



دوماهنامه علمی- پژوهشی

د ۱۰، ش ۶ (پیاپی ۵۴)، بهمن و اسفند ۱۳۹۸، صص ۱۳۹-۱۷۱

طراحی و اعتبار بخشی آزمون استاندارد زبان فارسی (بر پایه چهار مهارت زبانی)

بهمن زندی^{۱*}، بلقیس روشن^۲، امیررضا وکیلی فرد^۳، لیلا گلپور^۴

۱. استاد گروه زبان‌شناسی، دانشگاه پیام‌نور، تهران، ایران.
۲. دانشیار گروه زبان‌شناسی، دانشگاه پیام‌نور، تهران، ایران.
۳. دانشیار گروه آموزش زبان فارسی، دانشگاه بین‌المللی امام خمینی^(د) قزوین، ایران.
۴. دانشجوی دکتری دانشگاه پیام‌نور، تهران، ایران.

پذیرش: ۹۷/۷/۲۴

دریافت: ۹۷/۰۲/۲۲

چکیده

هدف اصلی این پژوهش، طراحی آزمون مهارت (بسنده) زبان فارسی بر پایه چهار مهارت زبانی درک شنیداری، درک خوانداری، صحبت کردن و نگارش است. فرضیه این پژوهش بر این اصل استوار است که این آزمون به لحاظ تجزیه و تحلیل آزمون‌سازی دارای پایایی و روایی لازم، دارای سودمندی لازم، قابل اجرا و کاربردی است و در نهایت منطبق با معیارهای استاندارد شده، طراحی شده است. جامعه آماری این پژوهش، فارسی‌آموزان مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی^(د) بودند و ۱۳۰ نفر به عنوان نمونه آماری انتخاب شدند که این زبان‌آموزان حداقل ۴ ماه و حداکثر ۱۲ ماه در این مرکز مشغول یادگیری زبان فارسی بوده‌اند. به منظور استانداردسازی آزمون، پرسش‌های آزمون به صورت پیش‌آزمون در آزمون‌های پایان ترم بهار ۹۴ و تابستان ۹۴ مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی^(د) استفاده و بعد از آسیب‌شناسی پرسش‌ها و تجزیه و تحلیل، برای این آزمون انتخاب شده است. پرسش‌های آزمون پس از اجرا تحلیل شدند و پایایی و روایی آزمون، دو مشخصه مهم آزمون استاندارد، اندازه‌گیری شدند. پس از بررسی آزمون طراحی شده، نتایج آماری نشان‌دهنده این موضوع بوده‌اند که این آزمون، ضریب دشواری و ضریب تمیز بهینه‌ای دارد و دارای پایایی و روایی قابل قبولی نیز است و پیش‌بینی می‌شود که این آزمون سودمند و دارای کاربردی نیز است.

واژه‌های کلیدی: آزمون‌شناسی، مهارت‌های زبانی، آزمون استاندارد زبان فارسی.

E-mail: zandi@pnu.ac.ir

* نویسنده مسئول مقاله:

۱. مقدمه

زبان فارسی یکی از قدیم‌ترین زبان‌های دنیاست که با پشتوانه تاریخی، علمی و ادبیات غنی و با ارزش، زبان سرزمین ایران و حتی اطراف آن بوده است. در حال حاضر نیز بسیاری از فارسی-آموزان خارجی به انگیزه آشنایی با ادبیات فارسی، نیاز شغلی، مهاجرت و زندگی در ایران، تحصیل و یا تدریس در رشته‌های مختلف دانشگاهی و حوزوی در ایران در حال یادگیری زبان فارسی هستند. بنابراین، یکی از راه‌های مؤثر، معرفی و گسترش زبان فارسی و فرهنگ ایرانی، همین دانشجویان خارجی در ایران هستند که می‌توانند پس از پایان تحصیلات دانشگاهی به کشورشان باز گردند و مبلغان خوبی برای فرهنگ ایرانی و زبان فارسی باشند. توسعه دوره‌های آموزش زبان فارسی به غیرفارسی‌زبانان و تدوین کتاب‌های مختلف آموزشی مرتبط در این زمینه، تحولی فرهنگی است که با وجود آثار مثبت آن، کاستی‌هایی را نیز به همراه دارد. مهم‌ترین این نقص‌ها، گستره آزمون‌سازی زبان و به‌ویژه کمبود آزمون مهارتی (بسندگی) زبان فارسی است. همان‌طور که می‌دانیم ارزشیابی و آزمون زبان جزو جدایی‌ناپذیر و مکمل‌کننده آموزش زبان است. ارتباط آموزش و آزمون به قدری به هم وابسته است که پرداختن به هر کدام از آن‌ها بدون سروکار داشتن دائم با دیگری غیرممکن است. بدون تردید، یکی از ابعاد مهم و تأثیرگذار آزمون‌ها، قلمرو سیاسی، فرهنگی و شکل‌دهی هویت ملی هر کشوری است. برای مثال، آزمون بسندگی زبان انگلیسی (تافل) نماد قدرت زبانی کشورهای انگلیسی زبان است که سالانه میلیون‌ها نفر در سراسر جهان تنها پس از دریافت نمره لازم در این آزمون، مجوز پذیرش ورود به دانشگاه را به‌دست می‌آورند. آموزش زبان فارسی و آزمون آن نیز از این قاعده مستثنی نیست. دانشجویان خارجی که برای تحصیل به ایران می‌آیند، سطح زبانی یکسانی ندارند. پس ما برای سنجش سطح زبانی دانشجویان خارجی نیازمند ملاک و معیار زبانی معتبری هستیم تا بتوانیم بر اساس آن اقدام به راهنمایی تحصیلی کنیم. همچنین، باید به گونه‌ای عمل کرد که هزینه صرف‌شده برای داوطلبان یادگیری زبان فارسی، به نحو مطلوبی به نتیجه برسد و بتوانند ضمن یادگیری، به رشد و توسعه زبان فارسی در کشور خودشان و علاقه‌مند کردن مردم آنجا به زبان فارسی کمک کنند. در بیشتر موارد ما بدون سنجش دانشجویان خارجی بر اساس معیار و ملاک علمی، آنان را به دانشگاه‌هایمان وارد می‌کنیم و این ناآگاهی ما نیز از اینجا نشئت می‌گیرد که معیار و ملاک معتبری در ایران وجود ندارد که بتوان میزان مهارت زبان فارسی‌آموزان را بر اساس آن

سنجید. با توجه به موارد یادشده، این پژوهش نیز بر آن است که با توجه به اصول و مبانی تهیه آزمون‌های زبان به تهیه و تدوین یک آزمون مهارت بر پایه چهار مهارت اصلی برای سنجش مهارت زبانی فارسی آموزان خارجی بپردازد تا این آزمون، ملاکی برای پذیرش دانشجویان غیر-ایرانی (از لحاظ سطح زبانی) باشد. پرسش‌هایی که در این پژوهش مطرح می‌شوند در اینجا آمده است:

*آزمون طراحی‌شده، به لحاظ تجزیه و تحلیل آزمون‌سازی دارای چه ویژگی‌هایی است؟

*آزمون طراحی‌شده تا چه اندازه منطبق با معیارهای استاندارد طراحی شده است؟

برای دستیابی به اهداف این پژوهش، دوفرضیه مطرح شده است:

فرضیه اول: این آزمون به لحاظ تجزیه و تحلیل آزمون‌سازی دارای ویژگی‌هایی است که می‌آوریم:

- دارای سودمندی لازم است.

- دارای پایایی لازم است.

- دارای روایی لازم است.

- قابل اجرا و کاربردی است.

فرضیه دوم: این آزمون به لحاظ تجزیه و تحلیل آزمون‌سازی منطبق با معیارهای استاندارد-شده، طراحی شده است.

۲. پیشینه پژوهش

در این بخش مطالعاتی که در گذشته در مورد آزمون‌های زبان فارسی برای غیرفارسی‌زبانان، انجام شده، بررسی شده است:

موسوی (۱۳۷۹) در پژوهشی با عنوان «بررسی راهکارهای نظری آزمون‌های بسندگی و تهیه یک آزمون بسندگی استاندارد زبان فارسی برای غیرفارسی‌زبانان» به مطالعه مبانی تهیه یک آزمون چهاربخشی پرداخته است که قصد دارد مهارت زبانی فارسی‌آموزان خارجی را بسنجد. چهار بخش آزمون مورد نظر، گوش دادن، قواعد ساختاری، واژگان و درک خواندن است. در این آزمون دو مهارت تولیدی صحبت کردن و نگارش لحاظ نشده است و برای بخش دستور همانند تافل بخش جداگانه‌ای در نظر گرفته شده است، در حالی‌که پژوهش حاضر دارای چهار مهارت

زبانی است و مانند پژوهش موسوی از نوع واحد - منفصل نیست و دستور و واژگان در پرتو مهارت‌های زبانی سنجیده شده است.

قنسولی (۱۳۸۹) در مقاله‌ای با عنوان «طراحی و روایی سنجی آزمون بسندگی زبان فارسی» از میان مدل‌های مختلف اجزای بسندگی زبان، مدل چند بعدی زبان را که با واقعیت شرایط آزمون تناسب دارد، انتخاب کرده است. در این مدل به دو مهارت اصلی زبان مانند خواندن، شنیدن و مقوله‌های دستور و واژگان توجه شده است. در طراحی این آزمون تنها به دو مهارت اصلی خواندن و شنیدن توجه شده و مهارت‌های گفتن و نوشتن به دلیل محدودیت‌های تحقیق مانند پژوهش موسوی، لحاظ نشده است. در این پژوهش ضمن ساخت یک نمونه آزمون بسندگی زبان فارسی با تأکید بر آزمون‌های شنیدن و خواندن و دستور، روایی و پایایی آزمون نیز سنجیده شده است. برای روایی سنجی آزمون، از فرایند دو مرحله‌ای شامل تحلیل بازنگری سه دانشجوی خارجی دوره دکتری زبان و ادبیات فارسی و بهره‌گیری از مدل فولچر و دیویدسون بر مبنای نظریه بحث در سنجش روایی بهره گرفته شده است. این مدل چارچوبی مفهومی است و اساساً روایی را فرایند بحث درباره معنای نمرات و کاربرد نمرات می‌داند. پایایی آزمون نیز به شیوه آلفای کرونباخ محاسبه شده است. نتایج این آزمون نشان داد که دانشجویان خارجی در بخش خواندن و درک مفاهیم بهترین عملکرد را داشتند و بخش دستور دشوارترین بخش برای آنان بود و پایایی آزمون نیز ۰/۷۰ تعیین شد. در این پژوهش نیز خلاف پژوهش حاضر جای مهارت‌های تولیدی نوشتار و گفتار خالی است.

جلیلی (۱۳۹۰) در پژوهشی با عنوان «آزمون مهارتی فارسی بر پایه چهار مهارت اصلی زبانی» تلاش کرده است، یک آزمون مهارتی برای زبان فارسی طراحی کند. تفاوت موسوی با جلیلی این است که موسوی قواعد ساختاری (دستور زبان) و واژگان (دایره لغات) را به عنوان دو بخش جداگانه در آزمون خود گنجانده است و دارای قالب آزمون تافل است؛ اما جلیلی آزمون مهارتی زبان فارسی را بر پایه آلیتس طراحی کرده و آن را آمفا نامیده است. آزمون آمفا این دو بخش را به طور ضمنی در بخش‌های دیگر می‌آزماید و نیازی به جداسازی آن‌ها نمی‌بیند. به علاوه دو بخش نوشتن و صحبت کردن که از مهارت‌های زبانی ضروری برای دانشجویان خارجی هستند در این آزمون وجود دارد؛ اما در پایان‌نامه موسوی نادیده گرفته شده است. این آزمون همانند آزمون موسوی اجرا و اعتبارسنجی نشده است و تنها در حد طراحی یک آزمون بر اساس

مبانی نظری آزمون‌سازی است، در حالی‌که پرسش‌های پژوهش حاضر واکاوی و اعتبارسنجی شده است.

شقایق و همکاران (۱۳۹۰) در مقاله‌ای با عنوان «هنجاریابی آزمون بین‌المللی زبان فارسی» به هنجاریابی آزمون استاندارد کانون زبان پرداخته است. به دلیل محدود بودن آزمودنی‌ها، در این پژوهش حجم جامعه و حجم نمونه برابر است و شامل ۵۰ نفر از زبان‌آموزان سطح متوسط بوده که حداقل ۳ ماه و حداکثر ۶ ماه به یادگیری زبان مشغول بوده‌اند. این آزمون به ترتیب شامل بخش‌های شنیدن (۳۰ پرسش)، دستور (۳۰ پرسش)، واژگان و خواندن (هر یک ۲۰ پرسش) است و مهارت مکالمه و نگارش در این آزمون نیز لحاظ نشده و برای دستور نیز بخش جداگانه‌ای در نظر گرفته شده است. نتایج این آزمون نشان داد پرسش‌هایی که همبستگی کمی دارند بیشتر یا به سنجش کلمات و عبارات کم بسامد پرداخته‌اند و یا عوامل دیگری مانند ضریب هوشی، میزان درک شنیداری، قوه استدلال و مانند آن‌ها در پاسخ‌گویی مؤثر بوده‌اند. در این آزمون ضریب دشواری، ضریب تمیز و روایی آزمون ارزیابی نشده است.

ممقانی (۱۳۹۰) در پژوهشی با عنوان «ارزیابی روایی و پایایی آزمون وزارت علوم و تحقیقات با هدف دستیابی به یک آزمون مهارت برای زبان فارسی» به ارزیابی پایایی و روایی آزمون زبان انگلیسی وزارت علوم و تحقیقات و فناوری MSRT پرداخته است. هدف از این پژوهش این بوده است که آیا می‌توان با ارزیابی نتایج روایی و پایایی آزمون (MSRT) به یک الگوی مناسب برای طراحی یک آزمون مهارت زبان فارسی رسید. در همین راستا، از دو آزمون دیگر، یعنی آمفا (آزمون مهارتی فارسی) و آزمون کانون زبان ایران نیز استفاده شده و روایی و پایایی این دو آزمون نیز در این پژوهش بررسی شده‌اند. از داده‌های این پژوهش این نتایج به دست آمد که از آزمون MSRT نمی‌توان برای طراحی یک آزمون مهارت فارسی استفاده کرد؛ چون فاقد مهارت صحبت کردن است و مانند آزمون‌های گذشته که واحد - منفصل بودند، دو بخش دستور و واژگان را جداگانه سنجش کرده است. بنابراین، نتیجه‌گیری شد که از میان دو آزمون آمفا و کانون زبان ایران، آزمون آمفا مینای مناسب‌تری برای ارزیابی مهارت‌های زبانی غیرفارسی - زبانان است؛ زیرا شامل چهار مهارت اصلی زبان است.

۳. مبانی نظری پژوهش

اساساً آزمون‌ها به منظور دستیابی به ۲ هدف کلی به‌کار گرفته می‌شود. یکی ارزشیابی پیشرفت و دریافت یافته‌ها و دیگری پیش‌بینی توان آزمودنی در آینده است. هدف از آزمون دریافت یافته‌ها، سنجش مهارت یا توانی است که آزمودنی تا زمان انجام آزمون به‌دست آورده است، در حالی که هدف از آزمون پیش‌بینی، تعیین میزان مهارت یا توانی است که آزمودنی در زمان معینی در آینده می‌تواند دارا باشد.

جدول ۱: کارکردهای آزمون زبان
Table 1. Language Testing Functions

کارکردهای آزمون‌های زبان					
ارزشیابی پیشرفت (تعیین میزان دریافت یافته‌ها)			پیش‌بینی		
پیشرفت: تشخیصی عمومی	مهارت (بسندگی)	دانش	استعداد	جایابی گروه‌بندی تعیین سطح	گزینش

علاوه بر تقسیم‌بندی جدول شماره ۱، تقسیم‌بندی‌های دیگری هم برای آزمون وجود دارد از قبیل آزمون‌های عینی و آزمون‌های ذهنی، آزمون‌های کلان و آزمون‌های خرد، آزمون‌های تکوینی و آزمون‌های مجموعی، آزمون نحوه کاربرد و آزمون کاربرد، آزمون سرعت و آزمون توانایی و آزمون معیار - مرجع و آزمون هنجار - مرجع (Coombe et al., 2007, xvii). اما از آنجا که آزمون طراحی‌شده این پژوهش آزمون مهارت (بسندگی) است تنها به بررسی این نوع از آزمون پرداخته می‌شود.

۳-۱. آزمون مهارت (بسندگی)

آزمون مهارت نوعی از آزمون‌های زبان است که کاربرد و هدف آن ارزشیابی پیشرفت است. فرهادی و همکاران (2007:148) آزمون مهارت را این‌گونه تعریف کرده‌اند: آزمون مهارت به منظور تعیین میزان توانایی آزمودنی در زبان به‌کار می‌رود. چنین آزمونی با نگرش به گذشته و

با هدف سنجش آنچه که تا زمان برگزاری آزمون فراگرفته شده، انجام می‌گیرد و توجهی به شیوه زبان‌آموزی آزمودنی ندارد. بسندگی زبان، به آموزش در یک درس یا دوره محدود نمی‌شود؛ بلکه توصیفی است از آنچه آزمودنی می‌تواند در زمان اجرای آزمون انجام دهد. آزمون‌های مهارت این را می‌سنجند که یک شخص چه مقدار از یک زبان را یاد گرفته است. این آزمون‌ها به دنبال پاسخ به این پرسش هستند: با این مقدار یادگیری، زبان‌آموز چه کاری می‌تواند انجام دهد. این آزمون‌ها برای اندازه‌گیری توانایی افراد در یک زبان و بدون توجه به هر نوع آموزشی، طراحی می‌شوند. بنابراین، محتوای یک آزمون مهارت، مبتنی بر محتوا یا اهداف دوره‌های زبانی نیست (ربکا، ۱۳۷۳: ۱۰).

همچنین، یک آزمون مهارت می‌تواند به‌عنوان یک آزمون تعیین سطح، برای تقسیم آزمون‌دهندگان به سطوح مهارتی مشابه به‌کار رود و یا به‌عنوان یک آزمون گزینشی، برای اطمینان از مهارت زبانی یک شخص به منظور آغاز یک دوره یا تصدی یک حرفه یا شغل مورد استفاده قرار گیرد (Brindly, 2003: 317).

از آنجا که برای شروع طراحی آزمون، لازم است تا با بررسی مدل‌های آزمون بسندگی زبان، یک مدل زبانی انتخاب شود. بنابراین، در ابتدا انواع مدل‌های آزمون بسندگی زبان بررسی می‌شود:

۳-۲. مفهوم کلی بسندگی زبان

در ادبیات آزمون‌سازی، تعاریف متعددی برای بسندگی کلی زبان مطرح شده است که هر یک برگرفته از چارچوب نظری حاکم بر دوره‌های خاص مکتب زبان‌شناسی کاربردی است. برای نمونه در دوره حاکمیت ساختارگرایی در زبان‌شناسی که متأثر از مکتب روان‌شناسی رفتارگرایی است، زبان و دانش کلی زبان تعریفی برگرفته از نوعی سازوکار شرطی شدن محض دارد. در این دیدگاه، توانایی کلی زبان را می‌توان در سابقه و شرایط یادگیری پیشین آن جست‌وجو کرد. از این رو، بسندگی کلی زبان با توجه به سطح و میزان آشنایی قبلی زبان‌آموز تعیین می‌شود. در این دیدگاه، دانش زبان در حد مجموعه‌ای از مهارت‌ها محدود شده است و ماهیت پیچیده زبان تنزل پیدا می‌کند. علاوه بر روش آزمون‌های واحد - منفصل که در دوره ساختارگرایی حاکم بود، روش دیگری که مقبولیت بیشتری در میان اهل نظر داشت، مطرح شد. این روش بر طراحی مدلی

از دانش زبانی تأکید می‌کند که بر اساس دانش گویش بومی زبان استوار است. پرسش اصلی در این مدل آن است که دانستن یک زبان و ویژگی‌های آن چیست. چهار ویژگی در پاسخ به این پرسش مطرح است که عبارت‌اند از:

۱. چیرگی بر صورت‌های زبان؛

۲. چیرگی بر معانی شناختی، عاطفی، فرهنگی - اجتماعی؛

۳. ظرفیت لازم در به‌کارگیری زبان برای ایجاد ارتباط؛

۴. خلاقیت در استفاده از زبان. (Stern, 1983)

از دیدگاه کارتر و نونان (2001: 137) هم‌بستگی زبان توانایی به‌کارگیری زبان دوم با اهداف ارتباطی و در بافت واقعی است که با نظریه استرن همسو است. هیوز نیز بیان می‌کند که بسندگی زبان، داشتن تسلط زبانی کافی به‌منظور رسیدن به یک هدف خاص است که با وجود تفاوت‌هایی که میان محتوا و سطح دشواری آزمون‌های آن وجود دارد، به‌طور مشترک بر اساس دوره‌های آموزشی و محتوای درسی طراحی نمی‌شود. چارچوب نظری این پژوهش، بر اساس مفهوم چند بعدی نظریه استرن مبنی بر چهار مهارت اصلی شنیدن، گفتن، خواندن و نوشتن و چیرگی بر صورت‌های زبان و ظرفیت به‌کارگیری زبان است (Hughes, 2004).

۴. روش‌شناسی پژوهش

در این بخش متغیرهای پژوهش، جامعه آماری، ابزارهای پژوهش، روش انجام پژوهش، ساختار آزمون و معیارهای نمره‌دهی، بحث و بررسی می‌شود.

۴-۱. متغیرهای پژوهش

متغیرهایی که در این پژوهش بررسی می‌شود عبارت‌اند از: پایایی و روایی آزمون. در تعریف روایی گفته شده است، آزمونی روایی دارد که فقط آنچه را می‌خواهیم بسنجد و نه چیز دیگری را. در تعریف پایایی نیز گفته می‌شود که پایایی یک آزمون به معنی درستی و همسانی نتایج است. به عبارتی آزمونی که در تکرار آن نتایج همسانی به‌دست آید دارای پایایی لازم است (سیف، ۵۵۷:۱۳۹۰).

۴-۲. جامعه آماری پژوهش

جامعه آماری این پژوهش، زبان‌آموزان مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی^(۵) است. تعداد کل این جامعه آماری ۲۰۰ نفر است، نمونه آماری در این پژوهش ۱۳۰ نفر (زن - مرد) از این جامعه آماری بوده که حداقل ۴ ماه و حداکثر ۱۲ ماه در این مرکز مشغول به یادگیری زبان فارسی بوده‌اند. میانگین سنی آن‌ها بین ۱۸ - ۳۸ سال است. روش نمونه‌گیری به صورت تصادفی است.

۴-۳. ابزارهای پژوهش

آزمون طراحی‌شده، یکی از ابزارهای این پژوهش است که از ۴ بخش درک شنیداری، درک خوانداری، صحبت کردن و نگارش تشکیل شده است. پرسش‌نامه، ابزار دیگر این پژوهش است. به منظور تحلیل کیفی روایی آزمون و تعیین ارتباط پرسش‌های آن با سازه‌های دانش زبان، ناگزیر از به‌کارگیری روش بازنگری^۲ شدیم. در پایان آزمون، پرسش‌نامه‌ای به آزمودنی‌ها ارائه شد تا دیدگاه آزمودنی نیز درباره آزمون بررسی شود. نرم افزار (اس پی اس اس)^۳ برای محاسبه ضریب آلفای کرونباخ برای تعیین پایایی درونی آزمون و برای محاسبه روایی سازه‌ای، از روش همسانی درونی یا تعیین روایی درونی استفاده شده است که این روش مستلزم محاسبه ضریب همبستگی پیرسون، بین اجزای آزمون است. شاخص دشواری تست‌ها نیز، بر اساس فرمول کسر پاسخ‌های صحیح بر تعداد آزمون‌دهندگان محاسبه شده است. شاخص تمییز تست‌ها نیز از طریق فرمول تفریق پاسخ‌های درست گروه ضعیف از پاسخ‌های گروه قوی بر تعداد آزمون‌دهندگان به دست آمده است.

۴-۴. روش انجام پژوهش

در این پژوهش، ابتدا به منظور ارائه مبانی و مفاهیم نظری برای مطالعه پژوهش‌های قبلی و انتخاب چارچوب نظری از روش توصیفی - تحلیلی استفاده شده است. چارچوب نظری این

پژوهش همان‌طور که در بخش قبلی هم بیان شد، بر اساس مفهوم چند بُعدی نظریه استرن است. بدین معنا که این آزمون به منظور ارزیابی توانایی کاربرد عملکردی زبان، توانایی آوایی، خطی، واژگانی و دستوری در ارتباط با مهارت‌های شنیدن، خواندن، گفتن و نوشتن طراحی شده است. برای ساخت آزمون معیار و استانداردسازی آن از روش پیمایشی توافق محور^۴ و به منظور اعتبار بخشی آزمون پیشنهاد شده، از روش میدانی استفاده شده است.

۴-۵. ساختار آزمون طراحی شده زبان فارسی

در جدول زیر ساختار آزمون، به تفکیک مهارت‌ها مشخص شده است:

جدول ۲: ساختار آزمون

Table 2. Test Structure

مهارت‌ها	تعداد پرسش‌ها	مدت زمان آزمون	بخش‌های آزمون
خواندن	۳۰ پرسش و ۳۰ امتیاز	۹۰ دقیقه	درک واژه، بندش، درک متن، دگرگفت و مرتب کردن
شنیدن	۳۰ پرسش و ۳۰ امتیاز	۴۵ دقیقه	درک جزئیات (زمان و مکان) دگرگفت، درک توالی یک رویداد، درک مقایسه کردن، درک طبقه بندی و.....
مکالمه	مصاحبه و ۲۰ امتیاز	۱۵ دقیقه	بخش عمومی بخش آکادمیک
نگارش	نوشتن و ۲۰ امتیاز	۶۰ دقیقه	بخش اول: نامه نگاری بخش دوم: انتخاب یک موضوع از بین دو موضوع و نوشتن مقاله‌ای ۲۰۰ - ۲۵۰
مجموع	۱۰۰ امتیاز	۲۱۰ دقیقه: ۳/۵ ساعت	

۴-۶. محتوای آزمون

با توجه به اینکه چارچوبی مانند چارچوب مرجع اروپا و پیکره سطح‌بندی‌شده مناسب در زمینه واژگان و ساختارهای دستوری برای آموزش زبان فارسی به غیرفارسی‌زبانان وجود ندارد و هنوز فهرست نظام‌مندی از ویژگی‌های کامل کارکردهای زبانی در اختیار نداریم. از این رو، برای طراحی این آزمون از زبان رسانه (روزنامه، مجله، رادیو و تلویزیون) استفاده شده است؛ زیرا این زبان همگانی است و علاوه بر سادگی دارای ضریب دشواری مناسبی است و از ترکیبات و اصطلاحات و ساختارهایی تشکیل شده است که در دیگر گونه‌های زبانی کمتر دیده می‌شود و دارای موضوعات گوناگونی است که درک و فهم آن نیاز به تخصص خاصی ندارد. بنابراین، محتوای آزمون درک خواندن و درک شنیدن از موضوعات مختلف رسانه‌ها انتخاب شد. در آزمون‌هایی که برای زبان فارسی طراحی شده است مانند آزمون بین‌المللی زبان فارسی کانون زبان ایران، مهارت نگارش وجود ندارد و در آزمون آمفا (آزمون مهارتی فارسی) نیز معیارهای لازم به‌منظور ارزیابی نوشته آزمون‌دهنده مشخص نشده است و از آنجا که ساختارها و الگوهای دستوری زبان فارسی، کاملاً با زبان انگلیسی منطبق نیست، نمی‌توان از معیارهای آزمون آیلتس که مهارت نوشتن دارد نیز استفاده کرد، بنابراین، برای رسیدن به این هدف، پرسش‌نامه‌ای تهیه شد و در اختیار اساتید باتجربه آزمون قرار گرفت. از مجموع اطلاعات پرسش‌نامه‌ها و با توجه به شاخص‌هایی که در کتاب *ارزیابی نگارش* (Weigle, 2002) وجود دارد، به‌منظور ارزیابی نگارش معیارهایی برای سنجش دستور (تطابق فعل با فاعل، کاربرد درست زمان افعال، استفاده درست از ترکیب‌های اضافی و وصفی، استفاده درست از حروف اضافه، ربط و نشانه مفعولی «را» و ...)، غنای واژگانی و کاربرد درست واژه‌ها با توجه به بافت، انسجام معنایی و استفاده مناسب از عوامل انسجام بین جمله‌ها، ظاهر متن (جمله موضوعی، مقدمه، بدنه، نتیجه‌گیری، رعایت اصول پاراگراف‌نویسی و علائم سجاوندی) مشخص شد، ارزیابی متون نگارش از نوع تحلیلی^۶ هستند نه کلی^۷. مقوله‌های ارزیابی مهارت صحبت کردن شامل تلفظ دستور، واژگان، روانی کلام و درک است (براون، ۲۰۱۱: ۴۰۶).

۵. اعتبارسنجی و اعتبار بخشی آزمون

در این بخش به این موضوع می‌پردازیم که آزمون طراحی شده به لحاظ تجزیه و تحلیل آزمون‌سازی دارای چه ویژگی‌هایی است و آزمون طراحی شده تا چه اندازه منطبق با معیارهای استاندارد است و به طور کلی نتیجه می‌گیریم که آیا این آزمون، آزمونی موفق بوده است؟

۵-۱. تحلیل پرسش‌های آزمون

هدف از تحلیل پرسش‌های یک آزمون واریسی تک تک پرسش‌ها و تعیین میزان دقت و نارسایی آن‌هاست. در تحلیل پرسش‌های آزمون، نقاط ضعف و قوت یک آزمون و کیفیت پرسش‌های آن تعیین می‌شود. بنابراین، لازم است که آزمونگر پس از اجرای هر آزمون، پرسش‌های آن را تحلیل کند و با استفاده از نتایج حاصل از آن به تجدید نظر در آزمون و بهبود کیفیت پرسش‌ها برای استفاده‌های بعدی اقدام کند (سیف، ۱۳۹۰: ۴۲۱). در فرایند تحلیل پرسش‌های یک آزمون، میزان ضریب دشواری، ضریب تمیز، همبستگی پرسش با کل آزمون و پایایی و روایی آن سنجیده می‌شود که در ادامه به این مباحث می‌پردازیم.

۵-۱-۱. محاسبه ضریب دشواری^۱

بنا به تعریف، درصد کل آزمون شوندگانی که به یک پرسش جواب درست می‌دهند، ضریب دشواری آن پرسش است که با حرف p نشان داده می‌شود و از طریق فرمولی که آوردیم محاسبه می‌شود:

$$P = \frac{\text{تعداد افرادی که به پرسش جواب درست داده‌اند}}{100(R/T)}$$

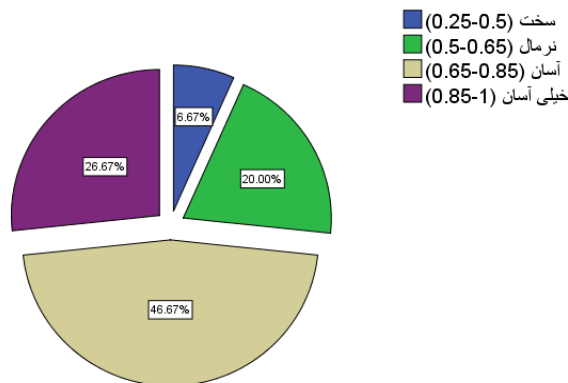
تعداد کل آزمون شوندگان

نتایج ضریب دشواری آزمون‌های شنیدن و خواندن در ادامه آورده می‌شود:

نمودار شماره ۱ توزیع فراوانی ضریب دشواری برای سطح کیفی پرسش‌های آزمون در بخش شنیدن را نشان می‌دهد. بر اساس این نمودار از مجموع ۳۰ پرسش، ۲ پرسش معادل ۶/۷

درصد آزمون در بخش شنیدن سخت، ۶ پرسش معادل ۲۰ درصد نرمال، ۱۴ پرسش معادل ۴۶/۷ درصد آسان و ۸ پرسش معادل ۲۶/۷ درصد خیلی آسان ارزیابی شده است. قابل توجه است که هیچ یک از پرسش‌ها از لحاظ ضریب دشواری خیلی سخت نبوده و فراوانی این سطح صفر است. بیشترین درصد و فراوانی مربوط به سطح آسان است. در رتبه دوم بیشترین درصد به سطح خیلی آسان اختصاص یافته است. در حالت کلی می‌توان نتیجه گرفت که آزمون در بخش شنیدن از نظر دانشجویان آسان و راحت بوده است. پس از آسیب‌شناسی مشخص شد که دلیل آسان بودن پرسش‌ها متأثر از ملیت آزمودنی‌ها و مدت یادگیری زبان فارسی بعضی از آنان بوده است؛ زیرا از میان آزمودنی‌ها ۱۳ نفر تاجیک بودند و ۲۴ نفر هم به مدت ۲ سال در دانشگاه زبان‌های خارجی پکن مشغول یادگیری زبان فارسی بوده‌اند. از این رو، توانایی مهارت درک شنیداری آنان در سطح بالایی بود و همین امر روی میزان ضریب دشواری تأثیر داشته است، در صورتی که در پیش‌آزمون نتیجه چنین نبود.

سطح کیفی سوالات - شنیدن



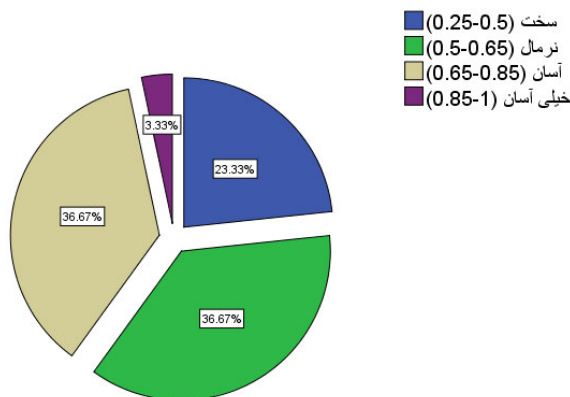
نمودار ۱: توزیع فراوانی ضریب دشواری پرسش‌های شنیدن

Graph 1. Frequency Distribution of Difficulty Coefficient of Listening Comprehension Questions

نمودار شماره ۲ توزیع فراوانی ضریب دشواری پرسش‌های آزمون در بخش خواندن را

نشان می‌دهد. بر اساس این نمودار، ۷ پرسش معادل ۲۳/۳ درصد آزمون در بخش خواندن سخت، ۱۱ پرسش معادل ۳۶/۷ درصد نرمال، ۱۱ پرسش معادل ۳۶/۷ درصد آسان و پرسش معادل ۳/۳ درصد خیلی آسان ارزیابی شده است. همانند بخش شنیدن برای هیچ‌یک از زبان‌آموزان این آزمون، بخش خواندن خیلی سخت نبوده است و فراوانی این سطح صفر است و دلیل آن نیز مانند مهارت شنیدن که در مطالب قبل عنوان شد، متأثر از ملیت تعدادی از زبان-آموزان و مدت آموزش آنان بود. سطوح نرمال و آسان بالاترین درصد را به خود اختصاص داده‌اند. در حالت کلی می‌توان نتیجه گرفت که ضریب دشواری آزمون در بخش خواندن نرمال و مناسب بوده است.

سطح کیفی سوالات - خواندن



نمودار ۲: توزیع فراوانی ضریب دشواری پرسش‌های خواندن

Graph 2. Frequency Distribution of Difficulty Coefficient of Reading Questions

۵-۱-۲. تفسیر ضریب تمیز^۹

از فعالیت‌های دیگری که برای تحلیل پرسش‌های چندگزینه‌ای انجام می‌شود، محاسبه ضریب تمیز است. ضریب تمیز که با *D* نشان داده می‌شود قدرت پرسش را در تمایزگذاری یا تشخیص بین گروه قوی و گروه ضعیف آزمون شونده‌گان مشخص می‌کند؛ یعنی معلوم می‌کند که پرسش تا چه

اندازه می‌تواند گروه قوی را از گروه ضعیف جدا سازد. برای محاسبه ضریب تمیز یک پرسش از فرمولی که آمده استفاده می‌شود (هریسون، ۱۳۷۳: ۱۶۲).

انتخاب درست گروه پایین - انتخاب درست گروه بالا = ضریب تمیز پرسش
تعداد افراد یک گروه (بالا یا پایین)

در این آزمون میانگین ضریب تمیز مهارت شنیدن ۰/۴۲ و میانگین ضریب تمیز مهارت خواندن ۰/۴۵ به دست آمد.

گفتنی است، از آنجا که آزمون مهارت صحبت کردن و نگارش از نوع ذهنی و تشریحی است امکان محاسبه ضریب دشواری و ضریب تمیز با استفاده از فرمول قبل وجود نداشت.

۵-۱-۳. سنجش پایایی آزمون

پایایی به دقت، اعتمادپذیری، ثبات یا تکرارپذیری نتایج آزمون اشاره می‌کند. از آنجا که این آزمون دارای گزینه‌های چند ارزشی است، به عبارت دیگر بیش از دو گزینه است، برای محاسبه پایایی این آزمون، از فرمول آلفای کرونباخ استفاده شده است. در روش ضریب آلفای کرونباخ هر پرسش آزمون با تک تک پرسش‌های دیگر مقایسه می‌شود. در محاسبه آلفای کرونباخ گفته شده است که اگر ضریب آلفا بیشتر از ۰/۷۰ باشد آزمون دارای پایایی قابل قبولی است (مؤمنی و قیومی، ۱۳۸۹: ۲۱۲). بنابراین، با توجه به جدول شماره ۳ و داده‌های بخش آماری، پایایی درونی آزمون‌های خواندن و شنیدن، مطلوب و قابل قبول است.

جدول ۳: پایایی مهارت خواندن و شنیدن

Table 3. Reading and Listening Skills Reliability

پایایی مهارت خواندن	پایایی مهارت شنیدن
۰/۷۸	۰/۷۰۲

برای سنجش پایایی مکالمه و نگارش این آزمون از روش پایایی بین مصحان^{۱۱} استفاده شده است. برای تعیین پایایی آزمون‌های تشریحی یا به‌طور کلی آزمون‌های غیرعینی که نمرات آن‌ها تحت تأثیر قضاوت مصحان قرار می‌گیرد، باید از دو یا چند مصحح برای ارزیابی استفاده کرد.

همبستگی بین نمرات مصححان مختلف، شاخص پایایی مصححان به‌شمار می‌آید که هر چه توافق بین مصححان یا نمره‌گذاران بیشتر باشد، این نوع پایایی بیشتر است. البته، این پایایی مربوط به شیوه کار تصحیح و نمره‌دهی است و نباید آن را با پایایی آزمون که مربوط به پرسش‌های خود آزمون است اشتباه گرفت.

۵-۱-۴. سنجش روایی^{۱۲} آزمون

روایی، ویژگی دیگری از یک آزمون موفق است. روایی اصطلاحی است که به هدفی که آزمون برای تحقق بخشیدن به آن درست شده است، اشاره می‌کند. به عبارت دیگر، آزمونی دارای روایی است که برای اندازه‌گیری آنچه مورد نظر است کافی و مناسب باشد. روایی، مفهومی آزمون وابسته‌تر از پایایی است و بیشتر از پایایی به آزمون متکی است. روایی بیشتر به خصوصیات آزمون بستگی دارد. بنابراین، تنها به‌دست آوردن یک تعداد نمره از یک آزمون، کاربران را قادر به ثبت روایی آن نمی‌کند. البته، روایی یک امر نسبی است. رابطه بین پایایی و روایی بدین‌گونه است که یک آزمون باید پایا باشد تا بتواند روا باشد؛ یعنی پایایی شرط روایی است؛ اما روایی برای پایایی ضروری نیست.

۵-۱-۴-۱. سنجش روایی محتوایی^{۱۳}

برای محاسبه روایی محتوایی و ظاهری (صوری)^{۱۴} یک آزمون، باید گفت که درواقع، هیچ‌گونه روش آماری برای تعیین آن وجود ندارد؛ بلکه برای تعیین روایی محتوایی و صوری یک آزمون، قضاوت متخصصان در این‌باره که پرسش‌های آزمون تا چه اندازه با هدف‌های برنامه مطابقت دارد، ملاک قرار می‌گیرد (Birjandi & Mosallanejad, 2010:166). در این آزمون نیز برای دستیابی به این مهم، فرم نظرسنجی‌ای در اختیار ۱۰ استاد زبان فارسی که در زمان برگزاری آزمون هم در جلسه حضور داشتند، قرار گرفت. معیار نظرسنجی بر اساس شاخص‌های عالی، خوب، متوسط و بد ارائه شده است که نتایج آن در ادامه آمده است.

جدول ۴: نظرسنجی روایی محتوایی
Table 4. Content Validity Survey

بد	متوسط	خوب	عالی	پرسش‌ها
	۳ نفر	۷ نفر		محتوای آزمون
	۱ نفر	۸ نفر	۱ نفر	تناسب سطح پرسش‌ها با اطلاعات زبانی آزمودنی
		۸ نفر	۲ نفر	زمان پاسخ‌دهی به پرسش‌ها
		۱ نفر	۹ نفر	ترتیب پرسش‌های آزمون (شنیدن، خواندن و نگارش)
		۱۰ نفر		تعداد پرسش‌ها و ساختار آزمون

۵-۱-۴-۲. سنجش روایی سازه‌ای

برای محاسبه روایی سازه‌ای از روش همسانی درونی یا تعیین روایی درونی استفاده شده است که این روش مستلزم محاسبه ضریب همبستگی بین اجزا آزمون است.

جدول ۵: همبستگی بین پرسش‌ها در دو بخش شنیدن و خواندن
Table 5. Correlation Between Questions in Listening and Reading

بخش خواندن	بخش شنیدن	آزمون
۰/۹۷۸	۱	ضریب همبستگی شنیدن
۱	۰/۹۷۸	ضریب همبستگی خواندن

مقدار ضریب همبستگی در جدول شماره ۵ نشان می‌دهد که همبستگی بین دو بخش بسیار بالا است؛ یعنی همپوشی بسیار خوبی بین پرسش‌های این دو بخش وجود دارد و همچنین، علامت مثبت ضریب همبستگی به این معنی است که همبستگی پرسش‌های دو بخش در جهت مستقیم است.

۵-۱-۳-۴-۱. ارزیابی عوامل مؤثر بر روایی آزمون

به منظور تحلیل و ارزیابی عوامل مؤثر بر روایی آزمون و تعیین ارتباط پرسش‌های آن با سازه‌های دانش زبان، ناگزیر از به‌کارگیری روش بازنگری^{۱۰} شدیم. در پایان آزمون، پرسش‌نامه‌ای به آزمودنی‌ها ارائه شد. در این پرسش‌نامه، به‌جای استفاده از میزان لیکرت^{۱۱} پنج‌گزینه‌ای معمول در بعضی از تحقیقات، پرسش‌نامه‌ای با چهارگزینه طراحی شده است؛ زیرا تجربیات تحقیقات پیشین نشان داده است که در غیر این صورت آزمودنی‌ها بیشتر تمایل دارند گزینه وسط یا گزینه سوم را بیشتر انتخاب کنند، بدون اینکه به مورد پرسش و نظر شخصی خود عمیق بیندیشند که این امر بر نتایج تحقیق تأثیر می‌گذارد. به همین دلیل در این پژوهش از قرار دادن گزینه‌های خنثی و یا کلی مانند نظر/ عقیده‌ای ندارم و یا تا حدی موافقم که نگرش و نظر شرکت‌کننده را به درستی نشان نمی‌دهد، اجتناب شده است. گفتنی است که این پرسش‌نامه فقط برای مهارت‌های شنیدن، خواندن و نگارش طراحی شده، چون مهارت صحبت کردن به‌صورت شفاهی و در روز دیگری برگزار شده است. پرسش‌های این پرسش‌نامه، شامل متناسب بودن سطح پرسش‌های آزمون با اطلاعات زبانی آزمودنی، کافی بودن زمان برای پاسخ‌گویی به پرسش‌ها، مناسب بودن ترتیب آزمون به صورت شنیدن - خواندن - نگارش، مناسب بودن تعداد پرسش‌ها و ساختار آزمون و سخت‌ترین مهارت زبانی از دیدگاه زبان‌آموزان بوده است. بعد از تجزیه و تحلیل فرم نظرسنجی، چهار پرسش اول مناسب و خوب ارزیابی شد و مهارت نگارش از دیدگاه آزمودنی‌ها سخت‌ترین مهارت آزمون است که این دیدگاه با نتایج میانگین آزمون نگارش همسویی دارد.

۵-۱-۵. شاخص‌های توصیفی نمرات

در اینجا به بررسی شاخص‌های توصیفی نمرات زبان‌آموزان (میانگین، میانه، مد، انحراف معیار، واریانس، مینیمم و ماکزیمم) در ۴ بخش نگارش، مکالمه، شنیدن و خواندن می‌پردازیم. جدول شماره ۶ این شاخص‌ها را نشان می‌دهد.

جدول ۶: شاخص‌های توصیفی نمرات زبان‌آموزان

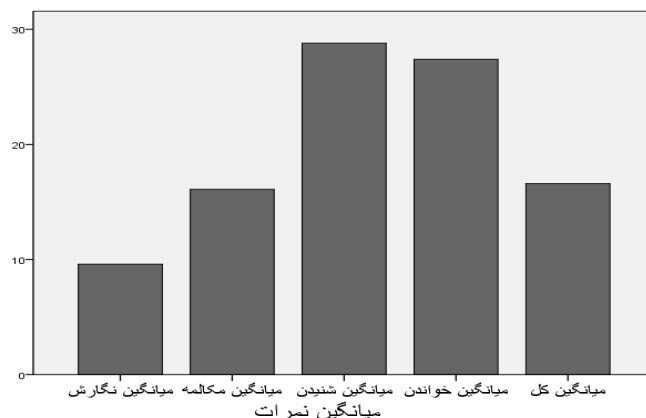
Table 6: Descriptive Indexes of Learners' Scores

شاخص‌ها	نگارش	مکالمه	شنیدن	خواندن
میانگین	۱۰/۵۷۵۰	۱۵/۳۲۱۲	۲۲/۲۹۲۳	۱۸/۲
میانه	۱۰/۸۷۵	۱۵/۷۵	۲۳	۱۸
مد	۱۰	۱۷	۲۲	۱۷
انحراف معیار	۴/۰۱۵۲	۲/۵۹۴۵	۳/۸۲۲۵	۴/۵۰۲۲
واریانس	۱۶/۱۲۲	۶/۷۳۲	۱۴/۶۱۲	۲۰/۳۷۰
مینیمم	۰	۸	۹	۶
ماکزیم	۱۸	۲۰	۳۰	۲۹
تعداد	۱۳۰	۱۳۰	۱۳۰	۱۳۰

این نکته مهم است که آزمون در بخش‌های نگارش و مکالمه از ۲۰ نمره و در بخش‌های شنیدن و خواندن از ۳۰ نمره محاسبه شده است. از جدول ۶ قابل ملاحظه است که میانگین نمرات زبان‌آموزان در بخش شنیدن از سایر بخش‌های آزمون بیشتر است و نشان‌دهنده این است که دانشجویان در مهارت درک شنیداری توانا هستند و عملکرد بهتری دارند. این نتیجه با نتیجه حاصل از میانگین ضریب دشواری مهارت شنیدن همسویی دارد. میانگین نمرات زبان‌آموزان در بخش نگارش از سایر بخش‌ها کمتر است و می‌توان چنین نتیجه گرفت که زبان‌آموزان در بخش نگارش نیاز به تمرین و ممارست بیشتری دارند. این نتیجه با نتیجه به دست آمده از پرسش پنجم فرم نظرسنجی دانشجویان منطبق است؛ زیرا زبان‌آموزان در فرم نظرسنجی، نگارش را سخت‌ترین مهارت زبانی دانستند و اینجا نیز ملاحظه می‌شود که میانگین نمرات در این بخش نسبت به سایر بخش‌ها بسیار پایین‌تر است.

از آنجا که مهارت‌های تولیدی (نگارش و مکالمه) دارای ۲۰ امتیاز و مهارت‌های درکی دارای ۳۰ امتیاز بوده است با مقایسه انحراف معیار و واریانس مشخص شده است که انحراف معیار و واریانس مهارت مکالمه نسبت به مهارت نگارش کمتر بوده است. بدین معنی که نمرات آزمودنی‌ها در بخش مکالمه نسبت به میانگین، دارای پراکندگی کمتری بوده و بنابراین، آزمودنی‌ها در مهارت مکالمه قوی‌تر بودند. علاوه بر میانگین نمرات در بخش‌های مختلف آزمون، میانگین کل آزمون ۱۶/۵۹۷۱ محاسبه شده است. با توجه به میانگین کل در می‌یابیم که زبان‌آموزان توانسته‌اند به

بیش از نیمی از پرسش‌های پاسخ صحیح دهند. نمودار شماره ۳ گویای این مطلب است.



نمودار ۳: میانگین نمرات نگارش، مکالمه، شنیدن، خواندن و میانگین کل آزمون

Graph 3: Average Scores of Writing, Speaking, Listening, Reading and the Average of the Whole Test

۵-۲. عملی بودن^{۱۷} آزمون

با وجود آنکه پایایی و روایی دو خصوصیت مهم یک آزمون هستند، قابل اجرا بودن و عملی بودن آن نیز مهم است. عملی بودن، یعنی اینکه آزمون باید سهولت برگزاری، سهولت نمره‌دهی و سهولت در تفسیر و کاربرد نتایج را داشته باشد. البته، باید توجه داشت که هدف آزمون به دلیل عملی بودن آن به خطر نیفتد (فرهادی و همکاران، ۲۰۰۷: ۱۵۹). آزمون طراحی‌شده، از لحاظ برگزاری، سهولت در نمره‌دهی و سهولت در تفسیر نتایج مناسب بود اگرچه نمره‌دهی آزمون نگارش وقت و دقت بسیاری را می‌طلبد و برگزاری سه مهارت هم زمان و طولانی شدن آزمون سبب خستگی آزمودنی‌ها شده بود.

۵-۳. سودمندی^{۱۸} آزمون

مهم‌ترین چیزی که باید در طراحی و ایجاد یک آزمون زبان لحاظ شود، سودمندی و کاربردی

بودن آن است که یک آزمون به آن منظور طراحی و ایجاد می‌شود. این آزمون هم کاملاً کاربردی است و از آن می‌توان برای تعیین سطح و پایه‌گزینی زبان‌آموز استفاده کرد و بدین وسیله سبب آگاهی فارسی‌آموزان از میزان توانایی خودشان در مهارت‌های زبانی چهارگانه در زبان فارسی و آگاهی استادان از وضعیت زبانی فارسی‌آموزان و طرح‌ریزی شیوه‌های تدریس متناسب با مهارت زبانی آن‌ها می‌شود و جهت‌دهی و جهت‌بخشی به مراکز آموزش زبان فارسی داخل و خارج از کشور را به‌دنبال خواهد داشت. (Bachman & Palmer, 2010:73)

۵-۴. اصیل بودن^{۱۹} آزمون

یکی از اصل‌های مهم در آزمون‌سازی زبان، اصیل بودن آزمون است. مزیک (3: 1996) اصیل بودن آزمون را این‌گونه تعریف کرده است: میزان ارتباط ویژگی‌های فعالیت آزمون زبان با ویژگی‌های فعالیت زبان هدف. هنگامی که ادعا می‌شود که فعالیت یک آزمون واقعی است بدین معنا است که این فعالیت در دنیای واقعی وجود دارد. در صورتی یک آزمون واقعی خواهد بود که زبان آزمون تا حد ممکن طبیعی باشد و پرسش‌ها بافت‌مدار باشند. از آنجا که برای طراحی این آزمون، از زبان رسانه استفاده شده است. پرسش‌ها واقعی و بافت‌مدار بوده و در مهارت مکالمه نیز سعی شده است که پرسش‌ها بر اساس موقعیت‌های واقعی و مورد نیاز گویشور زبان (دانشگاه، بازار، بیمارستان، کتابخانه و ...) و کارکردهای طبیعی پرسیده شود. در ارزیابی غنای واژگانی و دستوری آزمودنی‌ها نیز پرسش‌ها کاملاً بر اساس بافت بوده است و مقوله‌های واژگانی به‌طور جداگانه ارزیابی شدند (Messic, 1996: 18).

۶. تعیین سطح آزمودنی

با بررسی و مقایسه نمرات آزمون‌های مهارتی زبان فارسی کانون زبان، آیلتس و انواع تافل بر اساس شیوه‌نامه مهارت سنجی معتبر و رایج اکتفل^{۲۰} و تطبیق آن با سطوح چارچوب مرجع اروپا، نتایج جدول شماره ۷ به‌دست آمد:

جدول ۷: تعیین سطح آزمودنی
Table 7. Certifiable Proficiency Test

سطوح چارچوب مرجع اروپا نوع کاربر		آزمون تافل اینترنتی	آزمون تافل کامپیوتری	آزمون تافل کاغذی	آزمون آیلتس	آزمون بین‌المللی زبان فارسی
Proficient User	C2	۱۲۰ - ۱۲۰	۳۰۰ - ۲۶۲	۶۲۵ - ۶۸۰	۷/۵ - ۹	عالی: ۹۰ - ۱۰۰
		۱۰۰	۲۵۰	۶۰۰	۷	پیشرفته سطح بالا: ۷۹ - ۸۹
	C1	۹۰ - ۹۱	۲۳۲	۵۷۵	۶/۵	پیشرفته: ۶۸ - ۷۸
Independent User	B2	۷۹ - ۸۰	۲۱۳	۵۵۰	۶	متوسط سطح بالا: ۵۷ - ۶۷
		۶۹ - ۷۰	۱۹۶	۵۲۵	۵/۵	متوسط سطح میانه: ۵۶ - ۶۶
	B1	۵۹ - ۶۰	۱۷۳	۵۰۰	۵	متوسط سطح پایین: ۴۵ - ۳۵
Basic User	A2	۴۹ - ۵۰	۱۵۲	۴۷۵	۴/۵	مبتدی سطح بالا ۳۴ - ۲۴
		۳۹ - ۴۰	۱۳۳	۴۵۰	۴	مبتدی سطح متوسط ۱۳ - ۲۳
	A1	۲۹ - ۳۰	۱۱۳	۴۲۵	۳/۵	مبتدی سطح پایین: ۱ - ۱۲

در این آزمون نیز مانند آزمون آیلتس، تافل و... از مقیاس رتبه‌بندی استفاده شده است، مجموعه امتیاز این آزمون ۱۰۰ امتیاز است که آزمون شنیدن، خواندن، مکالمه و نوشتن به ترتیب ۳۰، ۳۰ و ۲۰ امتیاز است. از آنجا که امتیاز کل این آزمون مانند آزمون بین‌المللی کانون زبان فارسی از ۱۰۰ امتیاز محاسبه شده است، مبنای نمره‌گذاری و سطح‌بندی در این آزمون، از آزمون بین‌المللی کانون زبان فارسی الهام گرفته شده است. در این آزمون قبولی یا ردی مورد نظر نیست؛ بلکه پس از شرکت داوطلب در آزمون، مطابق نمره کسب‌شده، سطح مهارت‌های مختلف زبانی وی مشخص شده است و بر اساس این سطح می‌تواند در دانشگاه‌های ایران به تحصیل یا به کار مشغول شود. گفتنی است که شرکت‌کنندگان این آزمون، حداقل ۲ ماه و حداکثر ۱۲ ماه مشغول یادگیری زبان فارسی بوده‌اند. از میان این تعداد فقط یک آزمودنی بود که به سطح عالی رسیده است و او نیز سال‌هاست که در ایران زندگی می‌کند. ۷۰ درصد آزمودنی‌ها در سطح متوسط ۲۰ درصد سطح پیشرفته، ۵ درصد سطح پیشرفته بالا و ۵ درصد در سطح مبتدی قرار گرفتند. گفتنی است که به‌منظور نتیجه‌گیری دقیق‌تر، بهتر است که این آزمون با ملیت‌های گوناگون و توانایی زبانی متفاوت مجدداً برگزار شود.

به‌دلیل کثرت تعداد شرکت‌کنندگان (۱۳۰ نفر)، تنها سطح‌بندی تعدادی از آزمودنی‌ها به‌عنوان نمونه در جدول شماره ۸ ارائه می‌شود:

جدول ۸: سطح‌بندی آزمودنی‌ها

Table 8. Classification of Testees

		امتیاز کل	مکالمه (۲۰ نمره)	نگارش (۲۰ نمره)	نمره خواندن (۳۰ نمره)	شنیدن (۳۰ نمره)	آزمودنی	سطوح
C2	User Proficient	۹۴	۲۰	۱۷	۲۹	۲۸	عراقی	عالی: ۱۰۰ - ۹۰
		۸۷/۷۵	۱۷	۱۶/۷۵	۲۶	۲۸	ترک	پیشرفته سطح بالا: ۸۹ - ۷۹
		۸۶/۳۵	۱۹	۱۷/۲۵	۲۴	۲۶	چینی	
		۸۵/۷۵	۱۷	۱۶/۷۵	۲۴	۲۸	لبنانی	
		۸۲/۵	۱۹	۱۶/۵	۲۱	۲۶	سوری	

	Proficie User mt	امتیاز کل	مکالمه (۲۰ نمره)	نگارش (۲۰ نمره)	نمره خواندن (۳۰ نمره)	شنیدن (۳۰ نمره)	آزمودنی	سطوح
C1		۷۸	۱۷/۵	۱۱/۵	۲۴	۲۵	فلسطینی	پیشرفته:
		۷۴/۷۵	۱۷/۷۵	۱۰	۲۳	۲۴	هندی	۶۸ - ۷۸
		۷۳	۱۷/۵	۱۰/۵	۲۱	۲۴	تاجیک	
		۷۱	۱۶	۱۳	۱۸	۲۴	افغانی	
		۷۰	۱۶	۱۴	۱۹	۲۱	قزاقی	
B2		۶۲/۵	۱۸	۱۰/۵	۱۱	۲۳	تاجیک	متوسط سطح
		۶۰/۵	۱۶	۱۲/۵	۱۲	۲۰	سوری	بالا: ۶۷ - ۵۷
		۵۷/۵	۱۲	۸/۵	۱۶	۲۱	آذری	
	User Independent	۵۲	۱۰	۴	۱۶	۲۲	لبنانی	متوسط سطح
		۵۱/۵	۱۳	۸/۵	۱۳	۱۷	سوری	میانه: ۴۶ - ۵۶
		۴۶	۱۰	۵	۱۴	۱۷	لبنانی	
B1		۴۴	۱۰	۳	۱۲	۱۹	چینی	متوسط سطح
		۳۹	۸	۳	۱۴	۱۴	چینی	پایین: ۳۵ - ۴۵
		۳۸/۵	۱۵/۵	۰	۹	۱۴	صومالی	
A2	Basic Use	۳۳	۸	۵	۱۰	۱۰	چینی	مبتدی سطح بالا ۲۴ - ۳۴
		۲۳	۴	۴	۷	۸	چینی	مبتدی سطح متوسط: ۱۳ - ۲۳
A1								مبتدی سطح پایین: ۱ - ۱۲

۶- ۱. تطبیق سطح آزمودنی‌ها با نظر استادان

از آنجا که پیکره و چارچوب دقیقی برای تعیین سطح آزمودنی در زبان فارسی وجود ندارد و شیوه تعیین سطح در این پژوهش با استناد به معیارهای خارجی است، از این رو، به منظور اعتباربخشی بیشتر به آزمون طراحی شده، نمرات آزمون در اختیار اساتید آزمودنی‌ها قرار گرفت تا آنان با مقایسه نمره آزمودنی و سطح او در کلاس‌های درس، نظر خود را درباره نتایج آزمون بازگو کند. جدول شماره ۹ نشان‌دهنده متناسب بودن نمره آزمودنی با توانایی زبانی آزمودنی، از دیدگاه اساتید است:

جدول ۹: تطبیق سطح آزمودنی‌ها با نظر استادان

Table 9. Adjustment of the Level Testees with the Professors' Opinion

متناسب بودن نمره آزمون با توانایی زبانی آزمودنی	کمتر بودن نمره آزمون در مقایسه با توانایی آزمودنی	بیشتر بودن نمره آزمون در مقایسه با توانایی زبانی آزمودنی
٪۸۳/۸۷	٪۶/۴۵	٪۹/۶

۷. پیشنهادات

- با توجه به مسائلی که در این پژوهش بررسی شد، می‌توان موضوعاتی که در ادامه آورده می‌شود را به عنوان پیشنهاداتی برای تحقیقات آینده مطرح کرد:
۱. بررسی تأثیر محیط دوم یا خارجی در میزان موفقیت آزمون‌دهندگان در آزمون مهارت.
 ۲. بررسی تأثیر زبان مادری آزمون‌دهندگان در میزان موفقیت آزمون‌دهندگان در آزمون مهارت.
 ۳. تهیه آزمون مهارت بر اساس پیکره سطح‌بندی‌شده زبان فارسی و نیازهای عمومی غیر-فارسی زبانان.
 ۴. تهیه آزمون‌های پیشرفت در سه سطح پایه، متوسط و پیشرفته.

۸. نتیجه

پژوهش حاضر بر پایه مفهوم چند بعدی زبان بر گرفته از نظریه استرن، آزمون مهارت زبان فارسی را ارائه کرده است. این پژوهش خلاف پژوهش موسوی (۱۳۷۹)، قنصولی (۱۳۸۹) که فقط به مهارت‌های دریافتی زبان توجه کرده‌اند، دارای ارزیابی مهارت‌های تولیدی زبان (گفتار و نوشتار) نیز است و مانند دو پژوهش پیشین از نوع واحد - منفصل نیست و توانایی دستور و غنای واژگانی را در پرتو چهار مهارت زبانی می‌سنجد نه به صورت جداگانه. در این آزمون، میانگین نمرات زبان‌آموزان در بخش نگارش از سایر بخش‌ها کمتر است و می‌توان چنین نتیجه گرفت که زبان‌آموزان در بخش نگارش نیاز به تمرین و ممارست بیشتری دارند. این نتیجه با نتیجه به دست آمده از پرسش پنجم فرم نظرسنجی دانشجویان منطبق است؛ زیرا زبان‌آموزان در فرم نظرسنجی، نگارش را سخت‌ترین مهارت زبانی دانستند و اینجا نیز ملاحظه می‌شود که میانگین نمرات در این بخش نسبت به سایر بخش‌ها بسیار پایین‌تر است که این نتیجه با نتیجه - گیری قنصولی مبنی بر دشوار بودن بخش دستور تقریباً هم‌پوشانی دارد و چنین برداشت می‌شود که در آموزش دستور و مهارت‌های نگارش باید توجه خاصی مبذول شود. به منظور استانداردسازی آزمون، پرسش‌های آزمون در ابتدا به صورت پیش‌آزمون در آزمون‌های پایان ترم بهار ۹۴ و تابستان ۹۴ مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی^(۵) استفاده شد و بعد از آسیب‌شناسی، پرسش‌هایی که به سنجش کلمات و عبارات کم‌سامد پرداخته بود و یا گزینه انحرافی خوبی نداشت و دارای ضریب تمیز پایین بود و پرسش‌هایی که به نوعی با حافظه، بهره هوشی و استدلال و درک غیرزبانی نیاز داشت از کل پرسش‌ها حذف شد. با بررسی نتایجی که در بخش تجزیه و تحلیل آزمون و اعتبارسنجی آن به دست آمد، می‌توان چنین ادعا کرد که آزمون طراحی شده، مطابق با معیارهای آزمون‌سازی طراحی شده است؛ زیرا در بررسی ضریب دشواری مشخص شد که میانگین ضریب دشواری پرسش‌ها خواندن بر اساس معیارهای آزمون‌سازی مناسب و بهینه بوده است. نتیجه میانگین ضریب دشواری پرسش‌ها شنیدن آسان ارزیابی شد که پس از آسیب‌شناسی مشخص شد که این نتیجه متأثر از ملیت بعضی از آزمودنی‌ها و مدت یادگیری زبان فارسی بعضی از آنان بوده است؛ زیرا از میان آزمودنی‌ها ۱۵ نفر تاجیک بودند و ۲۴ نفر هم به مدت ۱ سال در دانشگاه زبان‌های خارجی پکن مشغول یادگیری زبان فارسی بوده‌اند. از این رو، توانایی مهارت درک شنیداری آنان در سطح بالایی بود و همین

امر روی میزان ضریب دشواری تأثیر داشته است. میزان ضریب آلفای کرونباخ آزمون ۰/۷ به- دست آمد که نشان‌دهنده پایا بودن آزمون است که ضریب آلفای کرونباخ این پژوهش با پژوهش قنسولی (۱۳۸۹) یکسان ارزیابی شده است. روایی محتوایی آزمون نیز که شامل ارزیابی محتوای آزمون، تناسب سطح پرسش‌ها با اطلاعات زبانی آزمودنی، زمان پاسخ‌گویی، ترتیب پرسش‌ها و تعداد پرسش‌ها و ساختار آزمون بوده است، بر اساس قضاوت اساتید زبان فارسی، خوب ارزیابی شده است. پس از محاسبه ضریب همبستگی پیرسون، روایی سازه‌ای بین مهارت خواندن و شنیدن نیز مناسب ارزیابی شد. پیش‌بینی می‌شود که آزمون طراحی‌شده سودمندی لازم را نیز داراست؛ زیرا از این آزمون، می‌توان برای تعیین سطح و پایه‌گزینی زبان‌آموز استفاده کرد و بدین‌وسیله فارسی‌آموزان از میزان مهارت خودشان و استادان از وضعیت زبانی فارسی‌آموزان آگاهی می‌یابند. یکی دیگر از ویژگی‌های یک آزمون خوب، قابل اجرا و عملی بودن آن است که آزمون طراحی‌شده نیز دارای سهولت در برگزاری، سهولت در نمره‌دهی و سهولت در تفسیر و کاربرد نتایج است. درنهایت، در پاسخ به پرسش این پژوهش می‌توان، چنین نتیجه گرفت که آزمون طراحی‌شده دارای معیارها و ویژگی‌های آزمون‌سازی و مطابق با معیارهای استاندارد طراحی شده است.

۹. پی‌نوشت‌ها

1. retrospection
2. Spss
3. Delphi
4. assessing writing
5. analytic scale
6. holistic scale
7. difficulty index
8. discriminative index
9. reliability
10. inter- examiner reliability
11. validity
12. content validity
13. face validity
14. - retrospection
15. Likert scale

- 16. practicality
- 17. usefulness
- 18. authenticity
- 19. the American council on the teaching of foreign languages

۱۰. منابع

- هریسون، اندرو (۱۳۷۳). *کتاب راهنمای ارزشیابی زبان*. ترجمه اسماعیل فقیه. تهران: الزهرا.
- براون، داگلاس (۲۰۱۱). *ارزشیابی زبان: اصول و عملکردهای کلاس*. ترجمه علیرضا فرح-بخش. تهران: رهنما.
- جلیلی، اکبر (۱۳۹۰). *آزمون مهارتی فارسی (آمفا) بر پایه چهار مهارت اصلی زبان*. پایان‌نامه کارشناس ارشد رشته آموزش زبان فارسی به غیرفارسی‌زبانان. دانشگاه علامه طباطبائی.
- ربکا، م والت (۱۳۷۲). *آزمون در آموزش زبان*. ترجمه گیتی دیهیم. تهران: مرکز نشر دانشگاهی.
- سیف، علی اکبر (۱۳۹۰). *اندازه‌گیری، سنجش و ارزشیابی آموزشی (ویرایش پنجم)*. تهران: دوران.
- شقاقی، ویدا و همکاران (۱۳۹۰). «هنجاریابی آزمون بین‌المللی زبان فارسی». پازند. ش ۲۴ - ۲۵. صص ۸۷ - ۹۹.
- قنصولی، بهزاد (۱۳۸۹). «طراحی و روایی‌سنجی آزمون بسندگی زبان فارسی». پژوهش زبان‌های خارجی. ش ۵۷. صص ۱۱۵ - ۱۲۹.
- ممقانی، هانیه (۱۳۹۰). *ارزیابی روایی و پایایی آزمون وزارت علوم و تحقیقات و فناوری (MSRT) با هدف دستیابی به یک آزمون مهارت برای زبان فارسی*. پایان‌نامه کارشناس ارشد رشته آموزش زبان فارسی به غیرفارسی‌زبانان. دانشگاه علامه طباطبائی.
- موسوی، ثریا (۱۳۷۹). *بررسی راهکارهای نظری آزمون‌های بسندگی و تهیه یک آزمون بسندگی استاندارد زبان فارسی*. پایان‌نامه کارشناس ارشد رشته آموزش زبان فارسی به غیرفارسی‌زبانان. دانشگاه شیراز.
- مؤمنی، منصور و قیومی، علی (۱۳۸۹). *تحلیل‌های آماری با استفاده از SPSS*. تهران: مؤلف.

[DOR: 20.1001.1.23223081.1398.10.6.13.5] IRST on Tuesday February 4th 2020

References:

- Bachman, L. & A. Palmer, (2010), *Language Assessment in Practice*. Oxford University Press. USA.P 73.
- Birjandi, P. & P. Mosallanejad, (2010), *An Overview of Testing and Assessment*. Esfahan: Sepahan Publication. P 166.[In Persian].
- Brindley, G. (2003). *Classroom-based Assessment*. in D. Nunan (ed.) *Practical English Language Teaching*. US: McGraw-Hill.P. 317.
- Carter, R. & D. Nunan, (2001), *The Cambridge Guide to Teaching English to Speakers of other Languages*. (4thed.).Cambridge: Cambridge University Press.p 137.
- Coombe, C. & K. Folse & N. Hubley, (2007), *A Practical Guide to Assessing English Language Learners*. US: The University of Michigan Press. Xvii.
- Douglas, B. (2011). *Language Evaluation: Principles and Classroom Practice*. Translated by Farahbakhsh, Alireza. Tehran: Rahnama Publ. [In Persian].
- Farhadi, H.; A. Jafarpour & P. Birjandi, (2007), *Testing Language Skills from Theory to Practice*. Tehran: SAMT. P 148-159.
- Ghonsoli, B. (2010). "Designing and validation of Persian language proficiency test". *Foreign Language Research*. No.57. Pp .115-129.[In Persian].
- Harrison, A. (1994). *Language Evaluation Guidelines*. Translated By Ismael Faghih. Tehran: Al-Zahra Publ. [In Persian].
- Hughes, A.(2004). *Testing for Language Teachers*. Cambridge University Press.P 11.
- Jalili, A. (2011). *Skill based Persian Language Test on the Basis of 4 Main Language Skills*. Thesis for obtaining MA degree in the Teaching Persian Language for Non-Natives Major. Allameh Tabatabaee University. [In Persian].
- Mamghani, H. (2011). *Validity and Reliability Assessment of Ministry of Science, Research and Technology's Examination with the Intention of Achieving to a Skill Based Persian Language Proficiency test*. [In Persian].

- Messic, S. (1996). *Validity and Washback in Language Testing* .Educational testing Service. princeton, Newjersey. Issue(1).Pp. 1-18.
- Moumeni, M.r & A. Ghayomi, (2010), *Analytical Analysis with the Utilization of SPSS*. Theran: Moallef Publ. [In Persian].
- Mousavi, S. (2000). *Investigating theoretical approaches of proficiency tests and producing a standard Persian language proficiency test. MA Dissertation in Teaching Persian language to non-native*. Shiraz uunuversity. [In Persian].
- Seif , A. A. (2011). *Pedagogical Measurement, Examination and Evaluation* (5th Edition). Tehran: Davaran Publ. [In Persian].
- Shagagi et. al., (2011), “Standardization of international Persian language proficiency test”. *Pazand Publ.* No.24-25. Pp.87-99.[In Persian].
- Stern, H.H. (1983). *Fundamental Concepts of Language Teaching* .Oxford university Press. Pp . 346-356.
- Walt M. Rebeca, (1993), *Examination in Language Teaching. Translated by Dihim,Giti*. Tehran. University Publication Centre. [In Persian].
- Weigle, S. C. (2008). *Assessing Writing*. Cambridge University Press. P 116

Planning and validating Persian language Standard test (On the Basis of Four Language Skills)

Bahman Zandi^{1*}, Belgheis Rovshan², Amirreza Vakilifard³,
Leila Golpour⁴

1. Professor of Linguistics , Department of Linguistics, Payam Noor University , Tehran, Iran.
2. Associate Professor of Linguistics , Department of Linguistics, Payam Noor University, Tehran , Iran.
3. Associate Professor , Department of Linguistics, Imam Khomeini International University, Qazvin , Iran.
4. Ph.D. Department of Linguistics, Payam Noor University, Tehran , Iran.

Received: 12/05/2018

Accepted: 16/10/2018

Abstract

Goal: This study aims to develop a language-based skill test revolving around four main language skills, with principals and basics of language test development in mind intended to evaluate Persian language proficiency of non-Iranian students (In terms of language level). One of the main goals of this study is to develop a Persian language skill (proficiency) test. The purpose of language proficiency test is to measure one's overall ability to utilize the desired language without having to undergo specific training courses. So the content of a language proficiency test is not based on the content or purpose of the language courses. Also in this study, measures are foreseen to evaluate the validity and reliability of the designed test in order to obtain the Credibility of the standardized test.

Research Question:

The questions raised in this study are as follows:

- What are characteristics of designed tests in terms of analysis of test making procedures?
- To what extent does the test designed conform to the standard criteria?

* Corresponding Author's E-mail: zandi@pnu.ac.ir

Research hypothesis:

This test possesses the following characteristics in terms of analysis of test making procedures:

This test incorporates the necessary efficiency

This test is reliable.

This test is valid.

This test is applicable and practical.

This test is designed according to standardized criteria.

Research method:

In this research a descriptive-analytical method is firstly used to study previous researches, and to select a theoretical framework in order to present the theoretical concepts and basics; to construct the criterion test and standardize it the “agreement-oriented scaling method” was used, and to determine the validity of the proposed test, the “field method” was used. The statistical population of this study is the learners of Persian Language Learning Centre of Imam Khomeini International University of Qazvin. The total population of this statistical population is 200 people. The sample size of this study will be 130 members of this statistical population who have been learning Persian language for at least 2 months and a maximum of 1 year. Their mean age is 38-38 years. The sampling method utilized is cluster random sampling.

Conclusion:

The current study has presented the Persian language proficiency test based on the multidimensional concept of language derived from Stern theory. Since it was found that the average difficulty coefficient of reading questions was appropriate and optimal compared to the criteria of test designing, It can be claimed that the designed test were in accordance with the test designing criteria. Difficulty coefficient of listening comprehension questions were assessed easy, which after flaw scrutinizing was found that it was influenced by the nationality of some of the testees and the duration of Persian language learning. Since 15 of the subjects were Tajik and 24 of them have been leaning Persian in Faculty of Foreign Languages of Beijing for 1 year, they possessed a high listening comprehension skill and this had an impact on the difficulty coefficient factor. Cronbach's alpha coefficient of the test was 0.7, indicating the reliability of the test. The content validity of the test was also evaluated well, based on the judgment of the Persian language teachers. After calculating Pearson's correlation coefficient, structural validity between reading and listening skills was also evaluated

well. The designed test can be used for placement since it also possesses the needed efficiency, it can be used to for placement of subjects and determine the base level of language learner, thus informing the Persian language learners of their own skill and making Professors cognizant of Students' Linguistic Status. Another attribute of a good test is its applicability and practicality, which the designed the test is also easy to conduct, easy to score and easy to interpret and apply. Finally, in response to the research question, it can be concluded that the designed test possesses the test designing criteria and features, and has been designed in accordance with standard criteria.

Key words: Testing, Language skills, Persian language standard test