصه در نظر گرفت که با شیوه‌های چهارگزینه‌ای و پرکردن جاهای خالی، می‌توان آن‌ها را سنجید. برای مثال‌های مختلف و کاربرد این شیوه می‌توان به صالحی (۲۰۱۲) نیز مراجعه کرد.

1. Cronbach, Meehl

2. Divergence

3. Convergence

4. Proficiency

## ۲- بررسی ساختار درونی آزمون‌ها:

در این شیوه، می‌توان از تحلیل عامل اکتشافی و تأییدی برای بررسی ساختارهای موجود در ابزارها استفاده کرد. صالحی (۲۰۱۱) از شیوه تحلیل عامل اکتشافی، برای بررسی ساختار درونی<sup>۱</sup> یک آزمون بسندگی زبان استفاده کرد.

## ۳- بررسی فرایندهای پاسخ دادن به سؤالات :

این شیوه، امروزه در زمره تحقیق کیفی قرار دارد و از اهمیت بسزایی برخوردار است. شاید کوهن<sup>۲</sup> (۱۹۸۴) اولین محققی بود که ما را به اهمیت بررسی اعتبارسازی یک آزمون از طریق بررسی فرایندهای پاسخ‌دهی آگاه کرد. از منظر کوهن، افراد ممکن است بنا به دلایل غلط، پاسخ‌های صحیح را انتخاب کنند یا برعکس. در حوزه آزمون‌سازی زبان، صالحی و باقری (۲۰۱۳) از این شیوه استفاده کردند تا تفاوت دو آزمون کلوز و سی‌تست (c-test) را بسنجند.

## ۴- بررسی خصوصیات فردی و ویژگی‌های جمعیتی:

عوامل گوناگونی چون جنسیت، رشته تحصیلی، آشنایی با رایانه می‌توانند در نمرات آزمون تغییراتی به وجود بیاورند که از آن‌ها با نام واریانس نامربوط به سازه یاد می‌شود. در این زمینه، تحقیقات گسترده‌ای انجام شده است؛ برای مثال گرانپایه و کونان (۲۰۰۹) تحقیقی در مورد رابطه بین نمرات یک آزمون و سن زبان‌آموزان انجام دادند. نتیجه تحقیق نشان داد که بعضی از سؤالات، نسبت به متغیر سن تبعیض قائل شده‌اند. صالحی و طیبی (۲۰۱۲) نیز تحقیقی درباره تأثیر جنسیت بر نمرات آزمون انجام دادند. نتیجه تحقیق این محققان نشان داد که جنسیت، تأثیر چندانی در نمرات آزمون ندارد.

## ۵-۲- آشنایی با رایانه

تیلر، کرش، جامیسن، و ایگور (۱۹۹۹) به اهمیت آشنایی با رایانه اشاره کردند. گیسلر و هاریج<sup>۳</sup> (۱۹۹۳) از دانشجویان خود خواستند تا در مورد استفاده از رایانه صحبت کنند. بگویند چگونه و چقدر از رایانه استفاده کردند. لی<sup>۴</sup> (۱۹۸۶) در تحقیق خود نشان داد که آشنا نبودن

1. Internal Structure

2. Cohen

3. Geissler, & Horridge

4. Lee

با رایانه، موجب اضطراب در آزمون می‌شد. از نظر دورندل و لایت بادی<sup>۱</sup> (۱۹۹۳) آشنایی با رایانه، به معنی دسترسی به رایانه در منزل، مدرسه و یا هر جای دیگر بود. لی از پرسشنامه آشنایی با رایانه استفاده کرد که شامل پنج سؤال از جمله تعداد دروس رایانه بود. در اینجا باید از ابزار دیگری نیز نام برد که مقیاس نگرش به رایانه است و لیاد و گراسارد<sup>۲</sup> (۱۹۸۴) طراحی کرده بودند. این مقیاس از سه عامل تشکیل شده بود: اعتماد به نفس، اضطراب و نگرش مثبت. کالارینا و والاس (۲۰۰۲) آزمون کاغذی و رایانه‌ای تافل را بررسی کردند. آن‌ها عواملی چون آشنایی با محتوا، آشنایی با رایانه، رقابت و جنسیت را در نظر گرفتند. دانشجویان دوره کارشناسی به تعداد ۱۰۵ نفر به دو گروه آزمون رایانه و آزمون کاغذی تقسیم شدند. نتایج تحلیل واریانس نشان داد که جنسیت، رقابت، و آشنایی با رایانه، ربطی به تفاوت عملکرد نداشت. اودو (۲۰۱۲) نتیجه گرفت که آشنایی با رایانه، تأثیر زیادی در عملکرد آزمون رایانه‌ای نداشت.

### ۳- روش

#### ۳-۱- شرکت‌کنندگان

در این پژوهش، محققان شرکت‌کنندگان را از دانشگاه‌ها و رشته‌های متفاوت انتخاب کردند تا رشته تحصیلی شرکت‌کنندگانی را بررسی کنند که از دانشگاه‌های مختلف آمده بودند. شرکت‌کنندگان، دانشجویانی از دانشگاه‌های مختلف شریف (N=30)، تبریز (N=30) و دانشگاه پیام نور تبریز (N=40) بودند. سن، جنسیت، رشته تحصیلی و زبان مادری، چهار متغیری بودند که براساس آن‌ها، محققان، شرکت‌کنندگان را دسته‌بندی کردند. حوزه مطالعاتی مشارکت‌کنندگان باعث شد که در نهایت، دو گروه مشخص شوند. در نتیجه، محققان شرکت‌کنندگان را در دو گروه مهندسی و علوم انسانی قرار دادند. در ادامه، اطلاعاتی در مورد ویژگی‌های متفاوت این دانشجویان آمده است.

گروه اول از مشارکت‌کنندگان شامل دانشجویانی بود که در حوزه علوم مهندسی و حوزه‌های مرتبط با ریاضیات تحصیل می‌کردند. مدارک این دانشجویان در حوزه ریاضیات محض (۱)، مهندسی صنعتی (۳)، مهندسی برق (۳)، علوم رایانه (۱)، علوم ماده (۳)، فیزیک

1. Durndell, Lightbody

2. Loyd, Gressard

(۳)، معماری (۳)، ساخت و ساز (۲۳)، مهندسی زیست پزشکی (۱) (N=50) بود. گروه دیگر از مشارکت کنندگان شامل دانشجویان علوم انسانی بود. این مشارکت کنندگان از رشته‌های تحصیلی متفاوتی مانند مدیریت (۷)، علوم اقتصاد (۹)، TEFL (۴)، حقوق (۱۵) و حسابداری (۱۵) (N=50) بودند.

متغیر دیگری که نقش مهمی در این پژوهش داشت، سن شرکت کنندگان بود. پیش‌تر اشاره شد که شرکت کنندگان شامل دانشجویان بودند. محدوده سنی شرکت کنندگان از ۲۰ تا ۳۰ بود. سه گروه سنی در این مطالعه تعریف شد (۲۰-۲۵، ۲۶-۲۹ و ۳۰-۳۵). همان‌طور که قبلاً اشاره شد، بیشتر شرکت کنندگان از دانشگاه تبریز و پیام نور تبریز بودند؛ در نتیجه، بیشتر شرکت کنندگان به زبان آذری (۷۰ درصد) صحبت می‌کردند، ۲۵ درصد فارسی زبان و ۵ درصد از آن‌ها هم کرد زبان بودند.

### ۳-۲-۳- ابزارها

#### ۳-۲-۱- تافل iBT

نگرش نسبت به فناوری و آشنایی با رایانه، دو مفهوم مهم و اساسی بودند که محققان در پژوهش حاضر، به بررسی آن‌ها پرداختند. تافل iBT ابزار مهمی است که محققان از طریق آن، به بررسی تأثیر متغیرهای مختلف اقدام کردند. پاسخ‌گویی به تافل iBT حدوداً ۳ ساعت و نیم تا ۴ ساعت طول می‌کشد و به چهار بخش تقسیم می‌شود: بخش اول خواندن است که حدوداً ۶۰ تا ۹۰ دقیقه طول می‌کشد؛ بخش دوم شنیداری است که حدوداً ۶۰ تا ۹۰ دقیقه طول می‌کشد. بعد از این مرحله ۱۰ دقیقه استراحت وجود داشت؛ بخش سوم مکالمه است که کوتاه‌ترین بخش است و حدوداً ۲۰ دقیقه طول می‌کشد؛ بخش نهایی، نوشتار است که حدوداً ۶۰ دقیقه برای آن در نظر گرفته شده بود. در این پژوهش، یک آزمون تافل iBT عرضه شده است. مشارکت کنندگان از لینک <http://www.testden.com/challenge/free-toefl.asp> استفاده کردند که محققان از طریق ایمیل، برای آن‌ها ارسال کردند. در این وب‌سایت، شرکت کنندگان نام و نشانی ایمیل‌شان را وارد می‌کردند و در نهایت، محل سکونتشان را تعیین می‌کردند. مرحله دیگر، آزمون اصلی به همراه راهنمایی‌های مربوط به آن بود. هر مشارکت کننده سه فرصت برای انجام این آزمون داشت. مشارکت کنندگان امکان پاسخ به یک بخش و ذخیره آن را داشتند تا در مرحله بعدی به بخش دیگر بروند و آن را ذخیره کنند. با پاسخ دادن به آخرین سؤال، فرایند آزمون به پایان می‌رسید و نمرات شرکت کنندگان اعلام می‌شد.

## ۲-۳- پرسشنامه‌ها

در این تحقیق محققان از دو پرسشنامه بهره بردند که ویژگی‌های آن‌ها همراه با خصایص آماری، توضیح داده می‌شود.

اولین پرسشنامه، «پرسشنامه آشنایی با رایانه» (CFQ<sup>۱</sup>) برای اندازه‌گیری سطح آشنایی شرکت‌کنندگان با رایانه به کار گرفته شد. این پرسشنامه حاوی ۱۰ پرسش بود که چندین مؤسسه مانند شورای تحقیقات آموزشی استرالیا (ACER)، نهاد ملی سنجش آموزشی هلند (CITO)، خدمات آزمون‌گیری آموزشی (ETS ایالات متحده) و نهاد ملی تحقیقات آموزشی (NIER، ژاپن) از آن استفاده کرده بودند. همان‌طور که کیرش، جمیسون، تیلور و ایگنور (۱۹۹۸) اشاره کرده‌اند همبستگی ۱۱ مؤلفه (پرسشنامه اصلی بدون تغییرات) با نمره کل از ۰،۵۹ تا ۰،۸۲ متغیر بود. پایایی آلفا مبتنی بر ۱۱ پرسش، معادل ۰،۹۱ و پایایی نیمه‌معادل<sup>۲</sup> برابر با ۰،۹۳ بود. در CFQ، پاسخ‌دهندگان با سؤالات متفاوتی مواجه می‌شدند از جمله: ۱- چند وقت یک‌بار در این مکان‌ها، رایانه‌ای در اختیار شما قرار می‌گیرد؟ و یا ۲- اگر خودتان را با افراد ۱۵ ساله دیگر مقایسه کنید، به توانایی‌تان در استفاده از رایانه چه نمره‌ای می‌دهید؟ این پرسشنامه دویخشی بود: بخش اول شامل ۴ سؤال و مرتبط با دسترسی به رایانه‌ها، راحتی استفاده و میزان استفاده از رایانه بود. بخش بعدی با ۶ سؤال جزئیات بیشتری را در برمی‌گرفت و سؤالاتی در مورد ویژگی‌های خاص کاربری رایانه مطرح می‌کرد. محققان این پرسشنامه را به فارسی ترجمه کردند.

پرسشنامه دیگر درباره رایانه و نگرش به رایانه (CAQ) بود که در آن، شرکت‌کنندگان دیدگاه‌ها و عقاید خود را درباره رایانه شرح می‌دادند. CAQ را زینک و کریستنسن (۱۹۹۶) و سایر محققان در مرکز فناوری آموزشی تگزاس طراحی کرده‌اند. همبستگی ۰،۸۰ به‌مثابه شاهدهی برای حداقل یک نوع پایایی پیشنهاد شد؛ با این حال، استانداردهایی در محدوده ۰،۵ تا ۰،۹ براساس کاربرد موردنظر و محیط کاربرد ابزار عرضه شده‌اند. این پرسشنامه به زبان‌های ژاپنی و اسپانیایی ترجمه شده است. جمع‌آوری داده در قالب CAQ خود گزارشگری است. پرسش‌های متفاوتی در این پرسشنامه گنجانده شد و پاسخ‌دهندگان از مقیاس ۵ امتیازی برای نشان‌دادن دیدگاه خود در مورد فناوری، به‌خصوص رایانه، بهره بردند. CAQ دارای سه بخش

1. Computer Familiarity Questionnaire

2. Split Half

با ۲۴ مؤلفه است. ۱۱ پرسش در بخش اول، ۳ پرسش در بخش دوم و ۱۰ پرسش در بخش آخر آمده است. در مورد اعتبار سازه این پرسشنامه، باید گفت که اعمال تجزیه و تحلیل عامل، شش زیرمجموعه معنی دار تولید کرده است.

نمونه‌ای از CAQ:

از انجام کارهای مختلف روی رایانه لذت می‌برم:

- ۱- بسیار مخالفم؛ ۲- مخالفم؛ ۳- تصمیمی نگرفته‌ام؛ ۴- موافقم؛ ۵- بسیار موافقم.
- نسخه ترجمه شده این پرسشنامه در اختیار شرکت کنندگان نیز قرار گرفت.

### ۳-۳- توزیع پرسشنامه‌ها

رویه جمع‌آوری اطلاعات در پژوهش حاضر، فرایندی سه‌مرحله‌ای بود که محققان طی این سه مرحله، به جمع‌آوری داده‌های موردنظر از یک‌صد شرکت‌کننده پرداختند؛ اولاً، پرسشنامه‌های مدنظر انتخاب و بخش‌های مختلفی از آن‌ها اصلاح و مؤلفه‌ها/پرسش‌های نهایی به فارسی ترجمه شدند. بعد از انجام این فرایندها، محققان از «سروی مانکی» (Survey Monkey) بهره بردند که مرکز جمع‌آوری داده‌های وب‌محور است که پرسشنامه‌ها را به صورت دیجیتالی درمی‌آورد. این وب‌سایت، لینک‌های متفاوتی برای هر یک از پرسشنامه‌ها عرضه کرد. این لینک‌ها برای مشارکت‌کنندگان ارسال شد. در این پرسشنامه‌ها، محققان سؤالاتی در مورد سن، جنسیت، رشته تحصیلی و زبان مادری مشارکت‌کنندگان مطرح کردند.

پرسشنامه اول (آگاهی از رایانه):

<https://www.surveymonkey.com/t/YHJQFTS>

پرسشنامه دوم (دیدگاه نسبت به رایانه):

<https://www.surveymonkey.com/t/RVNPWN3>

### انجام آزمون تافل iBT

مرحله بعدی فرایند جمع‌آوری داده، با تعیین نمره تافل iBT مرتبط بود. تافل با لینک <http://www.testden.com/challenge/free-toefl.asp> به عنوان آزمون مدنظر انتخاب شد. این آزمون شامل چهار بخش خواندن، شنیدار، گفتار و نوشتار بود و هر بخش، راهنمایی‌های

خاص خودش را داشت. شرکت‌کنندگان می‌توانستند آزمون را متوقف کنند و بعداً آن را ادامه دهند؛ بنابراین لزومی نداشت به کل آزمون‌ها در یک مرحله پاسخ بدهند. از آنجا که همه شرکت‌کنندگان دانشجو بودند، تقریباً همه آن‌ها تمایل به بررسی دانش زبان انگلیسی خود داشتند. در نتیجه، آن‌ها انگیزه کافی برای پاسخ دادن به این آزمون داشتند. بعد از پایان این آزمون، آن‌ها نمرات خود را از طریق بخش خاصی از پرسشنامه دریافت کردند.

#### ۴-۳- تحلیل داده‌ها

سه منبع داده از پرسشنامه‌ها و آزمون تافل اینترنتی وارد نرم‌افزار SPSS شدند تا تحلیل کمی صورت گیرد. در این بخش، در مورد شیوه تحلیل داده‌های جمع‌آوری شده صحبت می‌کنیم. در سؤال اول تحقیق، محققان معیار آشنایی با رایانه و رابطه احتمالی آن با نتایج آزمون را تحلیل کردند. برای پاسخ به این سؤال، محققان نمرات تافل را با نمرات به‌دست‌آمده از پرسشنامه آشنایی با رایانه مقایسه کردند؛ سؤال دیگر مرتبط با بررسی رابطه احتمالی بین رشته‌های تحصیلی مشارکت‌کنندگان و نتایج تافل آی بی تی شان بود. از آنجایی که بیش از دو رشته وجود داشت (۱۴ رشته)، محققان نتوانستند از آزمون t بهره ببرند. در نتیجه، بهترین راه برای تحلیل این سؤال استفاده از تحلیل واریانس یک‌سویه بود؛ سؤال سوم مرتبط با دیدگاه موجود نسبت به فناوری و رایانه بود. در اینجا، محقق خواهان اطلاعات بیشتری در مورد دیدگاه مشارکت‌کنندگان در مورد فناوری و رایانه به‌مثابه ویژگی مهم فناوری بود؛ این سؤال هم نیازمند رویه آماری توصیفی بود؛ در آخرین سؤال، سه متغیر (سن، جنسیت و زبان مادری) و یک متغیر وابسته (نتایج تافل iBT) وجود داشت. برای پاسخ دادن به این سؤال، بهترین رویه آماری آزمون تحلیل واریانس سه‌سویه بود.

#### ۴. نتایج

##### سؤال ۱:

اولین سؤال این بود که «آیا آشنایی یا عدم آشنایی با رایانه، باعث ایجاد واریانس نامربوط با سازه (CIV) در آزمون تافل می‌شود؟». در سؤال اول که مرتبط با دسترسی شرکت‌کنندگان به رایانه بود، ۹۵ شرکت‌کننده اذعان کردند که تقریباً هرروز به رایانه دسترسی دارند؛ سؤال دیگر مرتبط با دسترسی شرکت‌کنندگان در دانشگاه به رایانه بود. ۶۵ شرکت‌کننده اشاره کردند که چند بار در هفته، به رایانه دسترسی دارند. بخش دیگر این سؤال مرتبط با دسترسی به



کتابخانه دارای امکانات رایانه‌ای بود و ۴۲ شرکت‌کننده اشاره کردند که در هفته به چند بار از یارانه استفاده می‌کردند. یکی از مؤلفه‌ها (پرسش‌های) گنجانده شده در این پرسشنامه، مرتبط با حس شرکت‌کنندگان در زمان استفاده از رایانه بود. در مورد این سؤال، ۸۲ شرکت‌کننده اعلام کردند که در زمان استفاده از رایانه مشکلی ندارند. سؤال دیگر مهم بود و در مورد حس شرکت‌کنندگان در مورد پاسخ دادن به آزمون با رایانه بود. ۹۰ درصد از شرکت‌کنندگان اشاره کردند که در زمان انجام آزمون‌های رایانه‌ای، حس خیلی خوبی دارند. این امر نشان‌دهنده میزان آشنایی شرکت‌کنندگان با رایانه است؛ زیرا آن‌ها حس می‌کردند که اعتماد به نفس کافی، برای انجام آزمون رایانه‌ای دارند؛ به علاوه، نمرات تافل آی. بی. تی با نمرات پرسشنامه آشنایی با رایانه مقایسه شد. ارتباط این دو متغیر منفی بود ( $r = .08$ ).

## سؤال ۲

در این سؤال، محققان خواهان پاسخ‌گویی به این سؤال بودند که آیا رشته تحصیلی، تأثیری در نمره‌های آزمون تافل دارد. ۱۰۰ شرکت‌کننده در این مطالعه حضور داشتند و ۵۰ نفر از آن‌ها در بخش علوم انسانی و ۵۰ نفر دیگر در حوزه مهندسی و ریاضیات تحصیل می‌کردند. در مرحله پردازش داده، افراد براساس رشته تحصیلی خود، به دو گروه تقسیم شدند؛ سپس برای هریک از این گروه‌ها محققان آزمون آنالیز واریانس یک‌سویه انجام دادند تا به بررسی تأثیر متغیر مستقل بر متغیر وابسته موجود پردازند (امتیازات تافل iBT). شایان ذکر است که آزمون نرمال بودن داده‌ها انجام شد و شرایط برای انجام آزمون پارامتریک فراهم بود. یک تحلیل بین گروهی یک‌سویه بر روی واریانس، برای بررسی رابطه بین رشته مطالعاتی مشارکت‌کنندگان (رشته‌های انسانی) و امتیازات تافل iBT صورت گرفت. پنج گروه از شرکت‌کنندگان براساس رشته تحصیلی آن‌ها (مدیریت، علوم اقتصادی، آموزش زبان انگلیسی، حقوق، حسابداری) قابل تعریف بود. هیچ تفاوت معناداری در سطح  $P < 0.05$  در نمرات تافل آی بی تی برای پنج گروه وجود نداشت:  $F(4, 45) = 2.28, p = .075$ . این نتایج در جدول ۱ قابل مشاهده است.

جدول ۱. تجزیه و تحلیل واریانس یک‌سویه برای دانشجویان مهندسی

تجزیه و تحلیل واریانس					
نمرات تافل ای بی تی					
	مجموع توانها	درجه آزادی	میانگین توان	F	سطح معنی داری
بین گروه ها	1985.551	4	496.388	2.286	.075
میان گروه ها	9772.229	45	217.161		
مجموع	11757.780	49			

علاوه بر این، یک تحلیل بیناگروهی یک‌سویه برای بررسی رابطه رشته تحصیلی مشارکت‌کنندگان (رشته مهندسی) و نمرات تافل iBT آن‌ها صورت گرفت. ۹ گروه از مشارکت‌کنندگان براساس رشته تحصیلی آن‌ها قابل تعریف بودند: ریاضیات محض: ۱؛ مهندسی صنایع: ۳؛ مهندسی برق: ۳؛ علوم رایانه: ۱؛ علوم مواد: ۳؛ فیزیک: ۳؛ معماری: ۱۲؛ مهندسی عمران: ۲۳؛ مهندسی زیست پزشکی: ۱). هیچ تفاوت معناداری در سطح  $P < 0.05$  در امتیازات تافل iBT برای ۹ گروه وجود نداشت  $F(8, 41) = .692, p = .697$ .

جدول ۲: تحلیل واریانس برای دانشجویان علوم انسان

تجزیه و تحلیل واریانس					
نمرات تافل ای بی تی					
	مجموع توانها	درجه آزادی	میانگین توان	F	سطح معنی داری
بین گروه ها	1680.261	8	210.033	.692	.697
درون گروه ها	12452.859	41	303.728		
مجموع	14133.120	49			

### سؤال ۳

دیدگاه شرکت کنندگان در مورد رایانه، ویژگی مهم فناوری موضوع اصلی سؤال ۳ تحقیق حاضر است. برای پاسخ دادن به این سؤال، بهترین رویه آماری آمار توصیفی و تعیین فراوانی پاسخ های شرکت کنندگان بود. ۹۲ نفر از شرکت کنندگان اعلام کردند که از استفاده از رایانه، لذت می برند. این درصد بالا نشانه خوبی از دیدگاه مثبت آن ها به رایانه به مثابه یکی از ویژگی های برجسته این فناوری بود؛ سؤال بعدی که باید در قالب رویه آماری کدبندی معکوس شود، این است که «من از استفاده از رایانه خسته شده ام». ۷۵ درصد از شرکت کنندگان با این سؤال مخالف بودند و ۲۰ نفر از آن ها اعلام کردند که با این نظر، موافق هستند؛ سومین مؤلفه (پرسش) در این پرسشنامه، در مورد رابطه بین یافتن شغل و توانایی استفاده از رایانه بود. ۶۵ شرکت کننده بر این باور بودند که اطلاع از شیوه استفاده از رایانه به معنی یافتن شغل مناسبی بود؛ در مورد مؤلفه چهارم، شرکت کنندگان به بیان شیوه استفاده از رایانه می پرداختند: ۶۰ درصد از آن ها اعلام کردند که در زمان استفاده از رایانه تمرکز می کنند؛ بازی موضوع سؤال پنجم بود: ۶۱ مشارکت کننده اعلام کردند که خیلی از بازی های رایانه ای لذت می برند.

### سؤال ۴

سؤال آخر یک سؤال چندوجهی بود. در این سؤال تأثیر سه متغیر مستقل جنسیت، سن و زبان مادری بر نمره های تافل سنجیده شد. در مورد این سؤال، رویه آماری مناسب انجام آزمون ANOVA سه طرفه بود. شایان ذکر است که توزیع نمره ها نرمال بود؛ بنابراین شرایط برای انجام تجزیه و تحلیل واریانس سه طرفه مهیا بود.

جدول ۴. نتایج آزمون تجزیه تحلیل واریانس سه‌سویه برای بررسی تأثیر سه متغیر جنسیت، سن، زبان مادری در نمرات تافل

آزمون‌های تأثیرات بین موضوعی <sup>۱</sup>					
متغیر وابسته: نمرات تافل ای بی تی					
منبع	نوع سوم مجموع توانها <sup>۲</sup>	درجه آزادی	میانگین توان <sup>۳</sup>	F	سطح معنی‌داری
مدل تصحیح شده	3459.613 <sup>a</sup>	11	314.510	1.184	.310
عرض از مقطع <sup>۴</sup>	108427.452	1	108427.452	408.219	.000
گروه سنی	.741	2	.371	.001	.999
جنسیت	23.920	1	23.920	.090	.765
زبان مادری	1359.163	2	679.582	2.559	.083
سن * جنسیت	2.026	1	2.026	.008	.931
گروه سنی * زبان مادری	731.114	2	365.557	1.376	.258
جنسیت * زبان مادری	756.295	2	378.148	1.424	.246
گروه سنی * جنسیت * زبان مادری	392.353	1	392.353	1.477	.227
خطا	23373.777	88	265.611		
مجموع	540779.000	100			
مجموع تصحیح شده	26833.390	99			
a. (0.020 = آر مجذور تنظیم شده) = 0.129 آر مجذور					
b. = 0.05 محاسبه شده از طریق آلفا					

1. tests of between-subjects effects

2. sum of squares

3. mean square

4. intercept

جدول شماره ۴، حاوی نتایج آزمون تحلیل واریانس سه سویه هست که مرتبط با رابطه بین گروه سنی، جنسیت و زبان مادری و نمره تافل iBT هستند. اولین متغیر سن شرکت کنندگان است که به سه گروه تقسیم شدند: ۲۰-۲۵، ۲۶-۳۰ و ۳۰-۳۵. براساس نتایج (  $F(2-99) = .001$  )، هیچ رابطه معناداری بین سن شرکت کنندگان و نمرات آزمون وجود ندارد؛ متغیر دیگر در این بخش که باید به آن اشاره کرد، جنسیت شرکت کنندگان است. ۵۴ شرکت کننده مذکر و ۴۶ شرکت کننده مؤنث بودند. سطح معنادار این متغیر (  $F(1-99) = .090$  ,  $P = .765$  ) نشان داد که رابطه معناداری بین جنسیت آزمون دهندگان و نتایج آزمون آن‌ها وجود نداشت. آخرین متغیر در این آزمون آماری، زبان مادری مشارکت کنندگان بود. ۷۰ درصد از مشارکت کنندگان (  $N=70$  ) آذری زبان بودند، ۲۵ درصد آن‌ها (  $N=25$  ) فارسی زبان و ۵ درصد باقی مانده (  $N=5$  ) کردزبان بودند. در مورد این متغیر، سطح معناداری (  $F(2-99) = 2.559$  ,  $P = .083$  ) بود که نشان از نبود رابطه معناداری بین متغیر و نتایج آزمون تافل شرکت کنندگان داشت. تحلیل بیشتر جنسیت x گروه سنی (  $F(1-99) = .008$  ,  $P = .931$  )، زبان مادری x گروه سنی (  $F(2-99) = 1.376$  )، جنسیت x زبان مادری (  $F(2-99) = 1.424$  ,  $P = .246$  ) نشان دهنده رابطه‌ای با نتایج آزمون نبود. در نهایت، سطح معناداری سه متغیر گروه سنی، جنسیت و زبان مادری به صورت (  $F(1-99) = 1.477$  ,  $P = .227$  ) بود و در اینجا هم محققان هیچ رابطه معناداری بین این سه متغیر مورد بررسی با امتیازات تافل iBT شرکت کنندگان پیدا نکردند.

## ۵. بحث

محققان سؤالات تحقیق و روندهای آماری مشخص را ارائه دادند و تلاش کردند این سؤالات را مورد بررسی قرار دهند. سؤال اول در مورد آشنایی با رایانه و رابطه احتمالی آن با نتایج آزمون رایانه محور تافل بود؛ در سؤال دوم تحقیق، هدف اصلی و عمده بررسی رابطه بین رشته تحصیلی دانشجویان و نتایج آزمون تافل آن‌ها بود؛ سومین سؤال تحقیق، با ماهیت توصیفی خود، در تلاش بود تا دیدگاه افراد شرکت کننده در تحقیق را نسبت به رایانه و فناوری بسنجد؛ آخرین سؤال تحقیق یک سؤال چندمتغیری بود که در آن، محقق با رابطه بین سه متغیر سن، جنسیت و زبان مادری با نتایج آزمون رایانه محور تافل سروکار داشت. سؤالات تحقیق این پژوهش، به نوعی طراحی شدند که به واسطه استفاده از آن‌ها، محققان در تلاش بودند تا به این سؤال پاسخ دهند که آیا متغیرهای مختلف می‌توانند رابطه احتمالی با نتایج آزمون رایانه محور تافل داشته باشند.

یافته‌های این تحقیق، به‌طور کلی، همسو با دیگر تحقیقات صورت‌گرفته در زمینه رایانه هستند. کلارینا و والاس<sup>۱</sup> (۲۰۰۲) در تحقیق خود ادعا داشتند، همان‌طور که دانشجویان در مورد رایانه، اطلاعاتی به‌دست می‌آورند، دانش‌آموزان مستعدتر می‌توانند از این آشنایی سود برند و نتایج بهتری کسب کنند. زمانی که همه افراد شرکت‌کننده با رایانه آشنایی پیدا کنند، مفهوم آشنایی با رایانه دیگر اهمیت خود را از دست می‌دهد. تیلور و همکاران (۱۹۹۹) با تحقیق بر روی ۱۲۰۰ فرد شرکت‌کننده در آزمون تافل گزارش کردند که آشنایی با رایانه، نمی‌تواند بر عملکرد افراد در آزمون‌های زبان تأثیر داشته باشد. آن‌ها به این جمع‌بندی رسیدند که رابطه معناداری بین آشنایی با رایانه و سطح عملکرد افراد در آزمون رایانه‌محور زبان وجود ندارد؛ متغیر دیگری که آن‌ها مورد بررسی قرار دادند، متغیر جنسیت بود که در تلاش بودند تا تعامل بین جنسیت و میزان آشنایی افراد شرکت‌کننده در پژوهش خود را بیابند. آن‌ها بیان کردند که از لحاظ آماری، رابطه معناداری بین جنسیت و آشنایی با رایانه وجود ندارد. اودو<sup>۲</sup> (۲۰۱۲) بیان کرد که هرچند رابطه معناداری بین شاخص‌های آشنای با رایانه با عملکرد افراد وجود دارد، این متغیرها به‌طور قطعی و مشخص نمی‌توانند واریانس کلی را در این نوع آزمون شرح دهند. ساواکی<sup>۳</sup> (۲۰۰۱) مشاهدات مشابهی را در بررسی جامع خود درباره آشنایی با رایانه در زمینه‌های مختلف علمی عرضه کرد. رحمان (۲۰۱۱) در تلاش بود تا دیدگاه معلمان را درباره رایانه و فناوری مورد بررسی قرار دهد. نتایج تحقیق او نشان داد که دیدگاه معلمان، نسبت به فناوری و رایانه، به‌طور کلی مثبت است. به‌نظر می‌رسد که نگرش مثبت به رایانه‌ها، امری جهان‌شمول است. این می‌تواند به‌علت اجتناب‌ناپذیر بودن یارانه‌ها در زندگی روزمره باشد. بوسر و دوهرتی<sup>۴</sup> (۱۹۹۸) دیدگاه دانشجویان را نسبت به فناوری در برنامه‌های آموزشی فناوری‌محور مورد بررسی قرار دادند. نتایج این تحقیق بیان کرد که دیدگاه دانشجویان، می‌تواند تا حدی تحت تأثیر ۹ هفته در معرض روند آموزشی قرار بگیرد. تفاوت‌های معنادار در نتایج پیش‌آزمون و پس‌آزمون، در سه روش از چهار روش آموزشی دیده شدند. ویتتر، چودوبا و گوتک<sup>۵</sup> (۱۹۹۸) دیدگاه نسبت به رایانه و تأثیرات آن بر روی مفاهیم رفتاری را مورد بررسی

1. Clariana, Wallace

2. Odo

3. Sawaki

4. Boser, Daugherty

5. Winter, Chudoba, Gutek

قراردادند. آن‌ها بیان کردند که دیدگاه نسبت به رایانه می‌تواند دلیل خوبی برای سازمان‌هایی باشد که قصد دارند تا تعداد ساعاتی را که در حین آن‌ها کارمندان از رایانه استفاده می‌کنند، افزایش دهند. آن‌ها همچنین بیان کردند که دیدگاه نسبت به رایانه می‌تواند استفاده از رایانه را نیز توجیه کند.

تحلیل نتایج و بررسی آن‌ها براساس تحقیقات قبلی با همان دیدگاه، مبین این نکته بود که دیدگاه‌های مثبت نسبت به رایانه می‌تواند استفاده از رایانه را تخمین بزند و همان‌طور که در بخش‌های مشخصی از این مقاله اشاره شد، تجربیات مرتبط با استفاده از رایانه، می‌توانند منجر به آشنایی با رایانه شوند. در طرف دیگر، این تحقیق نشان داد که آشنایی با رایانه، نمی‌تواند تأثیری بر روی نتایج آزمون داشته باشد.

نتایج این تحقیق همچنین بر نتایج به دست آمده از تحقیقات پیشین در مورد سن و جنسیت، صحه می‌گذارد. نتایج به دست آمده از تحقیقات گرانپایه و کونان (۲۰۰۷) نشان داد که سن رابطه مشخصی با نمره‌های آزمون ندارد. نتایج تحقیق صالحی و طیبی (۲۰۱۲) هم مبین همین نکته است. به نظر می‌آید تحقیقات در خصوص این عوامل، نتایج قطعی را در برنخواهد گرفت. در دنیای امروزی و با تأکید بر برابری زن و مرد در جوامع مختلف، از جمله جامعه ایران، تفاوت‌های جنیستی در حال کم‌رنگ شدن است.

## ۶. نتیجه‌گیری

هر تغییری به واسطه کاربرد فناوری در جامعه، باعث ایجاد هنجارها و کاربردهای جدیدی شده است. آموزش در بین آن حوزه‌هایی است که تحت تأثیر این ویژگی فناوری قرار گرفته‌اند؛ زیرا فناوری گزینه‌ها و انعطاف چشمگیری در مورد حوزه‌های مختلف به همراه دارد، بسیاری از عقاید مرسوم و شیوه‌های تفکر و بررسی را تغییر می‌دهد. در مورد سایر حوزه‌ها، محیط آموزشی، می‌توان رایانه را پیش‌شرطی حیاتی برای هر تلاش نوآورانه‌ای در نظر بگیرند. در پژوهش حاضر، محقق در مورد این ویژگی فنی صحبت کرده است و به تشریح این مفهوم پرداخته است که ابزار مهمی مانند رایانه، می‌تواند باعث بروز مشکلات خاصی شوند. یک آزمون رایانه‌ای می‌تواند گزینه مناسبی برای این رویه باشد و از طریق آن، توانایی‌ها و خصلت‌های خاصی را می‌توان ارزیابی و سنجش کرد. ولی گاهی این ویژگی به نوبه خود، می‌تواند باعث مشکلاتی شود که در نهایت، نتایج آزمون را به انحراف بکشد. در مورد این منبع مشکل احتمالی، کارشناسان در این حوزه بررسی باید تصمیمات خاصی بگیرند و سعی

در تشریح راه‌ها و ابزارهایی داشته باشند که از طریق آن‌ها، می‌توانند این مشکلات را رفع بکنند.

رایانه یک ویژگی فنی رایج است که بسیاری از عقاید مرسوم در مورد مفاهیم مختلف را تغییر داده است. به خاطر این تغییر قابل توجه، فعالان حوزه‌های مختلف، سعی در بررسی نقش این مفهوم فنی در حوزه‌های مختلف مانند انجام آزمون دارند. در این پژوهش، کاربرد رایانه در محیط آزمون‌گیری، دغدغه اصلی به شمار می‌رفت. در نتیجه، محقق سعی در بررسی نقش رایانه در حوزه برگزاری آزمون کرد. آشنایی با رایانه به‌مثابه متغیر مهمی در این پژوهش در نظر گرفته شد و محقق خواهان بررسی رابطه بالقوه آن با نتایج آزمون بود. حوزه مطالعه، سن و جنسیت، سایر متغیرهای مستقلی بودند که محقق سعی در یافتن رابطه آن‌ها با نتایج آزمون داشت و آخرین متغیر نگرش شرکت‌کنندگان نسبت به رایانه بود.

هرچند نوآوری‌هایی مانند رایانه که از طریق عرضه راه‌حل‌های جدید یادگیری و آزمون، وارد محیط کلاس درس شده‌اند، بدون نقاط ضعف نیستند. به‌همین خاطر، بسیاری از فعالان حوزه آزمون‌سازی، نگران نقش این پدیده فنی هستند. این پژوهش به این نتیجه‌گیری انجامید که تجربه به‌کارگیری رایانه و آشنایی با رایانه، تأثیر معناداری بر آزمون رایانه‌ای مدنظرمان ندارند. در نتیجه، نکته اصلی در مورد نقش رایانه در حوزه آزمون‌سازی، تغییر در نگرش شرکت‌کنندگان درباره آزمون‌سازی و سنجش عملکرد به کمک رایانه است. بطور خلاصه می‌توان گفت:

الف: مقادیر محاسبه‌شده از نتایج به‌دست‌آمده نشان دادند که آشنایی با رایانه و تجربه کار با رایانه، هیچ رابطه معناداری با نتایج به‌دست‌آمده از شرکت‌کنندگان نداشت. در نتیجه، محقق به این نتیجه رسید که آشنایی با رایانه، نمی‌تواند باعث اختلال در نتایج آزمون شود. عوامل دیگر ممکن است منجر به تفاوت‌هایی در نتایج شوند که در تحقیقات آتی در این حوزه می‌توانند مورد بررسی قرار گیرند.

ب: رشته تحصیلی نمی‌تواند تفاوت‌هایی در نتایج آزمون به همراه داشته باشد.

پ: دانشجویان دیدگاه‌های مثبتی در مورد رایانه و فناوری داشتند. این نتیجه‌گیری به فعالان این حوزه نشان می‌دهد که آن‌ها باید به آزمون‌های رایانه‌ای توجه خاصی کنند و از این حس مثبت نهایت استفاده را ببرند.

تحلیل نتایج نشان می‌دهد که مشخصه‌های فرهنگی و پیشینه‌ای خاصی مانند سن، جنسیت و زبان مادری، باعث تفاوت در نتایج آزمون نمی‌شوند.



۷- منابع

- Alderson, J. C. (2000). *Assessing reading*. New York: Cambridge University Press.
- American Psychological Association. (1985). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association
- Ary, D., Jacobs, L., Sorensen, C., & Walker, D. (2014). *Introduction to research in education*. Belmont: Cengage Learning.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(5), 593-602.
- Choi, I. C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20(3), 295-320.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81.
- Cohen, A. (1984). On taking tests: what the students report. *Language Testing*, 1, 70-81.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings*. New York: Rand McNally.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington DC: American Council on Education.
- Durndell, A., & Lightbody, P. (1993). Gender and computing: change over time? *Computers & Education*, 21(4), 331-336.
- Eignor, D. (1999). Selected technical issues in the creation of computer-adaptive tests of second language reading proficiency. In M. Chalhoub-Deville (Ed), *Issues in computer-adaptive testing of reading proficiency* (pp. 167-181).UK: Cambridge University Press.
- Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179-197.
- Fulcher, G. (1999). Computerizing an English language placement test. *ELT Journal*, 53(4), 289-299.
- Geissler, J. E., & Horridge, P. (1993). University students' computer knowledge and commitment to learning. *Journal of Research on Computing in Education*, 25(3), 347-365.

- Geranpayeh, A., & Kunnan, A. J. (2007). Differential item functioning in terms of age in the certificate in advanced English examination. *Language Assessment Quarterly*, 4 (2), 190-22.
- Hambleton, R. K., (1984). Validating the test score. In R.A. Berk (Ed.), *A guide to criterion-referenced test construction*. (pp. 199- 230). Baltimore: Johns Hopkins University Press,.
- Kirsch, I., Jamieson, J., Taylor, C., & Eignor, D. (1998). Computer familiarity among TOEFL examinees. *ETS Research Report Series*, 1998(1), i-23.
- Lee, J. A. (1986). The effects of past computer experience on computerized aptitude test performance. *Educational and Psychological Measurement*, 46(3), 727-733.
- Lennon, R. T. (1956). Assumptions underlying the use of content validity. *Educational and Psychological Measurement*, 16, 294-304.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory: Monograph Supplement 9. *Psychological Reports*, 3(3), 635-694.
- Loyd, B. H., & Gressard, C. (1984). Reliability and factorial validity of computer attitude scales. *Educational and Psychological Measurement*, 44(2), 501-505.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30(10), 955.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Messick, S. (1993). Foundations of validity: Meaning and consequences in psychological assessment. *ETS Research Report Series*, 1993(2), i-18.
- Odo, D. M. (2012). Computer familiarity and test performance on a computer-based cloze ESL reading assessment. *Teaching English with Technology*, 12(3), 18-35.
- Rahman, M. A. (2011). Teacher educators' attitudes towards computer: Perspective Bangladesh. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.865.9821&rep>
- Rezaee, A., & Salehi, M. (2008). The construct validity of a language proficiency test: A multitrait multimethod approach. *TELL*, 2 (8), 93-110.
- Popham, W. J. (1978). *Criterion-referenced measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Salehi, M. (2011). On the construct validity of the reading section of the university of Tehran English Proficiency Test. *Journal of English Language Teaching and Learning, Faculty of Literature and Humanities of Tabriz University*, 22, 129-160.

- Salehi, M. (2012). The construct validity of a test: A triangulation of approaches. *Language Testing in Asia*. <https://doi.org/10.1186/2229-0443-2-2-102>.
- Salehi, M., & Bagheri Sanjareh, H. (2013). On the comparability of C-test and cloze test: A verbal protocol approach. *English for Specific Purposes World*, 14(39).
- Salehi, M. and Tayebi, A. (2012). Differential item functioning in in terms of gender in the reading sub-section of a high-stakes test. *Iranian Journal of Applied Language Studies*, 4(1), 135-168.
- Schmidt, F. L., Hunter, J. E., Pearlman, K., Hirsh, H. R., Sackett, P. R., Schmitt, N., ... & Zedeck, S. (1985). Forty Questions about Validity Generalization and Meta-Analysis. *Personnel Psychology*, 38(4), 697-798.
- Shulman, L. S. (1970). Reconstruction of educational research. *Review of Educational Research*, 40(3), 371-396.
- Taylor, C., Kirsch, I., Jamieson, J., & Eignor, D. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning*, 49(2), 219-274.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Houndmills, England: Palgrave Macmillan.
- Weir, C., & O'Sullivan, B. Jin Yan and Bax, S. (2007). Does the computer make a difference? Reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS Writing component: effects and impact. *IELTS Research Report*, 7, 311-347.
- Winter, S. J., Chudoba, K. M., & Gutek, B. A. (1998). Attitudes toward computers: when do they predict computer use? *Information & Management*, 34(5), 275-284.