

طراحی و تولید آزمون بسندگی زبان فارسی: مدل بدست آمده در دانشگاه الزهرا

وحید میرزائیان

چکیده

آزمون‌های بسندگی زبانی برای اکثر زبانهای دنیا بخصوص انگلیسی وجود دارد. با توجه به افزایش دانشجویان خارجی در ایران در سالهای اخیر و نیاز به سنجش سطح زبانی این دانشجویان، اخیراً توجه زیادی به این مهم صورت گرفته است. در همین راستا این مقاله به توضیح نحوه طراحی، تولید، اجرا و تعیین روایی و پایایی آزمون بسندگی زبان فارسی که در مرکز آموزش زبان فارسی به غیر فارسی زبانان دانشگاه الزهرا طراحی و اجرا شده است می‌پردازد. ابتدا پیشینه تولید چنین آزمون‌هایی بررسی می‌شود و مشخص می‌شود در این زمینه مدل خاصی وجود ندارد. گامهای برداشته شده برای تولید چنین آزمونی و مدلسازی بر اساس آن توضیح داده می‌شود. آزمون از سه بخش دستوری، خواندنی و شنیداری تشکیل شده است. سپس مراحل اجرای آزمون روی ۵۳ دانشجوی خارجی و نحوه محاسبه پایایی و روایی آن بیان می‌گردد. محاسبه روایی (محتوایی، ساختاری و همزمانی) و پایایی (با استفاده از دونیمه آزمون و آلفای کرونباخ) نشان می‌دهد آزمون از روایی و پایایی نسبتاً بالایی برخوردار است. در خاتمه توصیه‌هایی در خصوص نحوه تولید چنین آزمون‌هایی بر اساس تجارب بدست آمده ارائه می‌گردد.



کلید واژه‌ها: آزمون، بسندگی، زبان، فارسی،

۱. مقدمه
ارزیابی توانش زبانی، به لحاظ تأثیری که در سرنوشت آزمون‌دهنده دارد دارای اهمیت بسیار زیادی است. این نکته بخصوص در کشورهایی که نگرش فرهنگی خاص نسبت به نتیجه آزمون وجود دارد تشدید می‌شود (Cheng, 2005; Choi, 2008; Pan, 2013; Qi, 2007; Shohamy 2001, Watanabe, 2004). بر طبق نظر (Pan, 2014) با توجه به رشد فناوری‌های نوین و ارتباطات ملل با یکدیگر و نیز سهولت در جابجایی، تمایل به مهاجرت به کشورهای دیگر چه برای کار و چه ادامه تحصیل روز به روز گسترش می‌یابد. تقریباً اکثر دانشگاه‌های انگلیسی زبان دنیا از نتایج آزمون‌های بسندگی^۱ استاندارد مثل تافل و آیلتس برای پذیرش دانشجو استفاده می‌کنند و بسیاری از کشورهای انگلیسی زبان مهاجر پذیر نیز امتیاز ویژه‌ای را به نمره این آزمون‌ها اختصاص می‌دهند. از طرفی تقریباً تمامی کشورهای مهم همچون آلمان، فرانسه و روسیه نیز دارای آزمون‌های استاندارد بسندگی خاص خود هستند ((Pan, 2013; Roever & Pan, 2008; Tsai & Tsou, 2009). از طرفی اخیراً شاهد افزایش دانشجویان خارجی در ایران بوده ایم و از طرفی یکی از اولویتهای مهم دانشگاه‌های ایران جذب دانشجوی خارجی است. وزارت علوم تحقیقات و فناوری نیز شرکت در آزمون بسندگی زبان فارسی را جزو شروط اجباری پذیرش قرار داده است. لذا با توجه به محوریت کشور ایران بعنوان زادگاه و مهد زبان فارسی به نظر می‌رسد این موضوع مورد توجه قرار گرفته است و حتی آزمونی هم برای دانش‌آموزان در بدو ورود به دبستان طراحی گردیده که با واکنش منفی برخی اقلیت‌های قومی مواجه شده است.
آزمون‌های بسندگی زبانی علاوه بر اینکه یکی از شروط موفقیت برای مهاجرت یا ادامه تحصیل محسوب می‌شوند، از آنها برای قرارداد دانشجویان در کلاسهای متناسب با سطح آنها نیز استفاده می‌شود (Alderson, Clapham, & Wall, 1995; Brown, 2004; Hughes, 2003). علاوه بر این، از این آزمون‌ها به شیوه پیش آزمون - پس آزمون استفاده می‌شود تا موفقیت برنامه‌های آموزش زبان را نیز محک بزنند (Brown, 2004).

¹ Proficiency

از آزمونهای بسندگی زبانی برای مقاصد مختلفی استفاده می‌شود و تست‌های مختلفی در دنیا بخصوص برای زبان انگلیسی وجود دارد. لیکن هنوز آزمون‌های متعدد و استاندارد بسندگی برای زبان فارسی وجود ندارد و نیاز به کار در این حوزه احساس می‌شود. چنانچه بخواهیم چنین آزمونی حداقل برای زبان فارسی تولید کنیم الگوی خاصی را در پیش رو نداریم. هدف از نگارش این مقاله، توضیح فرایندی است که طی آن پژوهشگران در محل خدمت خود موفق به تولید آزمون استاندارد برای زبان فارسی شدند که دارای روایی^۱ و پایایی^۲ قابل قبول می‌باشد. ابتدا به فرایندی که منجر به تولید سوالات برای سه بخش زبانی دستور، خواندن و شنیدن شد می‌پردازیم. سپس، روش محاسبه روایی و پایایی توضیح داده می‌شود.

۲. پیشینه پژوهش

تعریف مشخص و اتفاق نظر قطعی در خصوص بسندگی زبانی وجود ندارد. بسیاری از پژوهشگران (Bachman & Palmer, 1996) کلمه توانایی^۳ را به بسندگی ترجیح می‌دهند چراکه واژه توانایی با این دیدگاه که اجزای زبان را بایستی جداگانه مورد آزمون قرار دهد سنجیت بیشتری دارد (Brown, 2004, p. 71). لیکن بر این نکته نیز اتفاق نظر وجود دارد که هر دو واژه توانایی و بسندگی دارای اجزایی هستند که می‌توان آنها را تبیین نموده و اندازه گیری نمود.

(McNamara, 2000) معتقد است بسندگی زبانی خود دارای اجزایی مستقل است. پس آزمون بسندگی مورد بحث در این پژوهش حول اجزای زبانی دستور^۴ و دو مهارت خواندن و شنیدن طراحی گردیده است. به همین نسبت، هدف از طراحی آن، سنجش بسندگی کلی زبان آموز بوده است نه سنجش مطالبی که در طول یک دوره تحصیلی توسط معلم و کتاب به زبان آموز آموزش داده شده است (Henning, 1987, p. 196).

موسوی (۱۳۷۹) در پژوهشی، آزمونی مشتمل بر چهار بخش جهت سنجش مهارت زبانی فراگیران خارجی زبان فارسی تهیه می‌کند. این چهار بخش شامل آزمون گوش دادن، قواعد ساختاری، واژگان و خواندن و درک مفاهیم است. برای هر یک از این بخشها، ماده‌هایی تهیه می‌شود و طی دو مرحله این ماده‌ها مورد واکاوی قرار می‌گیرد. ابتدا ۱۸۰ ماده برای این چهار بخش تهیه می‌شود که بعد از تایید روایی صوری، ۱۵۲ ماده باقی می‌ماند که به ۱۹ تن از دانشجویان خارجی دانشگاه علوم پزشکی شیراز ارائه می‌شود. بعد از واکاوی ماده‌ها که نامناسب بودند، ۱۰۰ ماده برای آزمون از همکاری ۵۳ تن از دانشجویان مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی بهره گرفته می‌شود. این دانشجویان در دو مقطع پایه و پیشرفته به یادگیری زبان فارسی در این مرکز اشتغال داشته و دارای ملیت‌های گوناگون بوده‌اند. بعد از اجرای آزمون، ماده‌های آزمون مورد واکاوی قرار می‌گیرد. برای تعیین پایانی این آزمون از دو روش پایانی دو نیمه و کودر ریچاردسون سود برده می‌شود و جهت تعیین روایی آزمون، روایی صوری، سنجی‌های و سازه‌ای آزمون مورد بررسی قرار می‌گیرد. سپس همبستگی میان بخش‌های مختلف آزمون اندازه‌گیری می‌شود. این پژوهش نشان می‌دهد که آزمودنی‌ها در مهارت خواندن و درک مفاهیم از کمترین توانایی و در حوزه قواعد دستوری از بیشترین توانایی برخوردارند.

قسنولی (۱۳۸۹) نیز ضمن ساخت یک نمونه آزمون بسندگی زبان فارسی، برای روایی سنجی آزمون خویش از فرایندی دو مرحله‌ای شامل تحلیل داده‌های بازنگری سه دانشجوی خارجی دوره دکتری زبان و ادبیات فارسی و بهره‌گیری از نظریه بحث در سنجش روایی بهره می‌گیرد. پایایی آزمون خویش را به شیوه آلفا کرونباخ محاسبه می‌کند و نیز به لحاظ ابعاد مهم و تاثیرگذار آزمون‌های بسندگی در قلمرو فرهنگی و سیاسی کشورها، به تولید و روایی سنجی آزمون ملی بسندگی زبان فارسی توجه خاصی معطوف می‌نماید.

جلیلی (۱۳۹۰) نیز در پژوهش خود سعی در تولید آزمون بسندگی زبان فارسی دارد که مهارت زبان آموزان را در عرصه کاربردی مورد سنجش قرار دهد. وی که آزمون خود را آمفا می‌نامد چهار مهارت زبانی اصلی را مورد آزمون قرار می‌دهد. وی مدعی است آزمونش از سنجش آشکار و مستقیم مقوله‌های واژگان، تلفظ و دستور زبان پرهیز می‌کند و این مقوله‌ها را ملزومات موفقیت در چهار مهارت اصلی می‌داند که میزان تسلط زبان آموزان بر آنها هنگام سنجش همان چهار مهارت، بروز و ظهور خواهد کرد. وی در خاتمه آزمون خود را مشابه آزمون آیلتس در زبان انگلیسی می‌داند.

¹ Validity

² Reliability

³ Ability

⁴ Grammar

حسینی (۱۳۹۱) نیز در پژوهشی دیگر، ضمن معرفی آزمونی به چگونگی طراحی آن در زبان فارسی به منظور اندازه گیری میزان بسندگی زبان فارسی می‌پردازد. در این پژوهش متون به صورت قاعده مندی انتخاب می‌شوند بدین معنا که از جمله عوامل موثر در انتخاب متون، تناسب متون با سن، میزان علاقه مندی و سطح بسندگی آزمودنی‌ها می‌باشد. نیمی از حروف هر دومین کلمه‌ی به کار رفته در جمله‌ی دوم به بعد در هر متن حذف شده و به یک نمونه آزمون تک تبدیل می‌شود. پایایی آزمون به شیوه‌ی آلفا کرونباخ محاسبه می‌شود که نتیجه بسیار مطلوبی (۰,۸۴) برای این آزمون تلقی می‌شود. واکاوی داده‌ها نشان می‌دهد این آزمون علاوه بر داشتن روایی ظاهری از روایی درونی (درجه تمیز: ۰,۲۳، درجه آسانی: ۰,۷۴) بالایی نیز برخوردار است.

مراصد صحرایی و جلیلی (۱۳۹۱) نیز پس از پرداختن به اهمیت وجود آزمون بسندگی زبان فارسی می‌کشند مبانی نظری و قالب کلی این آزمون را برای سنجش و ارزیابی مهارت زبانی فارسی آموزان خارجی ارائه نمایند. جلیلی و خدادادیان (۱۳۹۲) ضمن تاکید به ضرورت رویکرد نظام‌مند به آزمون سازی در زبان فارسی، ابتدا ۱۵۱ پاسخنامه مربوط به یکی از آزمون‌های پایان دوره‌ی خواندن و درک مطلب زبان فارسی سطح پایه (مقدماتی)، در مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی (ره) را بررسی کرده و پس از محاسبه‌ی ضریب سهولت هر یک از ۲۸ پرسش این آزمون، این ضرایب سهولت را با برآورد های ۳۰ نفر از مدرسان مقایسه می‌کنند. نتایج نشان می‌دهد که صرفاً با اتکا به تجربه یا برآوردهای شمی مدرسان نمی‌توان گفت که پرسش‌های یک آزمون، سخت، آسان یا ایده‌آل هستند و این نشان از ضرورت نهادینه سازی رویکرد علمی-آماری و نظام‌مند به آزمون سازی، در زمینه‌ی ارزشیابی و سنجش دانش زبانی فارسی آموزان دارد.

زند و همکاران (۱۳۹۳) ضمن توجه توأمان به مهارت شنیداری و گفتاری و درک ارتباط تنگاتنگ این دو با یکدیگر در تعاملات زبانی به دشوار بودن اندازه‌گیری درک شنیداری بدون در نظر گرفتن سایر مهارت‌های زبانی فارسی آموزان اذعان می‌کنند. آنان در پژوهش خود به دنبال یافت پاسخ این سوال هستند: آزمونی که گزینه‌های آن در قالب نوشتاری ارائه می‌گردد، برای ارزیابی مهارت شنیدن مناسب‌تر است یا آزمونی که گزینه‌ها در آن به صورت گفتاری طراحی و اجرا می‌گردد؟ بدین منظور، برای سنجش درک شنیداری، آزمونی طراحی می‌شود و برای اجرای این پژوهش، گزینه‌های آزمون در دو قالب نوشتاری و گفتاری ارائه می‌شود. نمونه‌ی آماری این پژوهش را ۳۶ فارسی آموز غیرایرانی مرکز آموزش زبان فارسی دانشگاه قزوین در رده‌ی سنی ۱۸ تا ۲۵ سال تشکیل می‌دهند که از لحاظ زبانی در سطح پیشرفته قرار دارند و به‌طور تصادفی انتخاب شده‌اند. روش انجام تحقیق از نوع میدانی بوده و برای تجزیه و تحلیل داده‌ها از نرم‌افزارهای آماری استفاده می‌شود. یافته‌های پژوهش نشان می‌دهد در مقایسه با قالب گفتاری، دیدگاه آزمودنی‌ها نسبت به دیدگاه نوشتاری، مثبت‌تر و رضایت‌مندی آنان نیز از این نوع قالب ارائه گزینه‌ها بیشتر است، اما آزمودنی‌ها در پاسخ به گزینه‌ها در قالب گفتاری عملکرد بهتری داشته و موفقیت بیشتری به دست آورده‌اند.

عباسی (۱۳۹۳) نیز با مقاله‌ای که در اولین همایش آموزش زبان و ادبیات فارسی آرایه می‌نماید گزارش می‌دهد دو نسخه آزمون برای زبان فارسی طراحی نموده و پس از برگزاری آزمایشی و اصلاحات نهایی، استاندارد سازی نموده است. آزمون حاوی پانصد سوال در بانک اطلاعاتی است که آزمون را هم می‌توان بصورت حضوری و هم مجازی اجرا نمود.

شایسته فر و بی‌جن خان (۲۰۱۶) نیز ساختار فاکتوری بخش شنیداری یک آزمون بسندگی زبان فارسی با نام PLPT را مورد بررسی قرار می‌دهند. آنها با استفاده از نرم افزار AMOS نسخه‌ی ۱۸ از مدل‌سازی معادله ساختاری^۱ استفاده می‌کنند تا پاسخ‌های ۱۲۰ زبان آموز فارسی را که در این آزمون شرکت کرده‌اند تحلیل نمایند. به منظور بررسی اینکه آیا عامل شنیداری با ساختار آزمون تطابق دارد، از سه مدل استفاده می‌شود. نتایج آزمون مدل سازی نشان می‌دهد مدل همبستگی مدعی وجود رابطه بین مهارت‌های خواندن و شنیدن، بهترین مدل‌سازی بر اساس اطلاعات موجود است.

جلیلی (۱۳۹۷) نیز سوالات طراحی شده در مرکز آموزش زبان فارسی دانشگاه قزوین را در کتابی تحت عنوان نمونه آزمون‌های زبان فارسی برای فارسی آموزان غیر ایرانی جمع آوری کرده است. این کتاب از دو فصل تشکیل شده است که از نظر دشواری به دو سطح یک و دو تقسیم شده است. هر سطح داری سه آزمون خواندن، دو آزمون شنیدن و سه آزمون نگارش می‌باشد.

گلپور (۱۳۹۷) نیز با توجه به اهمیت نوشتار در غیرفارسی زبانان، در پژوهشی آزمونی برای سنجش مهارت نوشتاری

^۱ Structural equation modeling

تدوین می‌کند. وی اهمیت زیادی برای مهارت نوشتن قائل است و معتقد است مهارت نوشتاری توانایی خلق کلمات و بیرون کشیدن ایده‌ها از ذهن و همچنین توانایی انتقاد کردن از آن می‌باشد. وی در پژوهش خویش سعی می‌کند راهکارهای ارزشیابی مهارت نوشتاری و کنش‌های نوشتاری مانند کنش تقلیدی، فشرده، واکنشی، گسترده و انواع سؤالات آزمون نگارش از قبیل سؤال پایه، سؤال چارچوب دار و سؤال متن محور را مورد بررسی قرار دهد. با توجه به اهمیت خطاهای نظام‌مند فارسی‌آموزان در امر سازمان‌دهی مواد آموزشی، انتخاب روش تدریس مناسب و آزمون سازی، ۲۰ خطای نحوی فارسی آموزان که دارای بسامد بالایی در برگ نوشته‌های فارسی زبانان است، مورد واکاوی قرار می‌گیرد که شامل حذف عناصر دستوری و واژگانی، اضافه کردن عناصر غیر ضروری یا نادرست، انتخاب یک عنصر نادرست و جایگزین کردن آن با عنصر در ست و چینش و ترتیب نادرست عناصر می‌باشد که باید در امر سازمان‌دهی مواد آموزشی، انتخاب روش تدریس مناسب و آزمون سازی به آن توجه کرد.

نقیبان و همکاران (۱۳۹۷) با توجه به مهارت خواندن در زبان فارسی و با هدف ساخت و رواسازی آزمون خواندن با استفاده از سنجش شناخت تشخیصی، اقدام به تولید آزمون می‌نمایند. ابتدا با دسته‌بندی خصیصه‌های مرتبط با خواندن در تحقیقات، اسناد ملی و بین‌المللی، مهم‌ترین خصیصه‌های خواندن را مشخص می‌کنند. پس از آن با استفاده از نظر متخصصان زبان و ادبیات فارسی و معلمان ابتدایی، آزمون اولیه را طراحی می‌کنند. پس از اجرای مقدماتی و تحلیل‌های اولیه و حذف پرسش‌های نامناسب، در نهایت آزمون نهایی با ۲۷ پرسش طراحی می‌گردد که چهار خصیصه‌ی درک اطلاعات صریح متن، استنباط، تفسیر، دانش واژگان را مورد سنجش قرار می‌دهد. با استفاده از روش نمونه‌گیری خوشه‌ای، آزمون روی ۸۱۶ نفر از دانش‌آموزان پایه‌ی ششم ابتدایی شهر قم اجرا می‌شود. تحلیل‌های آماری با استفاده از مدل شناختی تشخیصی افزایشی انجام می‌شود. شاخص‌های نسبی و مطلق برازش، نشان‌دهنده‌ی برازش مناسب مدل است. همچنین استفاده از شاخص‌های همسانی و دقت طبقه‌بندی، نشان‌دهنده‌ی همسانی و دقت مناسب طبقه‌بندی دانش‌آموزان توسط آزمون است. نتایج این تحقیق نشان می‌دهد این آزمون می‌تواند نیم‌رخ خصیصه‌های خواندن دانش‌آموزان را با استفاده از مدل سنجش شناختی تشخیصی شناسایی کند. لازم به ذکر این است این آزمون بر خلاف سایر آزمون‌های رایج، بر روی فارسی زبانان که زبان فارسی زبان مادری آنان است انجام می‌شود و نه غیر فارسی زبانان.

اولین آزمون رسمی استاندارد بسندگی زبان فارسی با عنوان سامفا در بهمن ماه ۱۳۹۷ به عاملیت سازمان سنجش و در دانشگاه‌های تهران، بین‌المللی امام خمینی، فردوسی مشهد، اصفهان و همدان برگزار می‌گردد. این آزمون به تقلید از آزمون آیلتس در زبان انگلیسی و DILF/DELF/DALF (فرانسسه) و ZD (آلمانی) در نوع دانشگاهی و عمومی طراحی شده است و چهار مهارت خواندن، نوشتن، شنیدن و صحبت کردن را می‌سنجد. علاوه بر این، نوع دانشگاهی نیز به چهار زیرشاخه علوم پایه و مهندسی، علوم انسانی، علوم پزشکی و علوم اجتماعی تقسیم شده است که هر داوطلب متناسب با رشته تحصیلی خویش در آن آزمون شرکت می‌کند.

از آغاز سال تحصیلی ۹۹، شرکت در این آزمون برای تمامی دانشجویان خارجی داوطلب تحصیل در ایران اجباری است و نیز خارجیانی که قصد تاسیس شرکت یا تجارت در ایران دارند نیز ملزم به شرکت در این آزمون می‌باشند. سامفا دارای ۲۴۰ نمره می‌باشد که به هر مهارت ۶۰ نمره اختصاص یافته است. بخش شنیداری و خواندن هر کدام ۳۰ سوال دارد و در قسمت نوشتاری دو موضوع برای انشا نویسی داده می‌شود (مجموعاً سه ساعت). علاوه بر این، داوطلبان ملزم به شرکت در مصاحبه هستند که به مدت ۱۵ دقیقه طول می‌کشد. البته به دانشگاهها اجازه داده شده است نمره قبولی را بر حسب شرایط خود تعیین نمایند. این آزمون سالی دو بار برگزار می‌شود و قرار است آزمون بعدی همزمان در دانشگاههای داخل و نمایندگی‌های آموزش زبان فارسی در خارج از کشور برگزار شود.

۳. الگوی مورد استفاده در این پژوهش

(Brown, 2004) شش مرحله را برای تولید یک آزمون پیشنهاد می‌کند که در شکل یک به تصویر کشیده شده است و برای تولید این آزمون بسندگی از همان شش مرحله استفاده شده است. هر یک از مراحل را در زیر شرح خواهیم داد.

تعیین هدف ← تعیین مختصات ← تولید ← ارزیابی ← تعیین نحوه تصحیح ← تعیین روایی و پایایی.

شکل ۱: مراحل تولید آزمون

۳-۱ تعیین هدف آزمون

تولید این آزمون بسندگی سه هدف را دربر داشت: اعطا یا عدم اعطای پذیرش به دانشجویان خارجی داوطلب در دانشگاه محل خدمت پژوهشگران، تعیین نقاط ضعف و قوت دانشجوی خارجی در زبان فارسی و آرایه کلاسهای متناسب و بلاخره تکرار این آزمون در انتهای دوره زبان آموزی به منظور تعیین میزان موفقیت فارسی آموزان در دوره فارسی آموزی. پس از بحث‌های فراوان و گرفتن نظر مشورتی اساتید درگیر، در مجموع مقرر گردید آزمون دارای سه مهارت دستور، خواندن و شنیدن باشد چراکه این سه مهارت نیازهای مطروحه در بالا را تامین می‌کند.

علاوه بر نکات بالا، دو محدودیت زبانی نیز وجود داشت که بایستی مورد توجه قرار می‌گرفت. نکته اول این بود که بایستی نتیجه آزمون فوراً به اطلاع کلیه دانشجویان داوطلب ادامه تحصیل در آن دانشگاه می‌رسید چراکه فاصله زمانی بین شرکت در آزمون و ثبت نام و حضور در کلاس نسبتاً کم بود. دوم مقدار زمانی که قابل اختصاص به این آزمون بود خیلی زیاد نبود.

با این محدودیت‌هایی که توسط عوامل اجرایی بر پژوهشگران تحمیل شده بود، تصمیم به تولید یک آزمون چهارجوابی دارای شصت سوال گرفته شد. هر تست دارای چهار گزینه بود که یکی از آنها صحیح بود. از گزینه چهارجوابی استفاده شد چراکه امکان برگزاری آزمون بصورت الکترونیکی در مرکز آزمون سازی دانشگاه مهیا بود و نتایج سریعاً توسط سیستم محاسبه و اعلام می‌شد. طبق نظر (Bailey, 1998) مزایای آزمون‌های چهار گزینه‌ای، سرعت در تصحیح، مقرون به صرفه بودن و عدم اعمال سلیقه مصحح در هنگام تصحیح آزمون می‌باشد.

۳-۲ تعیین مختصات آزمون

همانطور که گفته شد این آزمون قصد سنجش سه مهارت یعنی دستور، خواندن و شنیدن را داشت. بخش دستور که ۲۰ درصد از نمره کل آزمون را به خود اختصاص می‌داد شامل جملاتی بود که دارای یک یا دو جای خالی بود و چهار گزینه برای جای خالی پیشنهاد شده بود. بخش خواندن ۴۰ درصد از امتیاز را به خود اختصاص داده بود و دارای دو متن کوتاه بود که هر متن دارای پنج سوال درک مطلب بود. یک متن طولانی نیز وجود داشت که دارای ده سوال درک مطلب بود. در مجموع ۲۰ سوال وجود داشت که دارای ۴۰ امتیاز بود. بخش شنیدن نیز ۴۰ درصد از کل نمره را به خود اختصاص داده بود که شامل سه قسمت بود: مکالمه کوتاه (۷ سوال)، متن کوتاه (۷ سوال) و پاسخ مناسب (۶ سوال). جدول شماره یک، اطلاعات بالا را خلاصه می‌کند.

جدول ۱: قسمت‌های در نظر گرفته شده در آزمون

ردیف	نام مهارت	فرمت تست	تعداد زیر قسمت	تعداد سوال	درصد	زمان (دقیقه)
۱	دستور	سوال چهار جوابی	۱	۳۰	۲۰	۳۰
۲	خواندن	سوال چهارجوابی	۳	۲۰	۴۰	۴۰
۳	شنیدن	متن شنیداری و سوال چهار جوابی	۳	۲۰	۴۰	۴۰
کل	۳	چهارجوابی	۷	۷۰	۱۰۰	۱۱۰

نکات زیر در هنگام طرح سوالات چهارجوابی مورد توجه قرار گرفت: (۱) هر تست فقط یک نکته را بسنجد، (۲) صورت سوال و چهار گزینه ساده و قابل فهم باشد، (۳) فقط یک پاسخ صحیح وجود داشته باشد و (۴) هر چهار گزینه از نظر واژگانی و دستوری صحیح باشند، دارای ساختار دستوری یکسان بوده و طول چهار گزینه حتی الامکان برابر باشد.

۳-۲-۱ بخش دستور

(Hughes, 2003) معتقد است نتایج آزمون دستور به ما امکان می‌دهد زبان آموزان را به سطوح مختلف دسته بندی کرده و در کلاسهای زبان آموزی مناسب قرار دهیم. علاوه بر این، آزمون دستور به ما نشان می‌دهد زبان آموزان در چه نکات دستوری

دارای ضعف می‌باشند. هدف از بخش دستور، سنجش توانایی زبان آموز در تشخیص جملات دستوری در زبان استاندارد فارسی بود. تست‌های طراحی شده به شکل صحیح فعل، زمان فعل، تطابق فعل و فاعل، حرف ربط مناسب، کاربرد صحیح اجزای کلام (اسم، فعل، صفت، قید) شکل صحیح عبارت (اسمی) و نحوه چینش صحیح توصیف‌کننده‌ها قبل از اسم اختصاص داشت. قسمت دستور به گونه‌ای طراحی شد تا بر اساس نتایج آزمون، بتوان زبان آموزان را در کلاس‌های مختلف قرار داد. انتخاب واژگان، عبارات و نکات دستوری، بصورت تصادفی نبود بلکه بر اساس مقوله‌هایی بود که توسط (Chapelle and Abraham, 1990) مطرح شده بود.

۲-۲-۳ بخش خواندن

بخش خواندن دارای دو متن کوتاه بود که هر کدام دارای ۲۰۰ الی ۳۰۰ کلمه بود و یک متن بلند که دارای بیش از ۵۰۰ کلمه بود. متن اول نسبتاً ساده، متن دوم متوسط و متن سوم نسبتاً دشوار بود. علت این ترتیب، امکان جدا کردن زبان آموزان ضعیف از زبان آموزان قوی در مهارت خواندن بود. متون دارای (۱) مقدمه‌ای ساده و قابل فهم (۲) پاراگرافهایی که ایده اصلی در جمله اول بیان شده بود و (۳) دارای نتیجه‌گیری در پاراگراف آخر بود. نوع متون توصیفی بود و مشابه متونی بود که در مجلات یا کتب درسی یافت می‌شد و به موضوعات علمی می‌پرداخت (یعنی حالت محاوره‌ای نداشت و فاقد اصطلاحات عامیانه یا گفتاری بود). علاوه بر این، طبق توصیه آلدرسون (۲۰۰۰) متون به حقایق علمی می‌پرداخت، دارای اطلاعات مفید بود و بصورت توصیفی نگاشته شده بود. ابتدای هر پاراگراف دارای فرو رفتگی بود و پاراگرافها شماره گذاری شده بود و کلمات مهم مورد سوال با زیر خط کشی نشان داده شده بود.

طبق نظر آلدرسون (۲۰۰۰) توانایی خواندن خود از زیر مهارتهای مختلفی تشکیل شده است که آن را می‌توان هم با سوالات ریز و هم سوالات کلی مورد سنجش قرار داد. سوالات کلی به سوالاتی اشاره دارد که درک کلی زبان آموز از متن را می‌سنجد و سوالات ریز به درک زبان آموز از کلمات یا جملات خاص اشاره دارد (براون، ۲۰۰۴؛ هیوز، ۲۰۰۳). متون بخش خواندن دارای این سوالات کلی بود: منظور کلی متن، منظور اصلی هر پاراگراف و استنتاج. سوالات ریز نیز شامل این موارد میشد: جزئیات و معنی کلمه با توجه به متن.

۳-۲-۳ بخش شنیداری (Brindley, 1998) معتقد است اگر تمرینات متنوعی در بخش شنیداری گنجانده شود می‌توان تصویر واضحی از توانش شنیداری زبان آموز بدست آورد. لذا بخش شنیداری این آزمون از سه قسمت تشکیل شده بود: مکالمه کوتاه، متن کوتاه و پاسخ مناسب. هدف از مکالمه کوتاه و متن کوتاه، تست درک کلی شنیداری بود. هدف از بخش پاسخ مناسب، آرایه پاسخ مناسب با آنچه شنیده می‌شد بود. طبق نظر (Buck, 2001) اگر تست شنیداری شامل این سه قسمت باشد نشان می‌دهد زبان آموز (۱) توانایی درک گفتار عادی زبان را دارد (۲) می‌تواند ایده اصلی متن را تشخیص دهد (۳) اطلاعات مشخص داخل متن را تشخیص دهد و (۴) از متن، استنباط انجام دهد.

۳-۳ تولید تست

با توجه به فقدان چنین تست‌هایی و نبود بانک سوالاتی که بتوان از آنها چنین تست‌هایی را استخراج کرد، تولید تست بر عهده پژوهشگران بود. ابتدا نمونه سوالاتی تهیه شد. سپس مدرسین مرکز آموزش زبان فارسی به غیر فارسی زبانان (آزفا) و دپارتمان آموزش زبان انگلیسی آنها را بررسی کردند. تست‌ها به کمیته هماهنگی متشکل از سه مدرس آزفا که جزو تولیدکنندگان تست نبودند داده شد تا آنها نیز تست‌ها را مورد بررسی قرار دهند. در انتها پس از انجام یک تست آزمایشی، برخی از تست‌ها حذف و در برخی تغییراتی اعمال گردید.

۴-۳ ارزیابی و بازبینی سوالات

ابتدا سوالات به یک گروه آزمایشی با تعداد ۵۳ دانشجوی خارجی شاغل به تحصیل در دانشگاه داده شد و میزان دشواری، موفقیت تست در ایجاد تمایز بین دانشجویان ضعیف و قوی و نیز عملکرد گزینه‌های اشتباه مورد بازبینی قرار گرفت. در

محاسبه میزان دشواری فقط سوالاتی که بین ۵۵ الی ۶۶ درصد دارای دشواری بودند انتخاب و سوالات پائین تر و بالاتر از این مقدار حذف گردیدند. علاوه بر این، چینش سوالات برای آزمون نهایی بر اساس دشواری سوالات از آسان به سخت قرار داده شد تا چنانچه زبان آموزانی دچار کمبود وقت شوند نتوانند به سوالات دشوارتر پاسخ دهند.

۳-۵ تعیین نحوه تصحیح آزمون

بخش دستور دارای ۳۰ سوال بود و ۲۰ درصد امتیاز را به خود اختصاص داده بود. بخش خواندن ۴۰ درصد امتیاز کل آزمون را داشت و حاوی ۲۰ سوال دو امتیازی بود. بخش شنیداری نیز دارای ۴۰ درصد بود که بخش مکالمه کوتاه دارای ۷ سوال، متن کوتاه ۷ سوال و پاسخ مناسب ۶ سوال بود که جمعاً ۴۰ نمره داشت. آزمون بصورت الکترونیکی برگزار می‌گردید و سیستم بصورت خودکار نمرات را محاسبه می‌کرد.

۳-۶ تعیین روایی و پایایی آزمون

با توجه به سرنوشت ساز بودن نتایج این آزمون، لازم بود روایی و پایایی این آزمون محاسبه می‌گردید. در زیر نحوه محاسبه روایی و پایایی این آزمون توضیح داده می‌شود.

۳-۶-۱ روایی

روایی مفهومی پیچیده است ولی وجود آن برای تست فوق العاده مهم می‌باشد. بطور خلاصه می‌توان گفت آزمون دارای روایی، آزمونی است که آن چیزی را تست می‌کند که قرار است تست کند. برای اطمینان از وجود روایی در این آزمون از سه روش بهره گرفته شد. ابتدا از روایی محتوایی^۱ استفاده شد. سپس با استفاده از تحلیل فاکتور اکتشافی^۲ (EFA) و تحلیل فاکتور تاییدی^۳ (CFA)، روایی ساختاری^۴ مورد بررسی قرار گرفت. و در انتها با استفاده از همبستگی مقایسه‌ای متقاطع^۵، مقایسه‌ای بین این آزمون و سایر آزمون‌های دیگر انجام شد تا روایی همزمانی^۶ آزمون محاسبه شود.

۳-۶-۲ روایی محتوایی چهارمین همایش ملی آموزش زبان فارسی - ۱۳۹۹

طبق نظر (Hughes, 2003) اگر سه نفر از کارشناسان یک حوزه، آزمون را مورد بررسی قرار دهند و نظر مثبت خود را در خصوص آزمون اعلام نمایند می‌توان از روایی محتوایی آزمون اطمینان حاصل نمود. این آزمون به سه نفر از مدرسین دپارتمان آموزش زبان انگلیسی که جزو تیم طراح سوال نبودند و با مرکز آموزش فارسی به غیر فارسی زبانان همکاری داشتند داده شد و هر سه نظر مثبت خود را در خصوص روایی محتوایی اعلام داشتند. به عبارت دیگر به این جمع بندی رسیدند که آزمون همان مهارتهای دستوری، خواندن و شنیداری را تست می‌کند.

۳-۶-۳ روایی ساختاری

از تحلیل فاکتور برای تعیین روایی ساختاری آزمون استفاده شد. تحلیل فاکتور شامل یک فرایند سه مرحله‌ای است. ابتدا EFA انجام شد تا بهترین ساختار فاکتوری آزمون تعیین گردد. سپس بهترین راه حل از EFA با CFA مورد سنجش قرار گرفت. و در نهایت، با استفاده از شاخص برازش تطبیقی^۷ (CFI)، شاخص نیکویی برازش^۸ (GFI) و تقریب خطای جذری^۹ (RMSEA)،

¹ Content validity

² Exploratory factor analysis

³ Confirmatory factor Analysis

⁴ Construct validity

⁵ Cross-comparison correlation

⁶ Concurrent validity

⁷ Comparative fit index

⁸ Goodness of fit index

⁹ Root square error approximation

تغییر ناپذیری فاکتوریل^۱ مورد ارزیابی قرار گرفت تا انحرافات موجود در مدل بدست آمده مشخص گردد. طبق نظر (EFA (Bachman, 2004; Shin, 2005) فرایندی آماری است که هدف آن بررسی خوشه ها^۲ یا الگوهای اشتراک در متغیرهای قابل مشاهده است. هر بخش از این آزمون متناسب با ویژگیهای مد نظر طراحی گردید. پس سوالات آزمون بر حسب آن چیزی که می سنجد به متغیرهای مختلف دسته بندی شد. هدف از طراحی این آزمون سنجش سه مهارت بود: دستور (G)، خواندن (R)، شنیدن (L) و زیر ساختارهایی که دستور دارای زیر ساختار نبود (G1) و سه فاکتور خواندن که حاوی سه متن بود R1 R2 R3 و سه فاکتور شنیداری که دارای سه قسمت بود: مکالمه کوتاه^۳ (LSD)، متن کوتاه^۴ (LSP) و پاسخ مناسب^۵ (LAR). جدول دو حاوی متغیرهای این آزمون است.

جدول ۲: متغیرهای بکار رفته در این آزمون

ردیف	متغیر	شماره تست
۱	دستور (G1)	۱-۳۰
۲	خواندن متن اول (R1)	۳۱-۳۵
۳	خواندن متن دوم (R2)	۳۶-۴۰
۴	خواندن متن سوم (R3)	۴۱-۵۰
۵	شنیدن مکالمه کوتاه (LSD)	۵۱-۵۷
۶	شنیدن متن کوتاه (LSP)	۵۸-۶۴
۷	شنیدن پاسخ مناسب (LAR)	۶۵-۷۰

فرایند استاندارد EFA به شرح زیر بود. در گام نخست ماتریس همبستگی گشتاوری^۶ در بین متغیرها تعیین گردید. سپس از تحلیل مولفه‌های اصلی^۷ استفاده شد تا فاکتورهای اولیه استخراج گردد. نمودار سنگ ریزه^۸ و مقدار عددی نقطه عطف^۹ بدست آمده از استخراج‌های اولیه مورد بررسی قرار گرفت تا تعداد فاکتورهای موجود تعیین گردد. سپس از محورهای اصلی^{۱۰} استفاده شد تا تعداد فاکتورهای برابر با یکی بالا و یکی پائین تر از تعداد فاکتورهای مشخص شده توسط آرنج نمودار سنگ ریزه^{۱۱} تعیین گردد. این استخراج‌ها چرخش داده شدند^{۱۲} تا راه حل‌های قائم و مورب^{۱۳} بدست آید. گام آخر تعیین تعداد بهینه فاکتورها از ساختارهای ساده و تفسیرهای معنی دار بود. بر اساس نتایج EFA، سه فاکتور استخراج گردید و از راه حل سه فاکتوری برای تعیین قابلیت توضیح و درک داده‌ها استفاده شد. این سه فاکتور بر حسب الگوهای باری فاکتور، دسته بندی گردید. فاکتور اول فاکتور دستور بود چراکه این متغیر دستور روی فاکتور اول وزن سنگینی داشت. فاکتور دو، فاکتور خواندن بود که سه فاکتور خواندن دارای بار زیادی بود. فاکتور سوم، فاکتور شنیداری بود چراکه سه متغیر شنیداری وزن سنگینی روی این فاکتور سوم داشتند. نتایج FEA برای این آزمون در جدول سوم ارائه گردیده است.

¹ Factorial invariance

² Clusters

³ Listening short dialogue

⁴ Listening short passage

⁵ Listening appropriate response

⁶ Matrix of product-moment correlation

⁷ Principle components analysis

⁸ Scree plot

⁹ Eigenvalue

¹⁰ Principle axes

¹¹ Elbow of the scree plot

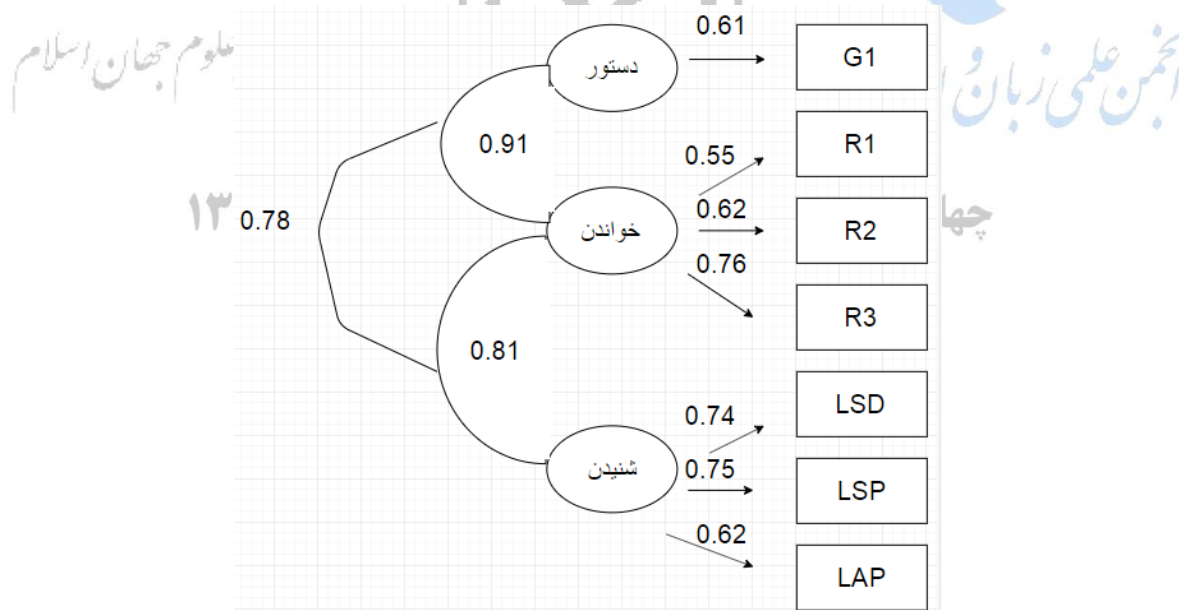
¹² Rotate

¹³ Orthogonal and oblique solutions

جدول ۳: نتایج EFA آزمون

فاکتور			متغیر
۳	۱	۱	دستور (G1)
۰,۳۱	۰,۳۳	۰,۶۳	
۰,۱۸	۰,۸۴	۰,۰۴	خواندن متن اول (R1)
۰,۱۴	۰,۶۴	۰,۳۸	خواندن متن دوم (R2)
۰,۴۱	۰,۵۳	۰,۴۱	خواندن متن سوم (R3)
۰,۷۴	۰,۲۹	۰,۲۱	شنیدن مکالمه کوتاه (LSD)
۰,۷۰	۰,۳۱	۰,۲۳	شنیدن متن کوتاه (LSP)
۰,۸۵	۰,۰۵	۰,۱۴	شنیدن پاسخ مناسب (LAR)

از CFA برای تعیین موارد زیر استفاده گردید: (۱) بارهای هر مسیر که از یک ساختار به یک متغیر مشاهده شده منتهی می‌گردد (۲) کل واریانس هر متغیر مشاهده شده که با یک متغیر متناظر نهفته توضیح داده می‌شود (R^2) و (۳) نشانگرهای انطباق مدلی^۱ بخصوص GFI, CFI, RMSEA. مدل سه فاکتوری بدست آمده از EFA و بارگذاری مسیری^۲ بدست آمده از CFA در شکل دو نشان داده شده است. همبستگی بین فاکتورها و متغیرها (بارگذاری مسیر) بین ۰,۵۵ و ۰,۷۸ متغیر بود. بارگذاری‌های مسیری خواندن (R1: 0.55; R2: 0.62; R3: 0.76) از بخش اول آزمون به بخش سوم رو به افزایش بود. بخش‌های شنیداری دارای بیشترین بارگذاری مسیری بودند (LSD: 0.74; LSP: 0.75; LA: 0.62). شاخص‌های GFI و CFI به ترتیب (0.995 و 0.996) بودند که بالاتر از آستانه ۰,۹۰ بودند در حالیکه مقدار RMSEA حدود ۰,۰۲۶ بود. این مقدار، پائین‌تر از آستانه 0.05 بود (بیرن، ۱۹۹۴). نتایج این سه شاخص نشان می‌دهد که مدل ساختاری سه فاکتوری، الگوی مناسبی برای مجموعه داده‌ها می‌باشد.



شکل ۲: بارگذاری مدل سه فاکتوری

هر دوی BFA و CFA نشان می‌دهند این آزمون سه مهارت را بدرستی مورد سنجش قرار داده است: دستور، خواندن و شنیدن. علاوه بر این، ساختار فاکتوری نشان داد که در این آزمون، متغیرهای مربوطه بدرستی مورد سنجش قرار گرفته‌اند: سی تست دستوری، دستور را سنجیده‌اند؛ سه متن خواندن، خواندن را مورد ارزیابی قرار داده‌اند و سه بخش خواندن،

¹ Fit model indicators

² Path loading

خواندن را مورد ارزیابی قرار داده اند. این راه حل، تایید نمود ساختار پیشنهادی آزمون صحیح است و نشان داد نتایج بدست آمده از آزمون دارای روایی می‌باشند. به بیان ساده، ساختار آزمون بدست آمده از تحلیل فاکتور نشان داد که این آزمون همان چیزی را می‌سنجد که طراحان مد نظر داشته اند.

۴-۶-۳ روایی همزمانی

(Brown, 2004) پیشنهاد می‌کند که همبستگی آماری یک آزمون با آزمون مستقل دیگر، روش مطمئن دیگری برای تعیین روایی آزمون است. نتایج همبستگی مقایسه‌ای ۵۳ دانشجویی که هم در این آزمون شرکت کردند و هم درس فارسی عمومی را گذرانده بودند نشان داد همبستگی بالایی ($r=0.89$) بین این آزمون و آزمون فارسی عمومی وجود داشت. بخش شنیداری با ۰,۸۳ و دستور و خواندن با ۰,۸۳ نیز دارای ضریب همبستگی بالا بودند. تحلیل دیگری از همین دانشجویان با معدل پایان ترم آنها نیز باز نشان داد همبستگی بالا ($r=0.76$) بین این دو برقرار است. بطور خلاصه، نتایج همبستگی نشان داد آزمون بسندگی ما دارای روایی همزمانی بسیار بالا با نمره درس زبان عمومی و روایی همزمانی نسبتاً بالا با معدل پایان ترم دانشجویان داشت.

۵-۶-۳ پایایی

آزمونی دارای پایایی است که با استفاده مکرر از آن، نتایج یکسان بدست آید. به عبارت دیگر اگر زبان آموزی دوبار در یک آزمون شرکت کند و نتایج شبیه همدیگر باشد این آزمون دارای پایایی است. از روش دو نیمه آزمون^۱ برای تخمین پایایی محتوایی این آزمون و از روش آلفای کرونباخ برای تعیین پایایی واریانس تست‌ها استفاده شد. برای دو نیمه آزمون، آزمون به دو قسمت مساوی تقسیم شد. برای مثال در قسمت دستور اگر تستی زمان را مورد سنجش قرار میداد تست مشابهی که زمان را مورد سنجش قرار می‌داد در نیمه دیگر قرار داده شد. در بخش خواندن، یک تست که ایده اصلی متن را سوال می‌کرد در یک نیمه و تست مشابه در نیمه دیگر قرار گرفت. همین کار برای سوالات درک جزئیات و لغات داخل متن و سوالات استنتاجی انجام شد. همین کار برای بخش شنیداری نیز مورد استفاده قرار گرفت. ضریب همبستگی پایایی دو نیمه آزمون اسپیرمن براون مورد محاسبه قرار گرفت و $r=0.873$ بدست آمد و ضریب همبستگی آلفای کرونباخ $r=0.868$ گزارش شد. بر طبق نظر (Hughes, 2003) بر اساس مقادیر بالا این آزمون دارای پایایی بالایی بود.

۴. بحث

نتایج حاصل از EFA، راه حل سه فاکتوری و نتایج حاصل از CFA، مدل سه فاکتوری آرایه نمودند که بهترین انطباق را با داده‌های موجود داشتند. این مدل نشان داد ساختار پیشنهادی آزمون توسط طراحان تست (دستور، خواندن و شنیدن) ساختاری مناسب است و آزمون دارای روایی ساختاری است. نتایج تحلیل فاکتور نیز نشان داد سه بخش این آزمون کاملاً مستقل عمل کرده و هر کدام مهارتهای خاص خود را می‌سنجند. درحالیکه بسیاری از صاحب‌نظران معتقدند بخش شنیداری را می‌توان از خواندن و دستور جدا کرد (Bae & Bachman, 1998; Hale, Rock, & Jirele, 1989; Shin, 2005; Song, 2008) جداکردن دستور از خواندن مناقشات بیشتری را به خود جلب کرده است. برای مثال، (Tomblin and Zhang, 2006) دریافتند که ایندو از هم مستقل هستند درحالیکه (Romhild, 2008) دریافت ایندو بهم شدیداً مرتبط هستند. علت اینکه در پژوهش ما دستور و خواندن از همدیگر مستقل ارزیابی شدند را می‌توان در محتوای آزمون ما جویا شد. تست‌های قسمت خواندن به ایده‌های کلی، جزئیات و لغت در داخل متن می‌پرداختند و نیازی به درک عمیقی از دستور نداشتند. به عبارت دیگر، تست‌های دستوری بیشتر با شکل صحیح فعل، زمان فعل، تطابق فعل و فاعل، حرف ربط مناسب، کاربرد صحیح اجزای کلام (اسم، فعل، صفت، قید) شکل صحیح عبارت (اسمی) و نحوه چینش صحیح توصیف‌کننده‌ها قبل از اسم سروکار داشتند و اینها نقش زیادی در درک مطلب ایفا نمی‌کنند.

از روش دو نیمه آزمون و واریانس تست برای تخمین پایایی داخلی آزمون استفاده شد. اولاً از آنجائیکه این تست طوری طراحی شده بود تا مهارتهای مختلف (دستور، خواندن و شنیدن) و جنبه‌های زبانی متنوع (نکات دستوری، و مهارتهای

¹ Split half

شنیداری و خواندنی) را بسنجد، به نظر لازم می‌رسید از ضریب پایایی دو نیمه آزمون اسپیرمان براون (Bachman, 2004) استفاده شود. دوما از آنجائیکه این آزمون مهارتهایی را با استفاده از تست‌های مختلف بصورت موازی با واریانس‌های مشابه مورد سنجش قرار می‌داد، استفاده از روش آلفای کرونباخ منطقی به نظر می‌رسید (Bachman, 2004). به بیان ساده تر، روش دو نیمه آزمون، پایایی را بر اساس محتوای آزمون می‌سنجید و ضریب پایایی آلفای کرونباخ، پایایی را بر اساس واریانس تک تک تست‌ها مورد سنجش قرار میداد. همانطور که در بخش نتایج اعلام شد هر دو پایایی‌ها بر اساس محتوای تست ۰,۸۷۳ و بر اساس واریانس ۰,۸۶۸ بود که هر دو مقادیر بالایی است.

اگر سابقه طولانی تولید آزمون‌های بسندگی بخصوص در آمریکا نگاهی بیاندازیم درخواهیم یافت که هنوز در آغاز راه هستیم و نیاز به فعالیت‌ها و پژوهش‌های زیادی در این حوزه است و تولید آزمون بسندگی مورد بحث نیز دقیقا در همین راستا بوده است. جالب است بدانیم در زبان عربی نیز که هم از لحاظ گستره جغرافیایی و هم تعداد سخنوران در جایگاه بسیار مهمی قرارداد و حتی جزو شش زبان مورد استفاده سازمان ملل متحد نیز می‌باشد کاری جدی در این زمینه صورت نگرفته است. هم اکنون دو آزمون بسندگی زبان عربی در کشور یکی توسط دانشگاه فردوسی مشهد و یکی دانشگاه تهران برگزار می‌گردد که بحث در خصوص این دو آزمون در این مقال نمی‌گنجد. مرکز آزمون سازی دانشگاه الزهرا با یک موسسه اردنی با نام التنال قرارداد همکاری بسته و آزمون بسندگی زبان عربی برگزار نموده است که این آزمون نیز دارای مشکلات خاص خود می‌باشد بخصوص برگزاری الکترونیکی آن نیازمند زیرساختارهایی است که نبود آنها مشکلاتی را ایجاد می‌نماید.

چنانچه بتوان به آزمونهای بسندگی دیگر در خصوص زبان فارسی بخصوص سامفا دست پیدا کرد مسلما مقایسه نتایج این آزمونها با همدیگر نیز می‌تواند نکات بسیار خوبی را روشن کند که راهگشای تولید آزمون‌های بهتر برای زبان فارسی باشد. تجربه پژوهشگران در آزمون‌های بسندگی زبان انگلیسی همچون MSRT نشان داده است داوطلبان با بخش شنیداری مشکلات جدی دارند. لیکن نتایج این آزمون نشان داد شرکت کنندگان در این آزمون مشکل خاصی با شنیدن مطالب به فارسی نداشتند. شاید یکی از دلایل این امر وجود فناوری‌های نوین باشد که به افراد امکان تماشای محتوای مختلف به زبانهای مختلف را بدهد و چون شرکت کنندگان در این آزمون همگی خارجی بوده اند ممکن است آشنایی بیشتری با جنبه‌های شنیداری زبان فارسی داشته اند.

در خصوص خواندن همانطور که می‌دانیم این مهارت یکی از اساسی ترین مهارتها در هر زبان بخصوص برای دانشجویان است. لیکن نکته‌ای که در اینجا مطرح می‌شود مرتبط بودن آزمون درک مطلب با رشته‌ی تحصیلی دانشجویان است. همانطور که گفته شد آزمون سامفا دارای متون مرتبط با رشته تحصیلی شرکت کنندگان است. در آزمون ما سعی شد متون، عمومی انتخاب شود لیکن ما تجربه آزمون‌های بسندگی مختلفی مثل آیلتس را داریم که دو نمونه عمومی و دانشگاهی دارند. لذا نیاز به پژوهش‌های بیشتری در این حوزه می‌باشد تا اولاً نسخ مختلف این آزمون (عمومی و دانشگاهی) تولید شود و دوما بررسی شود چرا وزارت علوم اقدام به تولید نسخه متفاوت دانشگاهی برای رشته‌های مختلف نموده است. این کار تقریبا در آزمونهای بین المللی نظیر تافل و آیلتس سابقه نداشته است و پژوهش در خصوص آن مسلما نکات خوبی را روشن خواهد ساخت.

در خصوص دستو، اگر نگاهی به نسخه جدید آزمون‌های بسندگی زبان انگلیسی بیاندازیم متوجه حذف این بخش می‌شویم. در آزمون ما بخش دستور وجود داشت و دلایل گنجاندن آن نیز به تفصیل بیان شد. سنجش دستور زبان آن هم به شیوه‌ی سنتی آن، در حال حاضر منسوخ شده‌است و در آزمون‌های امروزی تافل و آیلتس، دستور زبان بدین شیوه تست نمی‌شود. تست‌های توانش زبانی امروزی بیشتر به کاربردی بودن نزدیک شده‌اند. لذا لازم است در این خصوص بررسی‌های بیشتری صورت پذیرد و به نظر می‌رسد بر اساس نتایج این پژوهش و حذف بخش دستور زبان در آزمون‌های بین المللی، گنجاندن تست‌های دستور زبان آن هم بدین شیوه در آزمون‌های کنونی، جای تامل و بررسی دارد.

در حال حاضر دانشجویان خارجی در مقطع تحصیلات تکمیلی ملزم به نگارش پایان نامه به زبان فارسی هستند. بحث‌هایی در خصوص اجازه دادن به این دانشجویان برای نگارش پایان نامه به زبان انگلیسی یا زبان مادری خود مطرح گردیده است که دارای مخالفین و موافقین خاص خود است و هنوز به نتیجه قطعی نرسیده است. لذا به نظر می‌رسد در این شرایط بایستی بخش نوشتن به آزمون‌های بسندگی اضافه شود تا اطلاع دقیقی از توانش نوشتاری دانشجویان بخصوص در مقطع تحصیلات تکمیلی وجود داشته باشد. با توجه به برگزاری و تصحیح این آزمون توسط رایانه، گنجاندن نوشتار دارای

چالش‌های خاص خود می‌باشد که بایستی بدان پرداخته شود. پس می‌توان گفت ملاحظات اجرایی آزمون آنقدر قوی و گسترده‌است که بر تمامی جنبه‌های علمی آزمون‌های بسندگی سایه افکنده‌است و نیاز به بازنگری جدی دارد.

۵. نتیجه گیری

در این مقاله نتایج تولید یک آزمون بسندگی زبان فارسی و بررسی روایی و پایایی آن گزارش گردید. به هر حال چالش‌هایی در این میان وجود دارد فقدان وجود اطلاعات دقیق و پژوهش‌های متعدد در خصوص آزمون‌های توانش زبانی در کشور باعث می‌شود نتوان به استاندارد بودن آنها مطمئن بود. در ضمن با توجه به عدم دسترسی به سوالات این آزمون‌ها، امکان بررسی روایی و پایایی این آزمون‌ها و استفاده از آنها در شرایط مشابه برای کسب اطلاعات بیشتر و انجام تحلیل‌های آماری در خصوص کیفیت تست‌های این آزمون‌ها میسر نیست. عدم همکاری در ارائه نمونه سوالات سامفا، به پژوهشگران امکان بررسی دقیق سوالات آن آزمون و محاسبه روایی و پایایی آن را میسر نداشت.

با توجه به نوپا بودن تولید آزمون بسندگی برای زبان فارسی، می‌توان به بررسی جنبه‌های مختلف شرکت‌کنندگان در این آزمون‌ها نیز پرداخت. موضوعاتی از قبیل نگرش شرکت‌کنندگان نسبت به چنین آزمون‌هایی، بررسی رابطه‌ی رشته تحصیلی با آزمون‌های بسندگی و لزوم یا عدم لزوم نسخه‌های متعدد متناسب با رشته تحصیلی و جنیست شرکت‌کنندگان، مقایسه‌ی نتایج این آزمون‌ها با یکدیگر و نیز تفاوت آزمون کاغذی با الکترونیکی نیز از مواردی است که نیاز به پژوهش بیشتر دارد.

با توجه به الکترونیکی برگزار شدن این آزمون و روند الکترونیکی شدن همه آزمون‌ها بخصوص تافل، اهمیت واکنش شرکت‌کنندگان به این نوع آزمون جدید، می‌توان بر موضوعات دیگری همچون میزان سواد رایانه‌ای، میزان نگرش به رایانه، میزان به کارگیری فن آوری اطلاعات و ارتباطات توسط شرکت‌کنندگان در آزمون الکترونیکی و میزان اضطراب رایانه‌ای شرکت‌کنندگان در آزمون الکترونیکی و عملکرد آنان در آن آزمون را مورد بررسی و پژوهش قرار داد.

آزمون بسندگی ما علیرغم برگزار شدن توسط رایانه و استفاده از کیفیت خوب صدا بخاطر استفاده از قابلیت‌های الکترونیکی همچون انطباقی بودن تست و استفاده از سایر امکانات الکترونیکی همچون چند رسانه‌ای استفاده نکرده‌است فلسفه‌ی ما در الکترونیکی کردن این تست، کاملاً اجرایی بوده‌است یعنی حذف کاغذ، عدم نیاز به چاپ، تکثیر و حمل پر سخنامه، عدم نیاز به مراقبت‌های امتحانی برای جمع آوری پاسخنامه، عدم نیاز به ارسال پاسخنامه و عدم نیاز به محاسبه‌ی نتایج توسط تصحیح کننده.

طبق تقسیم بندی مطرح در این حوزه که آزمون‌های زبانی الکترونیکی را به دو دسته‌ی سطح بالا و سطح پائین دسته بندی می‌کنند می‌توان آزمون بسندگی ما را نوعی آزمون الکترونیکی سطح پائین دانست که در آن از تمامی قابلیت‌های الکترونیکی برای انجام بهتر و موثرتر آزمون استفاده نشده‌است ولی با توجه به اینکه، این گام، می‌تواند اولین گام در حرکت به سمت آزمون‌های واقعا الکترونیک باشد لذا می‌تواند گامی مثبت ارزیابی شود و نگارش این مقاله و مقاله‌های بعدی، قطعاً به پیشرفت موثر در این زمینه کمک شایانی خواهد کرد. لازم است فراموش نکنیم آزمون سامفا بصورت کاغذی برگزار می‌گردد و شواهد حاکی از آن است که قصد الکترونیکی کردن آن ولو در سطح پائین را هم ندارند.

یکی از مشکلات و محدودیت‌های این آزمون، نداشتن قسمت‌های نوشتن و صحبت کردن است. آزمون خوب بایستی قابلیت اجرا داشته باشد. بر طبق (Brown, 2004) تست عملی باید ۱ خیلی گران نباشد ۲ دارای بازه زمانی مناسب باشد ۳ اجرای آن آسان باشد و ۴ روش امتیازدهی به آن مشخص و سریع باشد (صفحه ۱۹). به همین دلایل است که امکان افزودن بخش‌های نوشتن و صحبت کردن به این آزمون‌ها فعلاً میسر نمی‌باشد. این آزمون اولین آزمونی بود که بصورت کاملاً الکترونیکی برگزار گردید. با توجه به جدید بودن سیستم برگزاری آزمون الکترونیکی، مسلماً فرمت الکترونیکی نیز دارای نکات و ملاحظات خاص خود است که پژوهشگران امیدوارند در پژوهش‌های بعدی بتوانند به آن بپردازند.

منابع

تقیان، حسن، خدایی، ابراهیم، بازرگان، عباس، مقدم زاده، علی، کبیری، مسعود. ساخت آزمون خواندن برای دانش‌آموزان پایه‌ی ششم ابتدایی با استفاده از انگاره‌ی سنجش شناختی تشخیصی. پژوهش نامه آموزش زبان فارسی به غیر فارسی زبانان، ۱۳۹۷، ۷(شماره یک-پیاپی ۱۵)، ۳-۳۰.

- جلیلی، سید اکبر، خدادادیان، مهدی. ضرورت رویکرد نظام‌مند به آزمون‌سازی در زبان فارسی: مقایسه‌ی برآوردهای مدرسان و واقعیت آماری. پژوهش نامه آموزش زبان فارسی به غیر فارسی زبانان، ۱۳۹۲، ۲ (شماره سه - پیاپی ۵)، ۳-۳۶.
- جلیلی، سید اکبر، آزمون مهارتی فارسی (آمفا) بر پایه چهار مهارت اصلی زبانی، پایان نامه کارشناسی ارشد، ۱۳۹۰.
- جلیلی، سید اکبر، نمونه آزمون‌های زبان فارسی برای آزمون غیر ایرانی. انتشارات دانشگاه بین‌المللی امام خمینی ۱۳۹۷.
- حسینی، هانیه، طراحی، اعتبارسازی و روایی‌سنجی آزمون تک به منظور اندازه‌گیری بسندگی زبان فارسی. پایان نامه کارشناسی ارشد دانشگاه فردوسی مشهد، ۱۳۹۰.
- زند، بهمن، روشن، بلقیس، وکیلی‌فرد، امیررضا، گلپور، لیلا. آزمون چندگزینه‌ای شنیدن: مقایسه‌ی تأثیر ارائه‌ی گزینه‌ها در دو قالب نوشتاری و گفتاری. پژوهش نامه آموزش زبان فارسی به غیر فارسی زبانان، ۱۳۹۳، ۳ (شماره دو - پیاپی ۷)، ۶۸-۴۷.
- صحرایی، رضا مراد، جلیلی، سید اکبر. مبانی تهیه و تدوین آزمون مهارتی فارسی (آمفا). پژوهش نامه آموزش زبان فارسی به غیر فارسی زبانان، ۱۳۹۱، ۱ (شماره یک - پیاپی ۱)، ۱۵۰-۱۲۳.
- عباسی، زهرا، تهیه و تدوین آزمون بسندگی زبان فارسی. مقاله رایج شده به اولین همایش آموزش زبان و ادبیات فارسی. انجمن آموزش زبان و ادبیات فارسی. ۱۳۹۳، تهران
- قسولی هزاره، بهزاد، طراحی و روایی‌سنجی آزمون بسندگی زبان فارسی پژوهش زبان‌های خارجی، ۱۳۷۹، شماره ۵۷؛ از صفحه ۱۱۵ تا صفحه ۱۲۹.
- گلپور، لیلا. طراحی آزمون مهارت نوشتاری ویژه‌ی غیرفارسی زبانان: راهکارها و تحلیل خطاها. پژوهش نامه آموزش زبان فارسی به غیر فارسی زبانان، ۱۳۹۷، ۷ (شماره دو - پیاپی ۱۶)، ۶۸-۴۵.
- موسوی، ثریا، بررسی راهکارهای نظری آزمون‌های بسندگی و تهیه یک آزمون بسندگی استانده زبان فارسی برای غیرفارسی‌زبانان، ۱۳۷۹
- پایان‌نامه. دولتی - وزارت علوم، تحقیقات، و فناوری - دانشگاه شیراز. ۱۳۷۹. کارشناسی ارشد
- Abbasi, Z. (2014) Preparation and Testing of Persian Language Proficiency Test. Paper presented at the first conference on teaching Persian language and literature. *Persian Language and Literature Education Association*. Tehran. (In Persian)
- Alderson, J. C. (2000). Assessing reading. Cambridge, England: *Cambridge University Press*.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). Language test construction and evaluation. *Cambridge, England: Cambridge University Press*.
- Bachman, L. F. (2004). Statistical analysis for language assessment. *Cambridge, England: Cambridge University*.
- Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice: Designing and Developing Useful Language test. *Oxford, England: Oxford University Press*.
- Bae, J., & Bachman, L. F. (1998). A latent variable approach to listening and reading: testing factorial invariance across two groups of children in the Korean/English two-way immersion program. *Language Testing*, 15, ۳۸۰-۴۱۴
- Bailey, K. (1998). Learning about Language Assessment. *Boston, MA: Heinle & Heinle*.
- Bijankhan, M., Shayestefar, P. (2015). Corpus-Based Insights into Modeling a Level-Specific Persian Language Proficiency Test (PLPT): Development and Factor Structure of the PLPT Listening Tasks. *Teaching Persian to non-persian speakers quarterly* (11) ۱۹-۴۲.
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171-191.
- Brown, H. D. (2004). Language assessment: Principles and classroom practices. *White Plains, NY: Pearson Education*.
- Buck, G. (2001). Assessing listening. *Cambridge, England: Cambridge University Press*.
- Byrne, B. M. (1994). Structural equation modeling with EQS and EQS? Windows. *Thousand Oaks, CA: Sage*.
- Chapelle, C. A., and Abraham, R. G. (1990). Cloze method: What does it make? *Language Testing*, 7, 121-146.
- Cheng, L. (2005). Changing language teaching through language testing: A washback study. *Cambridge, England: Cambridge University Press*.
- Choi, I. C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25, 39-62.
- Educational Testing Service. (2013). *Guidelines for best test development practices to ensure validity and fairness for international English language proficiency assessments*. Retrieved from https://www.ets.org/s/about/pdf/best_practices_ensure_validity_fairness_english_language_assessments.pdf
- Ghansouli Hazareh, B. (2010) Design and Validation of Persian Language Proficiency Test. *Foreign Languages Research*, 57; (115-129). (In Persian)

- Gulpur, L.. (1397). Designing a Non-Farsi Language Writing Skills Test: Solutions and Error Analysis. *Teaching Persian to Non-Persian Speakers*, 7 (2 - 16), 45-68. doi: 10.30479 / jtpsol.2018.1548. (In Persian)
- Hale, G. A., Rock, D. A., & Jirele, T. (1989). Confirmatory factor analysis of the TOEFL. *TOEFL research report 32*. Princeton, NJ: Educational Testing Service.
- Henning, G. (1987). A guide to language testing. *Los Angeles, CA: Newbury House*.
- Hosni, H. (2011). Design, and validation of a single test to measure Persian language proficiency. *Master's Thesis*. Ferdowsi University of Mashhad. (In Persian)
- Hughes, A. (2002). Testing for language teachers. *Cambridge, England: Cambridge University Press*.
- Jalili, S. (1979) Examples of Farsi Language Exams for Non-Iranian Persian Students. *Imam Khomeini International University Press*.
- Jalili, S. (2011) Persian Language Test (AMFA) Based on Four Basic Master's Thesis. *Allameh Tabatabaee University*. (In Persian)
- Jalili, S., Khodadadian, M. (1392). The Need for a Systematic Approach to Testing in Farsi: A Comparison of Teachers' Estimates and Statistical Reality. *Teaching Persian to Non-Persian Speakers*, 2 (3 - 5), 3 - 36. (In Persian)
- McNamara, T. F. (2000). Communication and design of language tests. In H. G. Widdowson (Ed.), *Language testing* (pp. 13-22). Oxford, England: Oxford University Press.
- Mousavi, S. (2000). Evaluation of theoretical approaches to proficiency tests and preparation of a standard Persian language proficiency test for non-Persian speakers. *Master's Thesis*. Shiraz University. (In Persian)
- Pan, Y. C. (2013). Does teaching to the exit? A case study of teacher's washback in Taiwan. *Journal of Asia TEFL*, 10(4), 185-213.
- Pan, Y. C. (2014). Learner washback variability in standardized exit tests. *The Electronic Journal for English as a Second Language*, 18(2), 1-30.
- Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education*, 14, 51-74.
- Roeber, C., & Pan, Y. C. (2008). Test review: GEPT: General English Proficiency Test. *Language Testing*, 25, 403-408.
- Romhild, A. (2008). Investigating the factor structure of the ECPE across different proficiency levels. *Spain Fellow Working Papers in Second of Foreign Language Assessment*, 6, 29-55.
- Sahari, R., Jalili, S.. (1391). Basics of preparation and formulation of Persian (AMFA) Proficiency test. *Teaching Persian to Non-Persian Speakers*, 1 (1), 150-112. (In Persian)
- Shin, S. K. (2005). Did they take the same test? Examinee language proficiency and the structure of language tests. *Language Testing*, 22, 31- ۵۷
- Shohamy, E. (2001). The power of tests. *Harlow, England: Pearson Education*.
- Sims, J. (2008). Tunghai freshmen English for non-major's program. *Tunghai Journal of Humanities*, 49, 353-373.
- Sims, J. (2012). Perceived changes in the English language ability of freshmen at a university in Taiwan. *Journal of Pan-Pacific Association of Applied Linguistics*, 16(1), 33-54.
- Sims, J., & Liu, J. (2013). Two decades of changes in the English ability of freshmen at a university in Taiwan. *Hwa Kang English Journal*, 19, 23-۵۱.
- Song, M. Y. (2008). Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. *Language Testing*, 25, 435-464. <https://doi.org/10.1177/0265532208094272>
- Taghiyan, H., Khodai, A., Bazargan, A., Moghadamzadeh, A., and Kabiri, M. (1397). Development of Reading Test for Elementary Grade 6 Students Using Cognitive Assessment Model. *Research in Farsi to Non-Farsi Language Education Research*, 7 (15), 3-30.(In Persian)
- Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Listening Research*, 49, 1193-1208. <https://doi.org/10.1044/1092-4388>
- Tsai, Y., & Tsou, C. H. (2009). A standardized English Language Proficiency test as the graduation benchmark: Student perspectives on its application in higher education. *Assessment in Education: Principles, Policy & Practice*, 16, 319-330. DOI: 10.1080/09695940903319711
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19-36).
- Zandi, B., Roshan, B., Vakilifard, A., Golpour, L. (1393). Multiple Choice Listening Tests: A Comparison of the Effect of Presenting Options in Two Written and Speech Formats. *Teaching Farsi to Non-Farsi Speakers Quarterly*, 3 (2 - 7), 47-68. (In Persian)