

ارزیابی روایی و پایایی آزمون تعیین سطح زبان فارسی برای غیرفارسی‌زبانان بنیاد سعدی

رضامراد صحرائی^۱، غلامحسین رضاپور^۲

تاریخ دریافت: ۹۸/۰۲/۱۳

تاریخ پذیرش: ۹۸/۰۶/۱۱

چکیده

از ویژگی‌های مهم یک آزمون موفق، داشتن روایی و پایایی مناسب است. در این پژوهش به بررسی روایی و پایایی آزمون تعیین سطح زبان فارسی بنیاد سعدی پرداخته شد. این آزمون از سه بخش اصلی دانش زبانی، نگارش و صحبت کردن تشکیل شده است که بخش دانش زبانی شامل سه زیربخش دستور، واژه و خواندن است. جامعه آماری این آزمون متشکل از ۱۴۰ زبان‌آموز از ۴۱ کشور دنیا بود. برای محاسبه روایی، از ضریب آلفای کرونباخ و برای بررسی روایی که از انواع آن، روایی سازه‌ای انتخاب شد، ضریب همبستگی پیرسون به کار گرفته شد. میزان پایایی در زیربخش‌های دانش زبانی برای دستور ۰/۷۱، برای واژه ۰/۸۱۷ و برای خواندن ۰/۸۶۱ و در کل ۰/۹۱۶ به دست آمد. پایایی دو بخش صحبت کردن و نوشتن ۰/۷۹۴ و کل آزمون ۰/۸۹۵ محاسبه شد. با توجه به اینکه میزان پایایی قابل قبول در روش آلفای کرونباخ حداقل ۰/۷ است، مشخص شد که آزمون در بخش‌ها و زیربخش‌های گوناگون، و کلیت آزمون، پایایی مناسب و قابل قبولی داشت. ضرایب همبستگی به دست آمده بخش‌های آزمون، بین دانش زبانی و صحبت کردن ۰/۱۷۵، دانش زبانی و نوشتن ۰/۱۸۶ و نوشتن و صحبت کردن ۰/۶۰۲ بودند. شاید در نگاه اول این آزمون از روایی مناسب و قابل قبولی برخوردار نباشد، ولی با مقایسه میزان همبستگی بین زیربخش‌های بخش دانش زبانی و بخش مهارت‌های صحبت کردن و نوشتن، می‌توان نتیجه گرفت که زبان‌آموزان در بخش مهارت‌های تولیدی ضعیف عمل کرده‌اند و این نشان می‌دهد که در تدریس آنها فقط به دانش محض زبانی پرداخته شده است و به دانش روندی توجه نشده است. بر این اساس پیشنهاد شد که در برنامه درسی زبان‌آموزان به ارتقای سطح دانش روندی آنها توجه بیشتری شود.

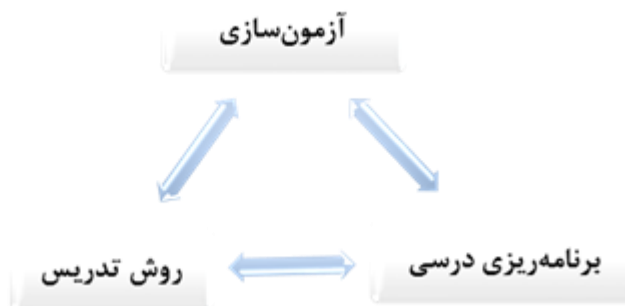
۱. دانشیار زبان‌شناسی، دانشگاه علامه طباطبائی، تهران، ایران.

۲. کارشناسی ارشد آموزش زبان فارسی به غیرفارسی‌زبانان، دانشگاه علامه طباطبائی، تهران، ایران. (نویسنده مسئول)

واژگان کلیدی: آزمون، آزمون تعیین سطح، روایی، پایایی، تأثیر پیامدی آزمون.

مقدمه

آموزش زبان دوم / خارجی یک حرفه تخصصی است که سه بخش اصلی و کلیدی دارد: روش تدریس^۱، برنامه‌ریزی درسی^۲ و آزمون‌سازی^۳. این سه بخش با یکدیگر رابطه فرایندی و چرخه‌ای دارند و به مثابه سه ضلع یک مثلث هستند که آموزش زبان دوم / خارجی بر آنها استوار می‌گردد.



شکل ۱. مثلث آموزش زبان دوم / خارجی

معلم زبان حرفه‌ای و موفق کسی است که علاوه بر آشنایی کافی با اصول و فنون تدریس و برنامه‌ریزی درسی، با اصول و مبانی آزمون‌سازی زبان و روش‌های تهیه آزمون نیز آشنا باشد. ارزشیابی و آزمون‌سازی جزء لاینفک حرفه معلمی زبان است، زیرا مبنای اصلی تصمیم‌گیری بهینه و عادلانه در نظام آموزش است. بر این اساس، معلم زبان باید با انواع آزمون و هدف هر یک آشنا باشد؛ اصول و فنون آزمون‌سازی را بداند؛ بر فرایند تهیه و اجرای آزمون مسلط باشد و شیوه‌های تفسیر و تحلیل نمرات آزمون را بشناسد. آزمون هم برای معلم اهمیت دارد و هم برای زبان‌آموز. آزمون به معلم کمک می‌کند تا نقاط قوت و ضعف روش تدریس خود را کشف کند؛ سطح زبان‌آموز را تشخیص دهد و برای ارتقای او به سطح بالاتر تصمیم بگیرد. آزمون برای زبان‌آموز اهمیت کلیدی دارد. آزمون

1. methodology
2. syllabus design
3. testing

خوب به زبان آموزان انگیزه مطالعه و یادگیری می‌دهد؛ در آنها احساس پیشرفت ایجاد می‌کند و مقیاس قابل اعتمادی برای مقایسه آنها با خودشان و سایر همکلاسی‌هایشان است (براون^۱، ۲۰۱۰).

روایی^۲ و پایایی^۳ از جمله مسائلی هستند که در هر آزمونی از اهمیت ویژه‌ای برخوردارند. به عبارتی هر آزمون استاندارد، باید دارای این دو ویژگی باشد. آزمونی که در این پژوهش روایی و پایایی آن مورد بررسی قرار گرفته، آزمون تعیین سطح زبان فارسی برای غیرفارسی‌زبانان بنیاد سعدی است. این آزمون با هدف تعیین سطح فارسی آموزان در بدو ورود به دوره دانش‌افزایی برگزار می‌شود. سطح سؤالات از آسان به مشکل طرح شده‌است و سطح‌بندی کلاس‌ها بر اساس استاندارد مرجع فارسی بنیاد سعدی^۴ (SPRF) است که در بخش‌های بعد به توضیح آن پرداخته می‌شود.

معمولاً روایی و پایایی آزمون‌های معتبر مورد ارزیابی قرار می‌گیرد. این کار برای آزمون‌های ارزیابی مهارت و تعیین سطح زبان دوم نیز انجام می‌شود. در ادامه، به صورت مختصر به چند پژوهش ایرانی و خارجی که به ارزیابی این ویژگی‌ها برای چند آزمون پرداخته‌اند، اشاره می‌شود. نیکلسون^۵ (۲۰۱۵) با پایش روایی و پایایی آزمون زبان انگلیسی برای ارتباط بین‌المللی^۶ (TOEIC) برای انگلیسی‌آموزان کره‌ای نتیجه گرفت که هر چند این آزمون کاربردی است، اما پایایی آن بحث‌برانگیز است. به اعتقاد وی TOEIC برای رسیدن به هدف خود، یعنی ارزیابی توانایی زبانی نامناسب است و هیچ شواهد مستقیمی وجود ندارد که نشان دهد این آزمون معیار خوبی برای ارزیابی بسندگی زبان انگلیسی است. حیدری، باقری، زمانیان، صادقی و یارمحمدی^۷ (۲۰۱۴) به بررسی روایی سازه‌ای بخش ساختار و بیان نوشتاری آزمون زبان سازمان سنجش ایران^۸ (TOLIMO) بر اساس نظریه پاسخ به مواد امتحانی^۹ (IRT) پرداختند و به این نتیجه رسیدند که بخش ساختار و

-
1. Brown
 2. validity
 3. reliability
 4. Sa'adi Persian Reference Framework
 5. Nicholson
 6. The Test of English for International Communication
 7. Heydari, Bagheri, Zamanian, Sadighi & Yarmohammadi
 8. The Test of Language by the Iranian Measurement Organization
 9. Item Response Theory (IRT)

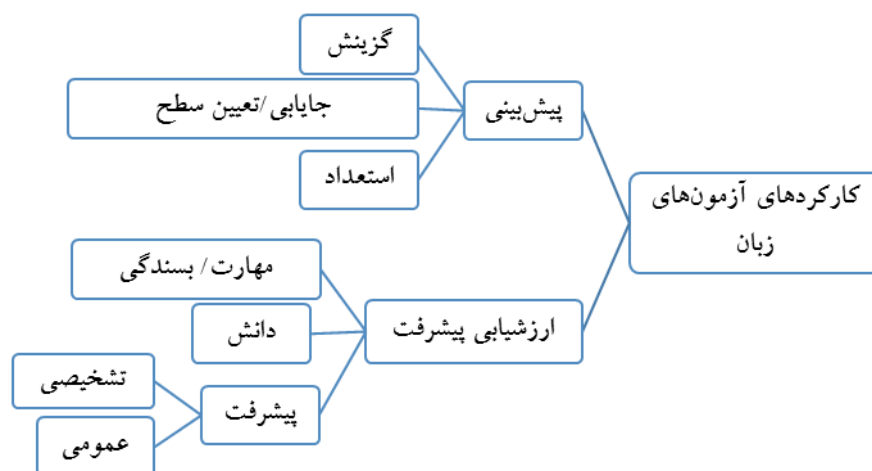
بیان نوشتاری TOLIMO می‌تواند توانایی زبانی آزمودنی‌ها را به‌خوبی ارزیابی کند، یعنی میزان روایی بالایی دارد. عربزاده و کدیور (۱۳۹۱) به بررسی پایایی، روایی و تحلیل عاملی مقیاس تأخیر رضامندی تحصیلی پرداختند. پایایی و همسانی درونی این پرسش‌نامه بر حسب ضرایب آلفای کرونباخ^۱ محاسبه شد و میزان قابل قبولی داشت. ممقانی (۱۳۹۰) به ارزیابی روایی و پایایی آزمون وزارت علوم، تحقیقات و فناوری^۲ (MSRT) با هدف دستیابی به یک آزمون مهارت برای زبان فارسی پرداخت. وی به این نتیجه رسید که این آزمون در سطح کلی دارای روایی و پایایی قابل قبول است اما در سطح تحلیل جزئی (بخش بخش) از روایی لازم برخوردار نیست. قنصولی (۱۳۸۹) ضمن طراحی یک نمونه آزمون بسندگی برای زبان فارسی با تأکید بر آزمون‌های شنیدن، خواندن، و دستور، روایی و پایایی آن آزمون را نیز مورد ارزیابی قرار داد. وی برای سنجش روایی این آزمون، از یک فرایند دو مرحله‌ای استفاده کرد. در مرحله اول، داده‌های بازنگری سه دانشجوی خارجی دوره دکتری زبان و ادبیات فارسی را تحلیل کرد و سپس با بهره‌گیری از نظریه بحث^۳ در سنجش، روایی آزمون را سنجید. وی پایایی آزمون را نیز با روش آلفای کرونباخ مورد ارزیابی قرار داد. نیکنامی، تقی‌پورظهیر و غفاری‌مجلج (۱۳۸۹) به بررسی پایایی و روایی پرسشنامه ارزیابی نوآوری سازمانی آمید و همکاران پرداختند. آنها برای این کار از ضریب آلفای کرونباخ برای ارزیابی همسانی درونی و استفاده کردند که با مقدار ۰/۸۳ میزان قابل قبولی بود.

اگر مفهوم آزمون زبان به درستی توصیف شود، باید گفت که آزمون، یک ابزار اندازه‌گیری است و در واقع تفاوت زیادی با یک خط‌کش منعطف لاستیکی یا یک دماسنج ندارد. آزمون زبان ابزاری برای سنجش توانایی زبانی است، حتی می‌توانیم ابعاد کمیته آزمون را بررسی کنیم: یک فرد چه میزان اطلاعات زبانی دارد؟ اما وقتی که می‌گوییم می‌خواهیم توانایی زبانی یا میزان اطلاعات زبانی فرد را بسنجیم، چه معنایی دارد؟ چگونه می‌توانیم مفهومی انتزاعی مانند توانایی زبانی را اندازه‌گیری کنیم؟ (براون، ۲۰۱۰). فرهادی، جفرپور و بیرجندی^۴ (۲۰۱۴) اینگونه عنوان می‌دارند که: «آزمودن، یک

-
1. Cronbach alpha
 2. Ministry of Science, Researches and Technology
 3. argument theory
 4. Farhady, Jafarpur & Birjandi

بخش مهم از هر برنامه آموزش زبان و زبان آموزی است. آزمون‌های خوش ساخت، که آموزگاران یا گروهی از آزمون‌سازان ماهر آماده می‌کنند، می‌توانند حداقل به دو شکل به آزمودنی‌ها کمک کنند؛ اول اینکه، آزمودن، آزمودنی‌ها را ترغیب به یادگیری خواهد کرد و آنها را در یادگیری موضوع آزمون برخواهد انگیزد، دوم آنکه، آزمودن به آزمودنی‌ها کمک خواهد کرد تا خودشان را آماده کنند و در نتیجه، مواد آموزشی را یاد بگیرند.»

انواع آزمون‌های زبان: آزمون‌های زبان برای اهداف مختلفی به کار می‌روند و متعاقباً کارکردهای متفاوتی دارند. جعفرپور (۱۳۷۵)، فرهادی و همکاران (۲۰۱۴) و بیرجندی و مصلی‌نژاد^۱ (۲۰۱۰)، دو هدف/کارکرد را برای آزمون‌های زبان عنوان و سپس بر اساس آن دو هدف و کارکرد، آزمون‌های زبان را تقسیم‌بندی می‌کنند. آزمون‌ها بر اساس کارکرد و هدفشان به دو دسته کلی آزمون‌های پیش‌بینی^۲ و آزمون‌های ارزشیابی پیشرفت^۳ تقسیم می‌شوند. به طور کلی و در یک نگاه نموداری، می‌توان انواع آزمون‌های زبان را بر اساس کارکردها/اهدافشان اینگونه نشان داد:



شکل ۲. انواع آزمون زبان بر اساس هدف/ عملکرد

1. Birjandi, & Mosallanejad
2. prognostic test
3. evaluation of attainment tests

آزمون‌های پیش‌بین: این آزمون‌ها رابطه مستقیمی با آموزش‌ها یا دوره‌های آموزشی قبلی ندارند و همچنین، با پیش‌زمینه تحصیلی افراد مرتبط نیستند و برای تصمیم‌گیری درباره جایگاه شغلی یا آموزشی مناسب برای آزمودنی به کار می‌روند (فرهادی و همکاران، ۲۰۱۴). آزمون‌های گزینش، تعیین سطح و استعداد از این نوع هستند.

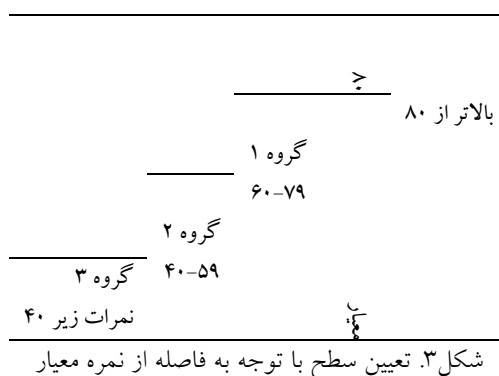
آزمون‌های گزینش^۱، برای به دست آوردن داده‌ها یا اطلاعاتی انجام می‌گیرد که بر که بر اساس آنها پذیرش آزمودنی‌ها در یک زمینه یا برنامه خاص رد یا تأیید شود. البته از آنجا که در بیشتر موارد، تعداد متقاضیان گذراندن یک آزمون، فراتر از ظرفیت برنامه‌های آموزشی می‌رود، بسیاری از داوطلبان، حتی با نمراتی بالاتر استاندارد مورد نیاز، پذیرفته نمی‌شوند (بیرجندی و مصلی‌نژاد، ۲۰۱۰).

آزمون تعیین سطح^۲ در تقسیم‌بندی آزمون بر اساس کارکرد/ هدف از نوع آزمون‌های پیش‌بین است که جعفری‌پور (۱۳۷۵) آن را آزمون گروه‌بندی نیز می‌نامد و می‌نویسد: «بر خلاف آزمون‌های گزینش، آزمون‌های گروه‌بندی برای اعلام قبولی یا ردی بکار نمی‌روند، بلکه هدف این آزمون‌ها گروه‌بندی و کلاس‌بندی همه پذیرفته‌شدگان است». این آزمون‌ها برای دسته‌بندی دانش‌آموزان یا دانشجویان تازه‌وارد، در گروه‌های نسبتاً متجانس و یک‌دست از لحاظ توان زمانی بکار می‌روند، به نحوی که آنها بتوانند یک دوره آموزشی را در سطح تقریباً مشابهی با دیگر افراد کلاس، آغاز کنند (بیرجندی و مصلی‌نژاد، ۲۰۱۰). آزمون‌های تعیین سطح، مناسب‌ترین کانال برای تصمیم‌گیری در مورد دسته‌بندی و گروه‌بندی آموزشی آزمودنی‌ها هستند. این آزمون‌ها بر خلاف آزمون‌های گزینش، تکیه‌ای بر پذیرش یا نپذیرفته شدن آزمودنی‌ها ندارند. هدف این آزمون‌ها، صرفاً تعیین میزان توانایی و سطح آمادگی زبانی آزمودنی برای یک دوره آموزش زبان مناسب است (فرهادی و همکاران، ۲۰۱۴).

آزمون‌های جایابی، یکی از پرکاربردترین آزمون‌هایی هستند که در سطوح مختلف آموزش زبان به کار گرفته می‌شود. به عنوان مثال، فرض کنید دانشجویان گروه آموزش زبان انگلیسی دانشگاهی، برای ادامه تحصیلات عالی نیازمند سطح بالایی از دانش زبان انگلیسی هستند. معمولاً همه دانشجویان، سطح زبانی مورد نظر را ندارند. بنابراین، گروه

-
1. selection test
 2. placement test

آموزش، اقدام به برگزاری یک آزمون جایابی می‌کنند. معیاری که گروه برای قبولی تغییر سطح و ادامه دوره آموزشی برای دانشجویان در نظر می‌گیرد، امتیاز ۸۰ و بیشتر از آن است. به عبارت دیگر، دانشجویانی که امتیاز ۸۰ یا بالاتر بیاورند، نیاز به آموزش بیشتر در سطح حاضر ندارند. همانطور که در شکل ۳ نشان داده شده است، نمرات دانشجویان گروه سه، از نمره معیار فاصله زیادی دارد و این نشان می‌دهد که نیاز به آموزش بیشتری دارند که به سطح مطلوب و مدنظر گروه برسند. گروه دو، در مقایسه با گروه سه، نیاز کمتری به آموزش دارند. به عبارت دیگر، هرچه فاصله نمرات از معیار بیشتر باشد، آموزش بیشتری مورد نیاز است.



هدف از برگزاری آزمون تعیین سطح، کمک به افرادی است که نیاز بیشتری به آموزش دارند. بنابراین، گام اول در تدریس مواد آموزشی، سطح‌بندی زبان‌آموزان است. تدریس هر سطح باید با توجه به اختلاف تا معیار، صورت پذیرد تا به ارتقای علمی زبان‌آموزان کمک شود. برای رسیدن به این هدف دو روش نه چندان متفاوت وجود دارد: اول اینکه با توجه به سطح مدت زمان کلی آموزش را تعیین کنیم. مثلاً برای سطوح ابتدایی یا مقدماتی سه ترم، سطح میانی دو ترم، و سطح پیشرفته فقط یک ترم، از نظر زمانی پیش‌بینی می‌شود. در روش دوم، میزان ساعتی که در هر ترم برای آموزش به کار گرفته می‌شود را با توجه به سطح تعیین می‌کنیم که هر چه سطح زبانی پایین‌تر باشد، زمان آموزش در هفته بیشتر لحاظ می‌گردد. مثلاً برای سطح مقدماتی، ده ساعت آموزش در هفته، سطح میانی، شش ساعت در هفته، و سطح پیشرفته سه ساعت در هفته آموزش در

نظر گرفته می‌شود. با توجه به موقعیت آموزشی، یکی از روش‌های مذکور یا هر دوی آنها به طور هم‌زمان به کار می‌روند (فرهادی و همکاران، ۲۰۱۴).

آزمون استعداد: این آزمون برای پیش‌بینی موفقیت یا شکست داوطلبان در دستیابی به اهداف معین شده به کار می‌رود. آزمون استعداد^۱، قابلیت، توانایی، و استعداد فرد را برای یادگیری زبان دوم یا خارجی می‌سنجد و در واقع، می‌تواند برای شناسایی یادگیرندگان که احتمال موفقیت بیشتری دارند، به کار رود (بیرجندی و مصلی‌نژاد، ۲۰۱۰).

آزمون‌های ارزشیابی پیشرفت: این آزمون‌ها برخلاف آزمون‌های پیش‌بین، مطالبی که قبلاً آموزش داده شده‌اند را می‌آزمایند. بنابراین، آزمون‌های ارزشیابی پیشرفت، بیشتر از آزمون‌های پیش‌بین به شرایط آموزش بستگی دارند (فرهادی و همکاران، ۲۰۱۴). این نوع آزمون‌ها مبتنی بر دوره‌های آموزشی گذرانده شده و یا کتاب‌های مطالعه شده توسط آزمودنی‌ها هستند (بیرجندی و مصلی‌نژاد، ۲۰۱۰). آزمون‌های دانش^۲، پیشرفت^۳، و مهارت^۴ از این نوع هستند.

آزمون دانش: آزمون دانش، برای سنجش توانایی زبانی افراد نیست، بلکه برای سنجش معلومات آنها در یک رشته خاص، در قالب زبانی غیر از زبان مادری‌شان به کار می‌رود. مثلاً یک آزمون فیزیک یا روانشناسی، به زبان انگلیسی، برای دانشجویان ایرانی، یک آزمون دانش تلقی می‌شود (فرهادی و همکاران، ۲۰۱۴).

آزمون پیشرفت: آزمون‌های به کار رفته برای اهداف این‌چنینی، میزان یادگیری دانش آموز را از یک مجموعه مطالب آموزشی معین شده، می‌سنجد. این آزمون‌ها نقش مهمی در محیط‌های آموزشی دارند، زیرا بیشتر آزمون‌های کلاسی در این طبقه قرار می‌گیرند. امتحان‌های کلاسی، نیمسال، و نهایی، نمونه کوچکی از این آزمون‌ها هستند. بیشتر آزمون‌های پیشرفت، به وسیله آموزگاران تهیه می‌شوند، زیرا این آزمون‌ها باید مبتنی بر مطالب یاد داده شده در کلاس باشند (فرهادی و همکاران، ۲۰۱۴).

آزمون مهارت: این آزمون‌ها برای اندازه‌گیری توانایی افراد در یک زبان، و بدون توجه به هر آموزشی که احیاناً آنها در زمینه آن زبان دیده‌اند، طراحی می‌شوند. بنابراین،

-
1. Aptitude test
 2. Knowledge test
 3. Achievement test
 4. Proficiency test

محتوای یک آزمون مهارت مبتنی بر محتوا یا اهداف دوره‌های زبانی‌ای که حیثاً آزمودنی‌ها دیده‌اند، نیست، بلکه مبتنی است بر تعریفی از آنچه که داوطلبان برای انجام دادن در زبان دارند (بیرجندی و مصلی‌نژاد، ۲۰۱۰).

ویژگی‌های یک آزمون موفق: در کتاب‌های آزمون‌سازی برای سنجش دانش زبانی زبان‌آموزان، ویژگی‌هایی برای یک آزمون ذکر شده است که اگر آزمونی آنها را کسب کند، به یک آزمون موفق تبدیل می‌شود (جلیلی، ۱۳۹۰). سودمندی: از نظر بکمن و پالمر^۱ (۱۹۹۶)، مهم‌ترین چیزی که باید در طراحی و ایجاد یک آزمون زبان لحاظ شود، سودمندی^۲، و کاربردی است که یک آزمون به آن منظور طراحی و ایجاد می‌شود. آنها اعتقاد دارند که، برای سودمند بودن، هر آزمون زبانی باید با یک هدف خاص و برای یک گروه معین از آزمون‌دهندگان و سنجش یک کاربرد خاص از زبان، طراحی و ایجاد شود (کومب، فالس و هابلی^۳، ۲۰۰۷). پایایی: بیرجندی و مصلی‌نژاد (۲۰۱۰) پایایی را انسجام نمرات آزمون می‌دانند و می‌نویسند «این انسجام نمرات، یعنی اینکه آنها چقدر از یک سنجش/آزمون تا سنجش/آزمون دیگر، استوار هستند، در واقع، سنجش پایایی، برآوردی است از میزان نوسانی که ما در شرایط متفاوت انتظار آن را داریم». فرهادی و همکاران (۲۰۱۴) اینگونه عنوان می‌دارند که «تکرار یک آزمون از یک فرد در برخی ویژگی‌ها بدون تغییر می‌ماند و نمره ثابت است. به این ثبات نمرات که در چند ارزشیابی حاکم است، پایایی گفته می‌شود». اما در بیانی جامع، تعریف پایایی را می‌توان اینگونه بیان کرد: «شاخصی است برای برآورد میزانی که یک آزمون، به آن میزان، نمرات منسجم و با ثبات تولید می‌نماید. این انسجام می‌تواند در طول زمان باشد (در یک بازه زمانی)، در دو نسخه (فرم) باشد، و یا در دو قسمت مساوی از یک آزمون». در این تعریف هر کدام از حالت‌های ذکر شده به یک نوع از روش‌های متداول سنجش پایایی اشاره دارد که در ادامه به آنها اشاره می‌شود:

-
1. Bachman & Palmer
 2. usefulness
 3. Coombe, Folse & Hubley

الف. روش آزمون - بازآزمون^۱ که پایایی از طریق برگزاری دوباره یک آزمون و محاسبه همبستگی^۲ بین دو نمره به دست آمده از هر بار برگزاری آن آزمون به دست می‌آید.

ب. روش نسخه‌های همتا^۳ (فرم‌های موازی) که در آن دو نسخه همتا از یک آزمون درباره یک گروه مشابه از آزمون‌دهندگان، در یک جلسه مشابه و تنها یک بار مورد آزمون واقع می‌شوند. سپس ضریب همبستگی^۴ بین نمرات به دست آمده از آن دو نسخه آزمون، برآوردی از پایایی نمره آزمون خواهد بود.

پ. روش آزمون-بازآزمون با نسخه‌های برابر^۵ که درباره این روش بیرجندی و مصلی‌نژاد (۲۰۱۰) می‌گویند: «در این روش دو نسخه متفاوت از یک آزمون با یک فاصله زمانی بین آنها، مورد آزمون واقع می‌شوند. ضریب پایایی در این روش، بیانگر خطاها در رویه آزمون، انسجام درباره نمونه‌های متفاوت سؤال‌ها، و پایداری و دوام روز به روز پاسخ‌های آزمون‌دهندگان باشد».

ت. روش دو نیمه کردن^۶ که در آن یک آزمون با سؤالات متجانس برای یک گروه از آزمون‌دهندگان برگزار می‌شود، سؤالات به دو نیمه مساوی تقسیم می‌شوند و ضریب همبستگی بین این دو نیمه، برآوردی از پایایی نمره آزمون است. برای محاسبه آن از فرمول زیر بهره می‌بریم:

$$r_{(total)} = \frac{2(r_{half})}{1+(r_{half})} \quad (\text{پایایی کل آزمون})$$

ث. روش کودر و ریچاردسون^۷ (KR-21) که عملی‌ترین، پرکاربردترین و راحت‌ترین شیوه محاسبه پایایی است. این روش که گاهی روش «تعادل عقلی»^۸ نیز نامیده می‌شود، یک روش صرفاً آماری برای محاسبه پایایی است که از فرمول زیر بهره می‌برد:

1. test-pretest method
2. correlation
3. parallel forms method
4. correlation coefficient
5. test-retest method with equivalent forms
6. split-half method
7. Kudar & Richardson method
8. rational equivalence

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

K= تعداد سؤالات آزمون

\bar{X} = نمره میانگین

V= واریانس

ج. ضریب آلفای کرونباخ^۱ (α) معادله‌ای است که برای ارزیابی همسانی درونی^۲ و پایایی مورد استفاده قرار می‌گیرد و معمولاً جایگزین روش دو نیمه کردن می‌شود و از فرمول زیر بهره می‌برد (براون^۳، ۲۰۰۸):

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

$r\alpha$ = ضریب پایایی کل آزمون

j= تعداد بخش‌های آزمون

σ^2 = واریانس نمران کل آزمون

σ_j^2 = واریانس نمرات هر بخش آزمون

روش آلفای کرونباخ در شرایطی که نمره‌های سؤالات 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 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 ضرایب همبستگی حاصل از دو نیمه ممکن یک آزمون ۵۰ 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 حالی که در روش ضریب آلفا، هر سؤال آزمون با تک تک سؤال‌های

-
1. Cronbach alpha coefficient (α)
 2. internal consistency
 3. Brown

دیگر مقایسه می‌گردد (مؤمنی و قیومی، ۱۳۸۹). روایی: از ویژگی‌های دیگر یک آزمون موفق، روایی آن است. روایی آزمون، بر خلاف پایایی آن، بیشتر به ماهیت آزمون و خود آزمون وابسته است و فقط وابسته به نمرات آن نیست. روایی، غالباً به خصوصیات و ویژگی‌های صوری آزمون نیز بستگی دارد. بنابراین، تنها به دست آوردن یک تعداد نمره از یک آزمون، کاربران را قادر به ثبت روایی آن نمی‌سازد. به عبارت دیگر، چارچوبی برای آزمون که بر اساس آن، موضوعی را اندازه‌گیری / ارزیابی کنند. یعنی سؤال روایی این است که «آیا آزمون مورد نظر، در حال دستیابی به آنچه که قصد دارد هست یا نه؟ آیا آزمون به هدفی که داشته، رسیده است یا خیر؟». بررسی روایی به سه گونه اجرا می‌شود (فرهادی و همکاران، ۲۰۱۴):

الف. روایی محتوایی^۱ که به محتوای آزمون با محتوای مطالب درسی‌ای که قرار است آزموده شود، اشاره دارد. برای آنکه آزمونی از لحاظ محتوا روا باشد، باید دو معیار را مد نظر قرار دهد. اول آنکه محتوای آزمون باید متناسب و منطبق با محتوای مطالب درسی‌ای که آزموده می‌شود، باشد. دوم اینکه، هدف از آزمون باید سنجش سطح متناسبی از آموخته‌های دانش‌آموزان باشد. ب. روایی ملاکی^۲ که بر حسب زمان اجرای آزمون، به دو نوع روایی همزمان^۳ و روایی پیش‌بینی^۴ تقسیم می‌شود، به انطباق بین نمرات حاصل از یک آزمون با نمرات حاصل از یک ملاک و معیار دیگر می‌پردازد. جعفرپور (۱۳۷۷) نیز می‌گوید: «مقصود از روایی سنجه‌ای (ملاکی)، تعیین همبستگی بین آزمون‌ها و یک سنجه (ملاک) دیگر است». پ. روایی سازه‌ای^۵ که 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.

روش‌های مختلفی برای تعیین روایی سازه‌ای وجود دارد از جمله همبستگی بین نمره‌ها، تعیین روایی درونی^۶ (تنها بر اساس موجودیت خود آزمون تعیین خواهد شد و به ضرایب همبستگی بین پرسش‌های آزمون و جمع نمرات آنان توجه می‌شود) و تحلیل

-
1. content validity
 2. criterion-related validity
 3. concurrent validity
 4. predictive validity
 5. construct validity
 6. Internal construct validation

عامل^۱ (روشی آماری که از طریق آن تعداد و ماهیت متغیرهایی را که یک آزمون اندازه می‌گیرد، مشخص می‌کنند) را می‌توان نام برد. عملی بودن: با وجود اینکه پایایی و روایی دو خصوصیت مهم یک آزمون هستند، قابل اجرا بودن و عملی بودن^۲ آن نیز مهم است. یعنی، یک آزمون باید عملی^۳ و کاربردی باشد. عملی بودن یعنی اینکه، آزمون باید سهولت برگزاری، سهولت نمره‌دهی، و سهولت در تفسیر و کاربرد نتایج داشته باشد (فرهادی و همکاران، ۲۰۱۴).

در این پژوهش تلاش شد تا به دو پرسش زیر پاسخ داده شود: ۱. میزان پایایی آزمون تعیین سطح زبان فارسی بنیاد سعدی در سال ۹۶ چقدر است؟ ۲. روایی آزمون تعیین سطح زبان فارسی بنیاد سعدی در سال ۹۶ چگونه است؟

روش

روش گردآوری اطلاعات در این پژوهش به صورت میدانی است که در آن از نمرات خام آزمودنی‌های آزمون تعیین سطح سال ۹۶ بنیاد سعدی استفاده شده است. روش پژوهش اینگونه بود که ابتدا میانگین و انحراف معیار نمره آزمودنی‌ها در هر بخش و سپس به صورت کلی محاسبه شد و در مرحله بعد پایایی آزمون مورد بررسی قرار گرفت. برای سنجش پایایی از پنج شیوه فوق‌الذکر، از محاسبه ضریب آلفای کرونباخ استفاده شد، چرا که از دیگر روش‌های محاسبه پایایی پذیرفته شده‌تر است. برای مثال، در تحلیل میزان پایایی آزمون آیلتس و تافل، هنوز هم ضریب آلفای کرونباخ معیار اصلی است. برای سنجش روایی سه نوع محتوایی، ملاکی و سازه‌ای ذکر شد که در این آزمون به سنجش روایی سازه‌ای پرداخته شد؛ چون آزمون شامل بخش‌های دستور، واژه، خواندن، گفت و گو، و نوشتن است. برای این منظور، از روش همسانی درونی و محاسبه ضریب همبستگی بین بخش‌های آزمون استفاده گردید. در این پژوهش از نرم‌افزار SPSS 23 برای تحلیل داده‌ها استفاده شده است.

-
1. factor analysis
 2. practicality
 3. practical

آزمون تعیین سطح بنیاد سعدی: آزمون تعیین سطح زبان فارسی بنیاد سعدی، تا کنون برای ۶۰۷ نفر از داوطلبان یادگیری زبان فارسی از ۵۱ کشور جهان، طی سال‌های ۹۴ تا ۹۶، در قالب دوره‌های دانش‌افزایی بنیاد سعدی، اجرا شده است. این آزمون توسط گروهی از استادان و کارشناسان با تجربه، در حوزه آموزش زبان فارسی به غیرفارسی‌زبانان طراحی شده است و از آنجا که معیار تولید سؤالات آزمون کتاب استاندارد مرجع آموزش فارسی در جها (صحرايي و همکاران، ۱۳۹۵) مصوب شورای علمی بنیاد سعدی است، این آزمون با نام «آزمون تعیین سطح بنیاد سعدی» شناخته می‌شود. این آزمون که مطابق آزمون‌های تعیین سطح میثیگان و آکسفورد آزمون دارای سه بخش اصلی است. بخش اول، بخش دانش زبانی است که دانش محض^۱ زبانی آزمودنی را اندازه می‌گیرد، یک آزمون کتبی ۵۰ سؤالی چهارگزینه‌ای است، که خود دارای سه بخش است و در جدول شماره یک نشان داده شده است:

جدول ۱. مواد بخش اول آزمون تعیین سطح بنیاد سعدی

نوع سؤالات	تعداد سؤالات
سؤالات واژه	۱۵ سؤال
سؤالات دستور زبان	۱۵ سؤال
سؤالات خواندن و درک مطلب	۲۰ سؤال

بخش دوم یک آزمون مصاحبه است که در آن دو آزمون‌گر، در چارچوب دستورالعمل آزمون‌های شفاهی با آزمودنی مصاحبه می‌کنند. بخش سوم، آزمون نگارش است که در آن از آزمودنی خواسته می‌شود درباره یک موضوع خاص، ۱۵۰ کلمه بنویسد. این موضوع از بین سه موضوع، که در اختیار آزمودنی‌ها قرار داده شده‌اند، توسط خود آزمودنی انتخاب می‌شود. این دو بخش میزان دانش روندی^۲ آزمودنی را می‌سنجد. پژوهشگر توانست به اطلاعات خام این آزمون دسترسی پیدا کند. بنابراین مطابق با اصول سنجش روایی و پایایی نتایج آزمون ۱۴۰ فارسی‌آموز خارجی، که در سال ۹۶ این آزمون

1. declarative knowledge
2. procedural knowledge

را گذرانده‌اند، مورد بررسی و ارزیابی قرار داده شد. با توجه به اطلاعاتی که از آزمون‌های تعیین سطح بنیاد سعدی در سال ۹۶ به دست آمد، جامعه آماری شامل ۱۴۰ نفر آزمودنی شرکت کننده در آزمون تعیین سطح زبان فارسی بنیاد سعدی بود. در جدول زیر آزمودنی‌های ما به تفکیک ملیت و جنسیت، مشخص شده‌اند؛ به این ترتیب که از مجموع ۱۴۰ نفر، ۹۰ نفر زن و ۵۰ نفر مرد، هستند.

جدول ۲. آزمودنی‌ها در سال ۱۳۹۶

ردیف	کشور	زن	مرد	جمع کل	ردیف	کشور	زن	مرد	جمع کل
۱	اتریش	۱	۰	۱	۲۲	تونس	۳	۰	۳
۲	آذربایجان	۰	۱	۱	۲۳	چین	۰	۱	۱
۳	ازبکستان	۲	۱	۳	۲۴	روسیه	۳	۶	۹
۴	آرژانتین	۱	۱	۲	۲۵	رومانی	۱	۰	۱
۵	ارمنستان	۴	۱	۵	۲۶	سوریه	۳	۱	۴
۶	اسپانیا	۲	۰	۲	۲۷	سوئیس	۱	۰	۱
۷	استرالیا	۰	۱	۱	۲۸	صربستان	۳	۰	۳
۸	اکراین	۷	۰	۷	۲۹	عراق	۱	۶	۷
۹	الجزایر	۰	۱	۱	۳۰	فرانسه	۱	۱	۲
۱۰	انگلستان	۳	۱	۴	۳۱	قرقیزستان	۲	۰	۲
۱۱	ایتالیا	۴	۱	۵	۳۲	قزاقستان	۲	۰	۲
۱۲	آلمان	۱	۰	۱	۳۳	قطر	۰	۱	۱
۱۳	بلاروس	۲	۰	۲	۳۴	گرجستان	۱	۱	۲
۱۴	بلژیک	۱	۰	۱	۳۵	لبنان	۷	۱	۸
۱۵	بلغارستان	۴	۰	۴	۳۶	لهستان	۴	۱	۵
۱۶	بنگلادش	۲	۳	۵	۳۷	مالزی	۱	۰	۱
۱۷	بوسنی	۱	۰	۱	۳۸	مجارستان	۵	۱	۶
۱۸	پاکستان	۵	۴	۹	۳۹	مصر	۲	۰	۲
۱۹	پرتغال	۱	۰	۱	۴۰	مقدونیه	۰	۱	۱
۲۰	تاجیکستان	۰	۱	۱	۴۱	هندوستان	۱	۱۰	۱۱
۲۱	ترکیه	۸	۳	۱۱					
	جمع کل	۸	۳	۱۱					۱۴۰

استاندارد مرجع بنیاد سعدی: بنیاد سعدی با هدف ایجاد یکپارچگی در سطوح آموزش زبان فارسی و با توجه به تنوع مخاطبان در موارد بالا و لزوم شناخت جامعه هدف، تجربه قریب به ۴۰۰ سال آموزش زبان فارسی و ویژگی‌های خاص نظام زبان فارسی، استاندارد مرجع آموزش فارسی را طراحی کرده است. در استاندارد مرجع بنیاد سعدی، هفت سطح برای آموزش زبان فارسی به شرح زیر در نظر گرفته شده است (صحرائی و همکاران، ۱۳۹۵):

جدول ۳. سطوح آموزشی زبانی

نام سطح	نشان اختصاری	ترجمه جهانی
۱ سطح نوآموز	N	Novice
۲ سطح مقدماتی	A	Elementary
۳ سطح پیش‌میانی	B1	Pre-intermediate
۴ سطح میانی	B2	Intermediate
۵ سطح فوق‌میانی	B3	Upper-intermediate
۶ سطح پیشرفته	C1	Advanced
۷ سطح ماهر	C2	Proficient

البته هر کدام از این سطوح دارای زیر سطح‌هایی هستند که در جدول شماره ۴ نشان

داده شده‌اند:

جدول ۴. زیرسطح‌های آموزش زبانی

نام سطح	نام زیر سطح
نوآموز	نوآموز
مقدماتی ۱	مقدماتی ۱
مقدماتی	مقدماتی ۲
پیش‌میانی	پیش‌میانی ۱
پیش‌میانی	پیش‌میانی ۲
میانی	میانی ۱
میانی	میانی ۲
فوق میانی	فوق میانی ۱
فوق میانی	فوق میانی ۲

پیشرفته ۱	
پیشرفته ۲	پیشرفته
فوق پیشرفته ۱ (خواندن و نوشتن)	ماهر
فوق پیشرفته ۲ (گوش دادن و صحبت کردن)	

آزمون تعیین سطح زبان فارسی بنیاد سعدی با توجه به معیاری که برای هر سطح مشخص شده است، سطح زبانی آزمودنی را مشخص می‌کند.

نتایج

برای محاسبه پایایی هر بخش به محاسبه ضریب آلفای کرونباخ با استفاده از نرم‌افزار SPSS 23 پرداخته شد که نتایج حاصل از این محاسبه در جدول شماره ۵، آورده شده است.

جدول ۵. ضریب آلفای آزمون سال ۱۳۹۶

بخش آزمون	ضریب آلفا (پایایی)
دستور زبان	۰/۷۱۰
واژه	۰/۸۱۷
خواندن	۰/۸۶۱
دانش زبانی	۰/۹۱۶
صحبت کردن و نوشتن	۰/۷۴۹
کل آزمون	۰/۸۹۵

برای محاسبه روایی سازه‌ای، باید میزان همبستگی نمرات در بخش‌های مختلف را به دست آورد. در پژوهش حاضر، از ضریب همبستگی پیرسون^۱ برای محاسبه همبستگی بین

نمرات بخش‌های مختلف آزمون استفاده شد. در جدول شماره ۶، میزان همبستگی بین نمرات بخش‌های مختلف آزمون بنیاد سعدی سال ۱۳۹۶ آورده شده است:

جدول ۶. ضریب همبستگی بین بخش‌های آزمون

بخش‌های آزمون	دانش زبانی	صحبت کردن	نوشتن
دانش زبانی	۱	۰/۱۷۵	۰/۱۸۶
صحبت کردن	۰/۱۷۵	۱	۰/۶۰۲
نوشتن	۰/۱۸۶	۰/۶۰۲	۱

در جدول ۷، میزان ضریب همبستگی زیربخش‌های دانش زبانی و بخش‌های صحبت کردن و نوشتن آورده شده است:

جدول ۷. ضریب همبستگی بین زیربخش‌های دانش زبانی و بخش‌های صحبت کردن و نوشتن آزمون

بخش‌های آزمون	دستور	واژه	خواندن	صحبت کردن	نوشتن
دستور	۱	۰/۶۱۵	۰/۶۳۸	۰/۱۶۱	۰/۱۹۰
واژه	۰/۶۱۵	۱	۰/۷۲۸	۰/۱۳۸	۰/۱۱۷
خواندن	۰/۶۳۸	۰/۷۲۸	۱	۰/۱۶۵	۰/۱۸۷
صحبت کردن	۰/۱۶۱	۰/۱۳۸	۰/۱۶۵	۱	۰/۶۰۲
نوشتن	۰/۱۹۰	۰/۱۱۷	۰/۱۸۷	۰/۶۰۲	۱

بحث و نتیجه‌گیری

برای محاسبه پایایی آزمون از ضریب آلفای کرونباخ استفاده شد. با توجه به اینکه حداقل میزان قابل قبول ضریب آلفا ۰/۷۰ است، در جدول شماره ۶ مشاهده می‌شود که تمامی بخش‌های این آزمون از پایایی قابل قبولی برخوردار هستند؛ چرا که بیشتر از حد قابل قبول ضریب آلفا هستند. همچنین این نتایج نشان می‌دهد که پایایی در بخش خواندن با میزان ۰/۸۶۱ بیشترین، و بخش دستور زبان با میزان ۰/۷۱۰ کمترین مقدار میزان را در بخش دانش زبانی، به خود اختصاص داده است. در کل آزمون، بخش دانش زبانی با میزان ۰/۹۱۶ بیشترین و بخش صحبت کردن و نوشتن با میزان ۰/۷۴۹ کمترین هستند. البته باید اذعان داشت که در کل آزمون و بخش‌های مختلف آن از پایایی قابل قبولی برخوردار هستند.

بنابراین پاسخ پرسش اول پژوهش تبیین شد. در بررسی روایی بخش‌های مختلف آزمون تعیین سطح بنیاد سعدی، با توجه به نتایج حاصل از محاسبه ضریب همبستگی که در جدول شماره ۷ مشاهده می‌شود، همبستگی بخش صحبت کردن و نوشتن با میزان ۰/۶۰۲ بیشترین مقدار را به خود اختصاص داده است. این مقدار بیانگر این است که این دو بخش از هم‌پوشی مناسبی برخوردار هستند. میزان همبستگی دو بخش دانش دستوری و صحبت کردن با مقدار ۰/۱۷۵ کمترین حد را در بین ضرایب داراست که هرچند مثبت است اما نمی‌توان این مقدار را برای یک آزمون قابل قبول دانست. همچنین، همبستگی بین بخش دانش زبانی و نوشتن نیز پایین است. شاید در نگاه اول، می‌توان اینگونه بیان داشت که آزمون در حالت کلی از روایی سازه‌ای خوب و قابل قبولی برخوردار نیست. اما وقتی به جدول ۷ دقت شود، میزان همبستگی زیربخش‌های بخش دانش زبانی و بخش‌های صحبت کردن و نوشتن بسیار پایین است ولی زیربخش‌های دانش زبانی همبستگی قابل قبولی دارند. این امر گواهی می‌دهد که بخش دانش زبانی آزمون، که دانش محض زبانی را می‌سنجد، روایی سازه‌ای خوبی دارد. بیشترین همبستگی، مربوط به زیربخش خواندن و واژه به میزان ۰/۷۲۸ است. بخش‌های صحبت کردن و نوشتن، که دانش روندی را ارزیابی می‌کنند، نیز با میزان ۰/۶۰۲ قابل قبول است و این دو بخش نیز از روایی سازه‌ای مناسب برخوردارند. با مقایسه میزان همبستگی بین زیربخش‌های بخش دانش زبانی و بخش مهارت‌های صحبت کردن و نوشتن، می‌توان به تأثیر پیامدی^۱ آزمون توجه کرد و اینکه زبان‌آموزان در بخش مهارت‌های تولیدی ضعیف عمل کرده‌اند که نشان دهنده این است که در تدریس آنها فقط به دانش محض زبانی پرداخته شده است و نه دانش روندی. به عبارت دیگر، چگونگی به کارگیری دانش محض را یاد نگرفته‌اند. بر این اساس پیشنهاد می‌شود که برنامه درسی زبان‌آموزان بیشتر متمرکز بر ارتقای سطح دانش روندی آنان و استفاده از دانش محض باشد. آموزش راهبردهای یادگیری را می‌توان در دستیابی به این امر، مفید دانست که البته نیازمند تحقیق و پژوهشی جامع است.

منابع

- براون، داگلاس. (۲۰۱۰). *درآمدی بر آزمون‌سازی زبان*. ترجمه رضامراد صحرائی و فائزه مرصوص (۱۳۹۶)، تهران: انتشارات نشرنویسه پارسی.
- جعفرپور، عبدالجواد. (۱۳۷۵). *مقدمه‌ای بر آزمون‌شناسی زبان*. شیراز: انتشارات دانشگاه شیراز.
- جعفرپور، عبدالجواد. (۱۳۷۷). *آمار زبان‌شناختی*. شیراز: انتشارات دانشگاه شیراز.
- جلیلی، سیداکبر. (۱۳۹۰). آزمون مهارتی فارسی (آمفا) پر پایه چهار مهارت اصلی زبانی، *پایان‌نامه کارشناسی ارشد*. دانشگاه علامه طباطبائی. تهران.
- صحرائی، رضامراد. و مرصوص، فائزه. (۱۳۹۵). *استاندارد مرجع آموزش زبان فارسی*. تهران: انتشارات دانشگاه علامه طباطبائی.
- عرب‌زاده، مهدی و کدیور، پروین. (۱۳۹۱). بررسی پایایی، روایی و تحلیل عاملی مقیاس تأخیر رضامندی تحصیلی. *فصلنامه اندازه‌گیری تربیتی*، ۳ (۹) ۱-۱۸.
- قنسولی، بهزاد. (۱۳۸۹). طراحی و روایی‌سنجی آزمون بسندگی زبان فارسی. *مجله پژوهش‌های زبان‌های خارجی*، ۵۷، ۱۱۵-۱۲۹.
- ممقانی، هانیه. (MSRT) با هدف دستیابی به یک آزمون مهارت برای زبان فارسی، *پایان‌نامه کارشناسی ارشد*. دانشگاه علامه طباطبائی، تهران.
- مؤمنی، منصور. و قیومی، علی. (۱۳۸۹). تحلیل‌های آماری با استفاده از SPSS، تهران: انتشارات مؤلف.
- نیک‌نامی، محمد.، تقی‌پورظهیر، علی. و غفاری‌مجلج، محمد. (۱۳۸۹). بررسی پایایی و روایی پرسشنامه ارزیابی نوآوری سازمانی امید و همکاران ۲۰۰۲. *فصلنامه اندازه‌گیری تربیتی*، ۱ (۳)، ۱۰۱-۱۱۹.
- نیک‌نامی، محمد.، تقی‌پورظهیر، علی. و غفاری‌مجلج، محمد. (۱۳۸۹). بررسی پایایی و روایی پرسشنامه ارزیابی نوآوری سازمانی امید و همکاران ۲۰۰۲. *فصلنامه اندازه‌گیری تربیتی*، ۱ (۳)، ۱۰۱-۱۱۹.
- Bachman, L. and A, Palmer. (2010). *Language assessment in practice*. UK: Oxford University Press.
- Birjandi, P. and P, Mosallanejad. (2010). *An Overview of Testing and Assessment*. Esfahan: Sepahan Publication.

- Brown, J. D. (2008). *Testing in Language Programs*. US: McGraw-Hill.
- Coombe, C., K. Folse and N. Hubley (2007). *A Practical Guide to Assessing English Language Learners*. US: The University of Michigan Press.
- Farhady, H., A. Jafarpur and P. Birjandi (2014). *Testing Language Skills from Theory to Practice*. Tehran: SAMT.
- Heydari, P., M.S. Bagheri, M. Zamanian, F. Sadighi, L. Yarmohammadi (2014). Investigating the Construct Validity of Structure and Expression Section of TOLIMO through IRT. *International Journal Language Learning and Applied Linguistics World*, 5, 105-123.
- McKinley, J. G. Thompson (2018). "Washback Effect in Teaching English as an International Language". *TESOL Encyclopedia of English Language Teaching*.
- Nicholson, S.J. (2015). "Evaluating the TOEIC in South Korea: Practically, Reliability and Validity". *International Journal of Education*, 7, 221-233.