

ارزیابی روایی و پایایی آزمون زبان انگلیسی وزارت علوم، تحقیقات و فناوری (MSRT)

رضامراد صحرائی^۱

هانیه ممقانی^۲

چکیده

زمینه: روایی و پایایی از جمله مسائلی هستند که در آزمون‌سازی از اهمیت ویژه‌ای برخوردارند. آزمون زبان انگلیسی وزارت علوم، تحقیقات و فناوری (MSRT) از جمله آزمون‌های مهارت زبان انگلیسی است که برای اعزام دانشجویان به خارج از کشور توسط معاونت دانشجویی وزارت علوم، تحقیقات و فناوری تقریباً هر دو ماه یک بار در ایران برگزار می‌شود. **هدف:** هدف از انجام این پژوهش که از نوع همبستگی و تجربی است این است که به ارزیابی روایی و پایایی آزمون MSRT پرداخته شود. **روش:** برای ارزیابی روایی این آزمون از روایی سازه‌ای (همسانی درونی) و برای ارزیابی پایایی درونی نیز از روش آلفای کرونباخ استفاده گردید و تمام این محاسبات از طریق نرم‌افزار (اس‌پی‌اس) انجام شد. **یافته‌ها:** سؤالاتی که در این پژوهش مطرح شدند به این صورت بودند که روایی و پایایی آزمون MSRT چگونه است، و اینکه آیا آزمون مهارت تافل می‌تواند مبنای مناسبی برای ارزیابی توانایی‌های زبانی زبان‌آموزان قرار گیرد؟ نتایج این پژوهش نشان داد: ۱. آزمون MSRT در سطح کلی دارای روایی و پایایی مناسبی است اما در سطح جزئی از روایی لازم برخوردار نیست. به عبارت دیگر، ضریب همبستگی بین بخش‌های شنیداری - ساختاری و شنیداری - خواندنی کمتر از بخش ساختاری - خواندنی است. ۲. آزمون مهارت تافل، آزمون مناسبی برای سنجش توانایی‌های زبانی زبان‌آموزان نیست چون فاقد بخش صحبت کردن است و همچنین دو بخش دستور و واژگان را جداگانه مورد سنجش قرار داده است.

واژگان کلیدی: روایی، پایایی، MSRT، آزمون مهارت.

۱. عضو هیئت علمی دانشگاه علامه طباطبائی Email: SahraeeReza@gmail.com

۲. کارشناس ارشد آموزش زبان فارسی به غیرفارسی‌زبانان

مقدمه

آزمون مهارت^۱ در زبان از جمله آزمون‌هایی است که به منظور تعیین میزان توانایی آزمودنی در زبان به کار می‌رود. چنین آزمونی با نگرشی به گذشته و با هدف سنجش آنچه که تا زمان برگزاری آزمون فرا گرفته شده است، اجرا می‌شود و توجهی به شیوه زبان‌آموزی آزمودنی ندارد (جعفرپور، ۱۳۷۵، ص ۱۵). آزمون MSRT از جمله آزمون‌های مهارتی زبان است که برای اعزام دانشجو به خارج از کشور توسط وزارت علوم، تحقیقات و فناوری تقریباً هر دو ماه یک بار در ایران برگزار می‌شود. اهمیت این آزمون برای کسانی است که نمره قابل قبولی در آزمون زبان خارجی مانند تافل (TOEFL) ندارند و قصد دارند در مقطع دکترا در داخل یا خارج از کشور ادامه تحصیل دهند. آزمون MSRT در واقع همان آزمون وزارت فرهنگ و آموزش عالی^۲ یا MCHE است که اخیراً به MSRT تغییر نام یافته است. دلیل این امر این است که MCHE از عنوان وزارت فرهنگ و آموزش عالی گرفته شده، این در حالی است که سال‌هاست این وزارتخانه تحت عنوان وزارت علوم، تحقیقات و فناوری خوانده می‌شود.

این آزمون که توسط وزارت علوم برگزار می‌شود از سه قسمت تشکیل شده است:

۱. درک شنیداری^۳: این بخش شامل ۳۰ سؤال است. سؤالات این بخش شامل تعدادی مکالمه کوتاه و سه سخنرانی است. آزمودنی‌ها باید در این بخش به ۳۰ سؤال در مدت ۱۵ دقیقه پاسخ دهند.
۲. ساختار و عبارات نوشتاری^۴: این بخش نیز از ۳۰ سؤال تشکیل شده است که ۱۵ مورد آن مربوط به ساختار و ۱۵ سؤال دیگر مربوط به عبارات نوشتاری می‌باشد. مدت پاسخگویی به این بخش نیز ۳۰ دقیقه است.
۳. درک مطلب خوانداری^۵: این بخش از ۴۰ سؤال تشکیل شده که معمولاً شامل ۴

1. proficiency test
 2. ministry of Culture and Higher Education (MCHE)
 3. listening comprehension
 4. structure and written expression
 5. reading comprehension

متن جداگانه است. ارزیابی دانش واژگانی آزمودنی‌ها نیز در قالب همین بخش صورت می‌گیرد. مدت پاسخگویی به سؤالات درک مطلب ۴۵ دقیقه می‌باشد.

آزمون MSRT نمره منفی ندارد و سالانه بیش از ۱۰۰۰۰ نفر شرکت کننده دارد و نمره این آزمون برای دانشجویان دکتری کلیه دانشگاه‌های دولتی و غیردولتی مورد قبول است. نمره قبولی در آزمون ۵۰ به بالا است که البته در مواردی بسته به دانشگاه مورد نظر متغیر است و از ۴۵ تا ۷۵ تغییر می‌کند.

از آنجایی که روایی^۱ و پایایی^۲ هر آزمون از جمله ویژگی‌های اساسی آن به شمار می‌روند و هر آزمون استاندارد بایستی از روایی و پایایی مطلوبی برخوردار باشد، بنابراین پژوهش به منظور ارزیابی روایی و پایایی آزمون MSRT صورت گرفته است.

سؤال‌هایی که در این پژوهش مطرح می‌شوند به صورت زیر است:

۱. روایی و پایایی آزمون MSRT چگونه است؟
۲. کدامیک از دو آزمون مهارت تافل (TOEFL) و آیلتس (IELTS) می‌تواند مبنای مناسبی برای سنجش توانایی‌های زبانی زبان‌آموزان قرار گیرد؟
فرضیه‌های این پژوهش عبارتند از:

۱. آزمون MSRT در سطح کلی دارای روایی و پایایی است اما در سطح تحلیل جزئی (بخش بخش) از روایی لازم برخوردار نیست.
۲. آزمون مهارت آیلتس (TOEFL) مبنای مناسبی برای سنجش توانایی‌های زبانی زبان‌آموزان نیست.

پیشینه مطالعات

در این بخش پیشینه مطالعات در زمینه روایی و پایایی آزمون مهارت مورد بررسی قرار می‌گیرد.

پژوهش‌های انجام شده در زمینه روایی و پایایی آزمون‌های مهارت زبان انگلیسی در ایران

شیخ‌الاسلامی (۱۳۷۸/۱۹۹۹) در رساله‌ای با عنوان «بررسی روایی آزمون ورودی تحصیلات تکمیلی در ایران^۱» به ارزیابی روایی آزمون ورودی دوره‌های کارشناسی ارشد پرداخته است. برای دستیابی به این منظور، وی به مقایسه نمرات به‌دست آمده از آزمون ورودی تحصیلات تکمیلی (دوره‌های کارشناسی ارشد ناپیوسته داخل) سالهای ۷۶ و ۷۷ با نمرات آزمون استاندارد تافل می‌پردازد. وی به دنبال این مطلب است که نمرات این دو آزمون تا چه اندازه با هم همبستگی دارند. نتایجی که از این پژوهش بدست آمده، بیانگر این است که بین آزمون تافل و سه بخش آزمون ورودی تحصیلات تکمیلی همبستگی وجود دارد، همچنین قسمت دستور آزمون ورودی تحصیلات تکمیلی ضریب همبستگی بالایی با قسمت دستور تافل دارد. اما دو قسمت دیگر آزمون یعنی واژگان و درک مطلب، ضریب همبستگی پایینی با دو قسمت واژگان و درک مطلب تافل دارند.

سعیدی (۱۳۸۶/۲۰۰۷) در رساله‌ای با عنوان «بررسی مقابله‌ای اعتبار آزمون C و آزمون cloze به عنوان آزمون‌های سنجش دانش زبان انگلیسی^۲» به بررسی مقایسه‌ای اعتبار آزمون C و آزمون cloze می‌پردازد. این دو آزمون از جمله آزمون‌های زبان هستند که سعی در سنجش تمامی مهارت‌های زبانی به طور همزمان دارند. در واقع آزمون C به عنوان جایگزینی مناسب‌تر برای آزمون cloze توسط طراحان آن پیشنهاد شد. وی همچنین در این تحقیق به همبستگی میان عملکرد زبان آموزان در این دو نوع آزمون با آزمون تافل می‌پردازد. نتایج به دست آمده در این تحقیق، تأییدی بر اعتبار آزمون C و آزمون cloze به عنوان ابزارهای آزمونی مفید جهت سنجش دانش زبانی زبان‌آموزان در سطح بالاتر از متوسط می‌باشد. نکته قابل توجه در میان یافته‌های این تحقیق، همبستگی بالای آزمون‌های

1. On the Validity of the TEFL MA. Entrance Examination Proficiency Tests in Iran
2. A Comparative Study on the Validity of C-test and Cloze Procedure as Integrative Tests of General English Proficiency.

c و cloze و قسمت ساختار زبان آزمون تافل می‌باشد.

پیشقدم (۱۳۸۱/۲۰۰۳) در رساله‌ای با عنوان «بررسی روایی پیش‌بین آزمون جدید کنکور زبان انگلیسی^۱» به بررسی این موضوع پرداخته است که از آنجایی که برنامه چهار ساله زبان انگلیسی در گرایشهای مختلف دبیری، ادبیات و مترجمی در دانشگاه‌های کشور موفق نبوده است، مسئولان سازمان سنجش کشور در سال ۱۳۸۱ تصمیم به تغییر آزمون کنکور سراسری در رشته زبان انگلیسی کردند و زیر آزمون جدیدی را با عنوان "زبان تخصصی" به مجموعه سؤال‌های زبان انگلیسی افزودند و این تحقیق به بررسی روایی پیش‌بین آزمون جدید زبان انگلیسی می‌پردازد. نتایج این پژوهش که با دو گروه از دانشجویان روزانه و شبانه انجام شد، نشان داد که دانشجویان روزانه هم در نمرات کنکور و هم در نمرات پایان ترم از دانشجویان شبانه بهتر عمل کرده بودند که این خود دلیلی است بر این ادعا که آزمون جدید زبان انگلیسی از قدرت پیش‌بینی بالایی برخوردار است و نتیجه دیگر این بود که از میان زیر آزمون‌ها تنها زیر آزمون جدید زبان تخصصی به عنوان قوی‌ترین پیش‌بینی‌کننده معرفی گردید و زیر آزمون‌های زبان عمومی، بینش اسلامی، فارسی و عربی هیچ یک رابطه معناداری با نمرات پایان ترم اول نداشتند.

قاسمی و رزنگه (۱۳۸۴/۲۰۰۵) در رساله‌ای با عنوان «بررسی روایی پیشگویی قسمت زبان عمومی آزمون ورودی کارشناسی ارشد مجموعه زبان انگلیسی^۲» به بررسی روایی پیشگویی قسمت زبان عمومی آزمون‌های ورودی کارشناسی ارشد مجموعه زبان انگلیسی که در سالهای ۱۳۸۱ و ۱۳۸۲ برگزار شد، پرداخته است. به این منظور تعداد ۶۷ نفر از دانشجویان مقطع کارشناسی ارشد در سه رشته آموزش-مترجمی و ادبیات انگلیسی که در دانشگاه‌های علامه طباطبائی، تربیت مدرس و اصفهان تحصیل می‌کردند انتخاب شدند. آزمون معیار انتخاب شده برای این منظور، آزمون IELTS بود که به طور فراگیری در سراسر جهان به

1. The Predictive Validity of the New Version of the State Universities Entrance Examination Admitting Candidates to English Major.
2. On the predictive Validity of the Proficiency Section of the M. A. Entrance Examination in English Language Major.

عنوان یک آزمون مهارت زبان انگلیسی استفاده می‌شود. نتایج این پژوهش نشان داد که اولاً قسمت زبان عمومی آزمون ورودی مقطع کارشناسی ارشد مجموعه زبان انگلیسی و زیرمجموعه‌های آن دارای روایی پیشگوی مناسبی هستند. دوم اینکه قسمت مصاحبه آزمون زبان عمومی که در سال‌های ۱۳۸۱ و قبل از آن مورد استفاده قرار می‌گرفت، پیشگویی مناسبی از عملکرد آتی داوطلبان در آزمون‌های مهارت زبان انگلیسی نیست.

مبانی نظری

روایی

مهم‌ترین ویژگی یک آزمون خوب روایی^۱ است. آزمونی، روایی دارد که فقط آنچه را که می‌خواهیم بسنجد و نه چیز دیگر. به عبارتی روایی به معنای راستگویی ابزار اندازه‌گیری است. مفهوم روایی نسبی است زیرا روایی هر وسیله یا آزمون را باید در ارتباط با کاربرد ویژه آن تعیین کرد. از این رو گفته می‌شود که هر ابزاری می‌تواند دارای روایی بسیار زیاد، متوسط یا ضعیف باشد (جعفرپور، ۱۳۷۵، ص ۸۴). لین و گرانلاند^۲ (۲۰۰۰، ص ۷۳) روایی را به صورت «یک ارزشیابی از کفایت و مناسبت تفسیرها و استفاده‌های نتایج سنجش» تعریف کرده‌اند. کاپلان و ساکوزو^۳ (۲۰۰۱، ص ۱۳۲) گفته‌اند «روایی را می‌توان به صورت توافق بین نمره آزمون و کیفیتی که قرار است اندازه بگیرد تعریف کرد» (نقل از سیف، ۱۳۹۰، ص ۴۱۴). روایی جنبه‌های گوناگونی دارد: روایی ظاهری (صوری)، محتوایی، سنجای یا ملاکی، سازه‌ای. همگی این روش‌ها به منظور تعیین رابطه بین کارکرد آزمون و واقعیت‌های قابل مشاهده در مورد توانایی یا مهارت مورد نظر است. البته، وجود هیچ‌یک از این جنبه‌ها به تنهایی، روایی آزمون را تضمین نمی‌کند. هر چهار ویژگی مکمل یکدیگرند. به عبارت دیگر، بر حسب نوع و هدف آزمون، یکی از این چهار جنبه، اولویت می‌یابد. برای نمونه در آزمون‌های پیشرفت، وجود روایی محتوایی الزامی است (جعفرپور،

1. validity
2. Linn & Gronlund
3. Kaplan & Saccuzzo

۱۳۷۵، ص ۸۴). در ادامه انواع روایی را بیشتر بحث خواهیم کرد.

روایی محتوایی

اینکه آزمون تا چه اندازه نمونه‌هایی از هدف‌های رفتاری^۱ مورد سنجش را در برمی‌گیرد، به روایی محتوایی آن مربوط می‌شود. یعنی محتوای آزمون تا چه اندازه مطالب مورد اندازه‌گیری را دربردارد. برای تعیین موجودیت روایی محتوایی، باید محتوای آزمون بازبینی شود و میزان روایی آن با توجه به هدف آزمون مشخص شود. آزمون خوب باید دربردارنده نمونه‌های کافی و مستقل در حوزه مهارت یا توانایی مورد نظر باشد. بدیهی است که نخست باید حوزه مورد سنجش به شیوه منظمی مورد واکاوی قرار گرفته باشد تا اطمینان به دست آید که ماده‌های آزمون، تمام جوانب مهم آن حوزه‌ها را بر حسب اهمیت در بردارند یا نه (جعفرپور، ۱۳۷۵، ص ۸۵).

برای محاسبه روایی محتوایی باید گفت که در واقع هیچ‌گونه روش آماری برای تعیین ضریب روایی محتوایی وجود ندارد. در عوض برای تعیین روایی محتوایی یک آزمون از قضاوت متخصصان در این باره که سؤال‌های آزمون تا چه میزانی معرف محتوا و هدف‌های برنامه هستند استفاده می‌شود. با این حال، "قضاوت‌های مربوط به روایی محتوایی نه قطعی هستند و نه نهایی؛ آزمون‌ها از سطوح مختلف روایی برخوردارند و متخصصان در قضاوت‌های خود همیشه با هم توافق نظر ندارند" (مرفی و دیوید شوfer^۲، ۱۹۹۴، ص ۱۱۰، نقل از سیف، ۱۳۹۰، ص ۴۱۶).

روایی ظاهری (صوری)

روایی صوری یا ظاهری به برداشت‌ها یا واکنش‌های آزمودنی‌ها، مدرس، ناظران اجرای آزمون^۳، مدیران برنامه آموزشی، و متخصصان آزمون‌شناسی مربوط می‌شود. بدین معنی که تا چه اندازه آزمون مورد بحث، به نظر این افراد، بهینه است. اگر شکل ظاهری و نوع

1. behavioral objectives
2. Murphy & Daivdshofer
3. proctors

ماده‌های آزمون‌ی به نظر آزمودنی‌ها بهینه نباشد، این برداشت ممکن است در کارکرد آنان تأثیر منفی بگذارد و هدف آزمونگر از انجام آزمون که تعیین توانایی به حق آن‌هاست، برآورده نخواهد شد (جعفرپور، ۱۳۷۵، ص ۸۵). مورفی و دیوید شوfer (۲۰۰۱، ص ۱۵۴) گفته‌اند "اصطلاح روایی صوری به درجه‌ای از منطقی بودن که یک آزمون به نظر کسانی که به آن جواب می‌دهند می‌رسد، گفته می‌شود" (نقل از سیف، ۱۳۹۰، ص ۴۱۶).

در تعیین روایی صوری می‌توان گفت از آنجا که روایی صوری نوع خاصی از روایی محتوایی است، همان روشی که برای تعیین روایی محتوایی به کار می‌رود، یعنی استفاده از نظر متخصصان، در اینجا نیز برای تعیین روایی صوری قابل استفاده است (همان، ص ۴۱۷).

روایی سنج‌های (ملاکی)

منظور از روایی سنج‌های (ملاکی) میزان ارتباط بین نمرات حاصل از یک آزمون با نمرات حاصل از یک آزمون دیگر یا وسیله اندازه‌گیری دیگر است. بنابراین هر زمان که از نمرات یک آزمون برای پیش‌بینی عملکرد افراد در آزمونی دیگر استفاده کنیم با روایی ملاکی سر و کار داریم. آزمون دوم که عملکرد فرد در آن پیش‌بینی می‌شود ملاک نام دارد (همانجا). روایی سنج‌های بر دو نوع است: همزمان^۱ و پیشگو^۲.

در بررسی روایی سنج‌های همزمان، کارکرد آزمودنی در آزمون با سنج دیگر مقایسه می‌شود که تقریباً همزمان برگزار شده است. مثلاً هنگامی که نمره‌های گروهی از آزمودنی‌ها در آزمون معلم ساخته‌ای را با نمره‌های همان گروه در یک آزمون استاندارد مقایسه می‌کنیم، در پی آن هستیم که بدانیم به چه میزان آزمون معلم-ساخته، همان چیزی را اندازه‌گیری می‌کند که آزمون استاندارد می‌سنجد. اما در بررسی روایی سنج‌های پیشگو، کارکرد آزمودنی در آزمون مورد بررسی با نتایج آزمونی مقایسه می‌شود که با فاصله زمانی برگزار شده و سنج تلقی می‌شود. (جعفرپور، ۱۳۷۵، ص ۸۶).

1. concurrent
2. predictive

روایی سازه‌ای

برخی از توانایی‌ها یا صفات، قابل مشاهده نیستند و بنابراین قابل اندازه‌گیری هم نیستند. اما از آنجایی که می‌توانیم پنداشت‌هایی از آن صفات یا توانایی‌ها را در ذهن داشته باشیم، این امکان وجود دارد که تکلیف‌ها و یا تمرین‌هایی در نظر بگیریم که پیدایش و بروز آن پنداشت‌ها را در آزمودنی‌ها پیش‌بینی کنند. در صورتیکه این پیش‌بینی‌ها درست از آب درآیند، روایی پنداشته‌شده یا فرضیه مورد نظر، تحقق یافته است. البته لازم نیست که تکلیف‌ها با توانایی‌ها یا صفات مورد پیش‌بینی، همسانی کامل داشته باشند، بلکه فقط کافی است که همبستگی پذیرفتنی بین آن دو وجود داشته باشد. بنابراین روایی سازه‌ای به این نکته توجه دارد که آزمون تهیه شده تا چه اندازه ویژگی یا توانایی یا سازه ویژه‌ای را اندازه‌گیری می‌کند. اندازه‌گیری بسندگی در یک زبان، علاقه به مطالعه، و نمونه‌هایی از این دست، سازه دانسته می‌شوند (جعفرپور، ۱۳۷۷، ص ۱۸۴). روش‌های مختلفی برای تعیین روایی سازه‌ای وجود دارد که در ادامه به توضیح هر یک از آنان خواهیم پرداخت.

الف) همبستگی بین نمره‌ها

این شیوه، یکی از روش‌های تعیین روایی است. با ملاحظه الگوهای همبستگی بین نمره‌ها به صورت مستقیم و یا غیرمستقیم، می‌توان گفت که ضریب همبستگی ناشی از رابطه بین نمره‌های دو آزمون، تا چه حد نشانگر واریانس یکی با دیگری است. طبیعی است که ضریب همبستگی به خودی خود قابل نتیجه‌گیری نیست. این پژوهشگر است که به استناد مقدار ضریب همبستگی نتیجه‌گیری می‌کند (همان، ص ۱۸۵).

ب) تعیین روایی درونی^۱

در این شیوه نیازی به استفاده از سنجه نیست. تنها به ضرایب همبستگی بین پرسش‌های آزمون و جمع نمره‌های آنان توجه می‌شود و روایی تنها بر اساس موجودیت خود آزمون تعیین خواهد شد. در این شیوه، فرض بر این است که اگر پرسش‌های آزمون دارای روایی

1. internal construct validation

سازه‌ای هستند، ضریب همبستگی ۲ رشته‌ای، مثلاً بین هر پرسش واژگان و نمره کل بخش واژگان باید بیشتر از ضریب همبستگی ۲ رشته‌ای بین هر پرسش واژگان و نمره کل بخش دستوری باشد. یعنی اگر ضریب همبستگی بین پرسش اول واژگان و بخش واژگان، $0/42$ است، باید این ضریب از ضریب همبستگی بین پرسش اول واژگان و بخش دستوری بالاتر باشد. در حقیقت اگر نتایج نشان دهد که همه پرسش‌های واژگانی دارای این ویژگی بوده‌اند، بنابراین دارای روایی سازه‌ای هستند (همان، صص ۱۸۵-۱۸۶). لازم به ذکر است که برای تعیین روایی آزمون MSRT از این روش استفاده شده است.

ج) تحلیل عامل^۱

راه دیگر تعیین روایی سازه استفاده از روش تحلیل عاملی است. تحلیل عاملی یا تحلیل عوامل یک روش آماری است که از طریق آن تعداد و ماهیت متغیرهایی را که یک آزمون اندازه می‌گیرد، مشخص می‌کنند. در تحلیل عامل، این کار از طریق ادغام تعداد زیادی از ماده‌ها یا متغیرها با همدیگر و ایجاد تعداد معدودی متغیر قابل بررسی به نام عامل انجام می‌شود. این روش روابط درونی میان داده‌های مورد تحلیل را تعیین می‌کند و برای ایجاد سهولت، متغیرها را به خوشه‌ها یا عوامل کاهش می‌دهد. بعد این خوشه‌ها یا عوامل را با توجه به آنچه به وسیله سؤال‌های آزمون اندازه‌گیری می‌شوند، نام‌گذاری می‌کنند. بنابراین، یک آزمون را می‌توان بر حسب عواملی که آن را تشکیل می‌دهند، نام‌گذاری کرد (سیف، ۱۳۹۰، صص ۴۲۴).

پایایی

از نظر بیرجندی و مصلی‌نژاد (۲۰۱۰، صص ۱۶۶)، پایایی در واقع یکدستی نمرات آزمون است. یکدستی نمرات، بدین معناست که این نمرات چقدر از یک سنجش/آزمون تا سنجش/آزمون دیگر، استوار و باثبات هستند؛ به عبارتی، سنجش پایایی، برآوردی است از میزان نوسانی که ما در شرایط متفاوت، انتظار آن را داریم. از آنجایی که برای تعیین پایایی

۱۰ آزمون MSRT از روش همسانی درونی استفاده شده است در ادامه به معرفی روش‌های محاسبه پایایی در این روش می‌پردازیم.

روش‌های همسانی درونی

در روش همسانی درونی، بر همسانی یا یکنواختی ماده‌ها یا اجزای تشکیل‌دهنده یک آزمون تأکید می‌شود. برای این منظور، آزمونی را که می‌خواهند ضریب پایایی‌اش را تعیین نمایند یک بار بر گروه واحدی از آزمون‌شوندگان اجرا می‌کنند و بعد با یکی از روش‌های زیر به تعیین ضریب پایایی آن می‌پردازند (سیف، ۱۳۹۰، ص ۴۴۹).

الف) روشِ دو نیمه کردن^۱

در این روش، نیازی به برگزاری دوباره یک آزمون و همچنین تهیه نسخه برابر دوم از یک آزمون مشابه نیست. برای استفاده از این روش، یک بار برگزاری یک آزمون، کافی خواهد بود. ایده اصلی این روش آن است که سؤالات یک آزمون، متجانس^۲ هستند، به عبارتی تمام سؤال‌های یک آزمون در صدد این هستند تا عناصر یک ویژگی خاص را مورد سنجش قرار دهند، از جمله زمان‌ها، حروف اضافه،... در حقیقت، فرض این روش بر این است که نوعی تجانس درونی بین سؤالات وجود دارد. در نتیجه، رابطه‌ای که بین سؤال‌ها در یک آزمون وجود دارد به نوعی از طریق پایایی نمرات آزمون با توجه به رابطه درونی آنان محسوب می‌شود. به همین دلیل است که گاهی به این روش تحت عنوان تجانس درونی نمرات آزمون نیز اشاره می‌شود. شیوه کار بدین صورت است که وقتی یک آزمون با سؤالات متجانس بر روی یک گروه از آزمون‌دهندگان اجرا می‌شود، سؤالات آزمون به دو نیمه مساوی تقسیم می‌شوند و ضریب همبستگی بین این دو نیمه، برآوردی از پایایی نمره آزمون است (فرهادی و دیگران، ۲۰۰۷، صص ۱۳۶ - ۱۳۷).

برای محاسبه پایایی کل آزمون از فرمول زیر تحت عنوان فرمول اسپیرمن استفاده می‌شود:

1. split-half method
2. homogeneous

$$r_{(total)} = \frac{2(r\ half)}{1 + (r\ half)}$$

r = پایایی کل آزمون

(r half) = ضریب همبستگی بین نمره‌های دو نیمه

ب) روش کودر و ریچاردسون

کودر و ریچاردسون دو متخصص معروف آمار هستند که یک مجموعه فرمول‌های ریاضی را برای محاسبات آماری ارائه داده‌اند. یکی از فرمول‌های آنان "KR-21" است که برای محاسبه پایایی نمرات آزمون بکار می‌رود. فرمول KR-21 بر اساس این فرضیه است که تمام سؤال‌های یک آزمون به منظور سنجش تنها یک ویژگی، طراحی می‌شوند. گاهی به این روش، روش "تعادل عقلی" نیز می‌گویند. (فرهادی و همکاران، ۲۰۰۷، صص ۱۳۸ - ۱۳۹).

$$(KR-21)r = \left[\frac{K}{K-1} \right] \left[1 - \frac{\bar{X}(K-\bar{X})}{KV} \right]$$

K = تعداد سؤال‌ها در یک آزمون

\bar{X} = نمره میانگین

V = واریانس

r = پایایی کل آزمون

فرض اساسی رابطه‌های مطرح شده توسط کودر-ریچاردسون این است که سؤال‌های آزمون را می‌توان به صورت ۱ برای پاسخ درست و ۰ برای پاسخ غلط تصحیح کرد. بنابراین نمی‌توان این روابط را برای محاسبه ضریب پایایی پرسشنامه‌ها یا آزمون‌هایی که طیف پاسخ‌های آنان بیش از دو گزینه است (مثلاً ۱ تا ۴ یا ۱ تا ۵) مورد استفاده قرار داد (مومنی و قیومی، ۱۳۸۹، ص ۲۱۱). به همین دلیل در این پژوهش از این روش برای تعیین ضریب پایایی استفاده نشده است.

ج) روش آلفای کرونباخ

سومین روش تعیین پایایی آزمون با تأکید بر همسانی درونی، روش ضریب آلفا نام دارد که به آن ضریب آلفا یا ضریب آلفای کرونباخ و یا حتی آلفا نیز گفته می‌شود. کرونباخ (۱۹۵۱) واضح این روش تعیین پایایی است. در این روش، اجزا یا قسمت‌های آزمون برای سنجش ضریب پایایی آزمون به کار می‌روند. اگر قسمت‌های آزمون همان سؤال‌های آزمون باشند و سؤال‌ها به صورت درست/غلط/۰ و غلط/۰ تصحیح بشوند، ضریب آلفا برابر خواهد بود با KR20. اما اگر به جای سؤال‌ها یا ماده‌ها، آزمون از بخش‌ها یا قسمت‌هایی تشکیل شده باشد، مثل زمانی که یک آزمون از تعدادی خرده آزمون تشکیل یافته است و بخواهیم از آن‌ها در محاسبه ضریب پایایی کل آزمون استفاده کنیم، آنگاه از روش آلفای کرونباخ مطابق با فرمولی که در ادامه آمده است، استفاده می‌کنیم (سیف، ۱۳۹۰، ص ۴۵۴):

$$ra = \frac{j}{j-1} \left(1 - \frac{\sum \sigma_j^2}{\sigma^2}\right)$$

ra= ضریب پایایی کل آزمون

j= تعداد بخش‌های آزمون

σ^2 = واریانس نمرات کل آزمون

σ_j^2 = واریانس نمرات هر بخش آزمون

روش آلفای کرونباخ در شرایطی که نمرات سؤال‌ها نه تنها برای گزینه‌های دو ارزشی ۰ و ۱، بلکه برای گزینه‌های چند ارزشی نیز قابل استفاده است، گفته می‌شود اگر ضریب آلفا بیشتر از ۰/۷ باشد، آزمون از پایایی قابل قبولی برخوردار است. اگر روش ضریب آلفا و روش دو نیمه کردن را با هم مقایسه کنیم، خواهیم دید که روش دو نیمه کردن آزمون، نوعی روش آلفاست. به عبارت دیگر، ضریب آلفا معرف میانگین ضرایب پایایی حاصل از تمامی اجزای ممکن یک آزمون است. یعنی اگر همه ضریب‌های همبستگی حاصل از دو نیمه‌های ممکن یک آزمون ۴۰ سؤالی محاسبه شوند، میانگین این

ضرایب برابر با ضریب آلفای کرونباخ خواهد بود. بنابراین تفاوت بین روش ضریب آلفا و روش دو نیمه کردن آزمون، تفاوت دو واحد است. در روش دو نیمه کردن، یک نیمه آزمون با نیمه دیگرش مقایسه می‌شود، در حالیکه در روش ضریب آلفا هر سؤال آزمون با تک تک سؤال‌های دیگر مقایسه می‌گردد (مومنی و قیومی، ۱۳۸۹، ص ۲۱۲). لازم به ذکر است که برای محاسبه پایایی ۱۰ آزمون اخیر MSRT از فرمول آلفای کرونباخ استفاده شده است.

تجزیه و تحلیل داده‌ها

در این قسمت به تجزیه و تحلیل داده‌ها از جمله میانگین نمرات، انحراف معیار و روایی و پایایی آزمون‌ها پرداخته می‌شود.

نمرات میانگین بخش شنیداری، ساختاری، خواندن و انحراف معیار کل ده آزمون اخیر MSRT در جدول ۱ آورده شده است.

جدول ۱. نمرات میانگین بخش شنیداری، ساختاری، خواندن و انحراف معیار کل ده آزمون اخیر

MSRT

| تاریخ آزمون | م ب شنیداری | م ب ساختاری | م ب خواندن | میانگین کل | انحراف معیار کل |
|-------------|-------------|-------------|------------|------------|-----------------|
| ۸۹/۴/۱۷ | ۸۵.۸ | ۳۱.۱۵ | ۳۲.۱۹ | ۴۹.۱۴ | ۷۵.۴ |
| ۸۹/۶/۲۶ | ۲۹.۹ | ۷۴.۱۳ | ۱۵.۱۸ | ۷۲.۱۳ | ۳۶.۴ |
| ۸۹/۹/۱۹ | ۶۷.۹ | ۲۱.۱۵ | ۱۲.۱۷ | ۱۴ | ۲۱.۴ |
| ۸۹/۱۰/۲۴ | ۵۶.۱۰ | ۵۰.۱۴ | ۵۳.۱۹ | ۸۶.۱۴ | ۷۶.۴ |
| ۸۹/۱۱/۲۱ | ۹۹.۸ | ۲۳.۱۳ | ۶۷.۱۷ | ۲۹.۱۳ | ۰۶.۴ |
| ۸۹/۱۲/۱۳ | ۳۰.۹ | ۲۵.۱۵ | ۰۲.۱۸ | ۴۹.۱۴ | ۴۴.۴ |
| ۸۹/۱۲/۲۱ | ۶۵.۹ | ۶۹.۱۴ | ۵۰.۱۹ | ۶۱.۱۴ | ۹۵.۴ |
| ۹۰/۱/۱۹ | ۶۶.۸ | ۱۱.۱۲ | ۲۷.۱۶ | ۳۴.۱۲ | ۵۵.۶ |
| ۹۰/۳/۲۰ | ۷۹.۹ | ۴۸.۱۲ | ۵۷.۱۶ | ۹۴.۱۲ | ۵۷.۶ |
| ۹۰/۶/۱۸ | ۴۷.۸ | ۸۷.۱۱ | ۵۰.۱۴ | ۶۱.۱۱ | ۸۸.۵ |

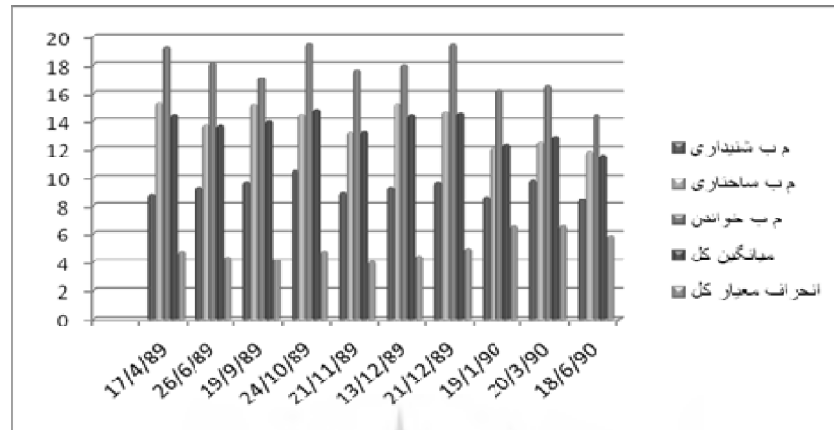
میانگین بخش شنیداری ۱۰ آزمون اخیر MSRT نشان می‌دهد که در ۴ آزمون اول یعنی آزمون‌های مورخ ۸۹/۴/۱۷، ۸۹/۶/۲۶، ۸۹/۹/۱۹ و ۸۹/۴/۲۴ شاهد پیشرفت عملکرد آزمون‌دهندگان هستیم. در آزمون بعدی یعنی ۸۹/۱۱/۲۱ شاهد کاهش عملکرد هستیم و بعد از آن نیز به صورت نوسانی عملکرد آزمون‌دهندگان افزایش و کاهش یافته است. اما به طور کلی می‌توان گفت که عملکرد آزمون‌دهندگان در این ۱۰ آزمون تفاوت چشمگیری با هم نداشته‌اند.

از میانگین بخش ساختاری ۱۰ آزمون اخیر MSRT اینگونه استنباط می‌شود که عملکرد آزمون‌دهندگان در این بخش نوسانات چشمگیری داشته است، به خصوص در سه آزمون آخر یعنی آزمون‌های مورخ ۹۰/۱/۱۹، ۹۰/۳/۲۰ و ۹۰/۶/۱۸ شاهد کاهش عملکرد آزمون‌دهندگان در این بخش هستیم.

از میانگین بخش خواندن ۱۰ آزمون اخیر MSRT اینگونه استنباط می‌شود که عملکرد آزمون‌دهندگان در این بخش نیز نوسانات چشمگیری داشته است، به خصوص در سه آزمون آخر یعنی آزمون‌های مورخ ۹۰/۱/۱۹، ۹۰/۳/۲۰ و ۹۰/۶/۱۸ شاهد کاهش عملکرد آزمون‌دهندگان در این بخش هستیم.

میانگین کل هر آزمون نشان می‌دهد که عملکرد آزمون‌دهندگان به صورت کلی خوب بوده است و آنان توانسته‌اند تقریباً به نیمی از سؤالات آزمون با موفقیت پاسخ دهند. از آنجایی که کوچک بودن نمرات انحراف معیار، نشان دهنده تفاوت کم نمرات آزمون با یکدیگر می‌باشد، با توجه به نمرات انحراف معیار می‌توان گفت که در ۷ آزمون اول نمرات تفاوت زیادی با هم نداشته‌اند اما در ۳ آزمون آخر یعنی آزمون‌های مورخ ۹۰/۱/۱۹، ۹۰/۳/۲۰ و ۹۰/۶/۱۸ تفاوت نمرات با یکدیگر چشمگیر است.

نمودار مقایسه‌ای کل آزمون‌ها در نمودار ۱ آورده شده است.



نمودار ۱. نمودار مقایسه‌ای کل ۱۰ آزمون اخیر MSRT

جدول کلی مربوط به پایایی درونی ۱۰ آزمون اخیر MSRT

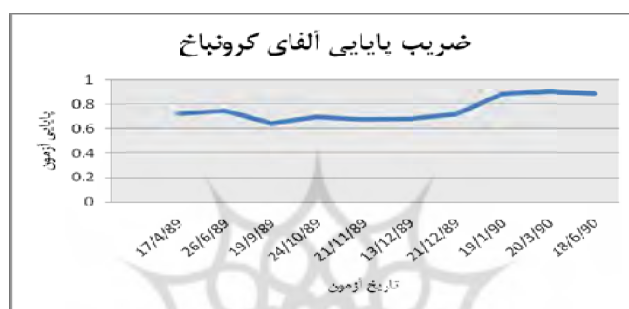
نمرات پایایی درونی ۱۰ آزمون اخیر MSRT که با روش آلفای کرونباخ که در ادامه آمده، محاسبه شده است در جدول ۲ آورده شده است.

$$ra = \frac{j}{j-1} \left(1 - \frac{\sum \sigma_j^2}{\sigma^2}\right)$$

جدول ۲. نمرات پایایی ۱۰ آزمون اخیر MSRT

| تاریخ آزمون | ضریب پایایی آلفای کرونباخ |
|-------------|---------------------------|
| ۸۹/۴/۱۷ | ۰.۷۱۹ |
| ۸۹/۶/۲۶ | ۰.۷۴۱ |
| ۸۹/۹/۱۹ | ۰.۶۳۸ |
| ۸۹/۱۰/۲۴ | ۰.۶۹ |
| ۸۹/۱۱/۲۱ | ۰.۶۷۳ |
| ۸۹/۱۲/۱۳ | ۰.۶۷۵ |
| ۸۹/۱۲/۲۱ | ۰.۷۱۴ |
| ۹۰/۱/۱۹ | ۰.۸۸ |
| ۹۰/۳/۲۰ | ۰.۹۰۳ |
| ۹۰/۶/۱۸ | ۰.۸۸۵ |

با توجه به جدول بالا، در ۱۰ آزمون اخیر MSRT، آزمون‌های مورخ ۸۹/۱۲/۱۳، ۸۹/۱۱/۲۱، ۸۹/۱۰/۲۴، ۸۹/۹/۱۹ از پایایی قابل قبولی برخوردار نیستند. سه آزمون آخر یعنی آزمون‌های مورخ ۹۰/۱/۱۹، ۹۰/۳/۲۰ و ۹۰/۶/۱۸ از ضریب پایایی بالایی برخوردارند و سایر آزمون‌ها دارای پایایی قابل قبولی هستند. نمودار (۲) نیز به روشنی گویای این مطالب است.



نمودار ۲. نمودار مقایسه‌ای پایایی ۱۰ آزمون اخیر MSRT

روایی سازه‌ای ۱۰ آزمون MSRT

در این پژوهش برای تعیین روایی سازه‌ای از ضریب همبستگی پیرسون استفاده شده است. ضریب همبستگی پیرسون، روشی پارامتری است و برای داده‌هایی با توزیع نرمال یا تعداد داده‌های زیاد استفاده می‌شود (مومنی و قیومی، ۱۳۸۹: ۱۱۰-۱۱۱). بنابراین هنگامی که داده‌ها به صورت مقیاس فاصله‌ای هستند، ضریب همبستگی، از فرمول پیرسون محاسبه می‌شود:

$$r = \frac{N(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[N\sum x^2 - (\sum X)^2][N\sum y^2 - (\sum Y)^2]}}$$

با توجه به محاسبات انجام شده، ضریب همبستگی بین دو بخش ساختاری - خواندنی به مراتب بیشتر از دو بخش شنیداری - ساختاری و شنیداری - خواندنی است. اما این وضعیت در سه آزمون آخر یعنی آزمون‌های مورخ ۹۰/۳/۲۰، ۹۰/۱/۱۹ و ۹۰/۶/۱۸ به چشم نمی‌خورد. به عبارتی، در این سه آزمون، ضریب همبستگی نه تنها بین دو بخش

ساختاری - خواندنی بالا است بلکه در بخش‌های دیگر یعنی شنیداری - ساختاری و شنیداری - خواندنی نیز بالاست. این امر بیانگر این مطلب است که سؤال‌های این دو آزمون کیفیت بهتری نسبت به سایر آزمون‌ها پیدا کرده‌اند، در واقع سعی شده است تا نوعی در هم تنیدگی بین سؤال‌ها برقرار شود.

بحث و نتیجه‌گیری

۱. آزمون MSRT در سطح کلی دارای روایی و پایایی مناسبی است اما در سطح جزئی از روایی لازم برخوردار نیست. به عبارت دیگر، ضریب همبستگی بین بخش‌های شنیداری - ساختاری و شنیداری - خواندنی کمتر از بخش ساختاری - خواندنی است. بنابراین فرض اول ما تأیید می‌شود.

۲. آزمون مهارت تافل، آزمون مناسبی برای سنجش توانایی‌های زبانی زبان‌آموزان نیست چون فاقد بخش صحبت کردن است و همچنین دو بخش دستور و واژگان را جداگانه مورد سنجش قرار داده است. بدین صورت فرض دوم ما نیز تأیید می‌شود.

۳. با توجه به میانگین کل نمرات بدست آمده در هر آزمون، می‌توان گفت که به طور کلی، عملکرد آزمون‌دهندگان رضایت بخش بوده است. اما اگر بخواهیم هر مهارت را به صورت جزئی مورد بررسی قرار دهیم، باید گفت که عملکرد آزمون‌دهندگان ابتدا در بخش درک مطلب (خواندن) و سپس ساختار خوب بوده است. در بخش شنیدن کمترین نمرات کسب شده است. به عبارت دیگر، آزمون‌دهندگان در بخش شنیدن از عملکرد ضعیفی برخوردارند که علت این امر می‌تواند یا به توانایی پایین آزمون‌دهندگان در مهارت شنیداری مربوط شود و یا نوع سؤالات شنیدن آزمون MSRT.

۴. در رابطه با پایایی آزمون MSRT می‌توان گفت که چون ضریب آلفا بیشتر از ۰/۷ بوده است، به طور کلی دارای پایایی قابل قبولی است.

منابع

- جعفرپور، عبدالجواد. (۱۳۷۵). *مقدمه‌ای بر آزمون‌شناسی زبان*. شیراز: انتشارات دانشگاه شیراز.
- جعفرپور، عبدالجواد. (۱۳۷۷). *آمار زبان‌شناختی*. شیراز: انتشارات دانشگاه شیراز.
- سیف، علی اکبر. (۱۳۹۰). *اندازه‌گیری، سنجش و ارزشیابی آموزشی*. تهران: نشر دوران.
- مومنی، م. و قیومی، ع. (۱۳۸۹). *تحلیل‌های آماری با استفاده از SPSS*. تهران: انتشارات مؤلف.
- Birjandi, P. and P. Mosallanejad (2010). *An Overview of Testing and Assessment*. Esfahan: Sepahan Publication.
- Farhady, H., A. Jafarpur and P. Birjandi (2007). *Testing Language skills from Theory to Practice*. Tehran: SAMT..
- Ghasemivarzaneh, sajad.(2005). "On the predictive Validity of the Proficiency Section of the M.A. Entrance Examination in English Language Major". MA thesis of TEFL.Allamehtabatabaee university.
- Pishghadam, Reza. (2003). "The Predictive Validity of the New Version of the State Universities Entrance Examination Admitting Candidates to English Major". MA thesis of TEFL.Allamehtabatabaee university.
- Saeedi, Masoud. (2007). " A Comparative Study on the Validity of C-test and Cloze Procedure as Integrative Tests of General English Proficiency". MA thesis of TEFL.Allamehtabatabaee university.
- Sheikholeslami, abbas.(1999). "On the Validity of the TEFL MA.Entrance Examination Proficiency Tests in Iran". MA thesis of TEFL.Allamehtabatabaee university.